Högskoleexamen med inriktning mot Programvaruteknik
med specialisering Webbprogrammering
Maj 2025

URI: urn:nbn:se:bth-27944

Webbaserad kommunikation via SMS

Integrerad LLM för effektiv meddelandehantering

Calle Andersson

Fakulteten för datavetenskaper, Blekinge Tekniska Högskola, 371 79 Karlskrona, Sverige

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Higher Education Diploma in Software Engineering with emphasis in Web Programming. The thesis is equivalent to 10 weeks of full-time studies.

The author declare that they are the sole author of this thesis and that they have not used any sources other than those listed in the bibliography and identified as references. They further declare that they have not submitted this thesis at any other institution to obtain a degree.

Contact Information:
Author:
Calle Andersson
E-mail: caas23@student.bth.se

External advisor:
Björn Skyttberg
46elks

University advisor:
Lecturer Philip Berglund
Department of Computer Science

Faculty of Computing
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

Artificial intelligence has become part of everyday life for most people, either directly or indirectly. More and more companies are seeing the benefits of including language models in various applications, and at the same time, research shows that this indeed is a way to increase turnover.

The purpose of this study is to investigate how a language model and an SMS API can be implemented together in a system, to enable more efficient communication between selling companies and their customers. In addition, the study aims to clarify how order information can be presented so that it is both easy to understand and easy to overview.

A requirements and development-based method is used to design and implement a prototype that includes both an API and a language model. A test group is used to test the prototype in an order scenario, after which they answer a survey to summarize their experiences. A literature study is conducted to provide a better understanding of how data can be presented in the given scenario.

The study shows that in-context prompting in a structured manner is a beneficial method for extracting information from SMS, which can then be interpreted and responded to. At the same time, the literature review shows that there is a lack of consensus on how data should be presented, as individual preferences make it difficult to give a general opinion.

The study shows that it is possible to effectively integrate an SMS API with a language model, but that further work is necessary to ensure that spontaneous misinterpretations, for example, can be avoided. In addition, the study shows that further work is needed to provide a clear picture of how the data should be presented, and that this may require the use of methods beyond the current one.

Keywords: language model, LLM, SMS API, prompting, data presentation

Sammanfattning

Artificiell intelligens har för de flesta kommit att bli en del av det vardagliga livet, antingen direkt eller indirekt. Fler och fler företag ser vinsterna med att inkludera språkmodeller i olika applikationer, samtidigt som undersökningar ger stöd för att detta är ett sätt att öka omsättningen.

Syftet med denna studie är att utreda hur en språkmodell och ett SMS-API kan implementeras i samspel med varandra, för att möjliggöra effektivare kommunikation mellan säljande företag och dess kunder. Dessutom syftar studien till att klargöra hur orderinformation kan presenteras på sådant sätt att denna enkelt kan överblickas.

En krav- och utvecklingsbaserad metod används för att designa och implementera en prototyp där både API och språkmodell finns integrerade. För att testa detta i ett orderscenario nyttjas en testgrupp, som efter genomfört test även besvarar en enkät för att sammanfatta hur de upplevde orderflödet. En litteraturstudie genomförs för att ge en bild av hur datan kan presenteras.

Studien visar att *in-context*-promptning enligt ett strukturerat schema är en gynnsam metod för att extrahera information från SMS, som sedan kan tolkas och besvaras. Samtidigt visar litteraturstudien att det saknas konsensus kring hur aktuell data bör presenteras och att individuella preferenser gör det svårt att generalisera på gruppnivå.

Arbetet visar att det på ett effektivt sätt går att integrera ett SMS-API med en språkmodell, men att fortsatt arbete är nödvändigt för att exempelvis säkerställa att spontana avvikelser i kommunikationen kan undvikas. Dessutom visar studien att fortsatt arbete krävs för att ge en tydlig bild av hur datan bör presenteras, samt att detta kan kräva nyttjande av metoder utöver den aktuella.

Nyckelord: språkmodell, LLM, SMS-API, promptning, datapresentation

Förord

Jag vill rikta ett stort tack till Björn Skyttberg och 46elks som gett mig möjligheten att genomföra mitt examensarbete hos dem.

Tack även till handledare Philip Berglund och kursansvarig Emil Folino som båda gett guidning och synpunkter under arbetets gång.

Tack till samtliga som tagit sig tid att bidra till detta arbete, genom testning och utvärdering.

Slutligen vill jag tacka Emmy som, genom hela utbildningen som i slutändan möjliggjort detta arbete, har visat stöd och uppmuntran i alla lägen.

Calle

Innehåll

Abstract i

Sammanfattning ii

Förord iii

1 Introduktion 1
- 1.1 Bakgrund . . . 1
- 1.2 LLM . . . 2

2 Forskningsfrågor 3
- 2.1 Mål . . . 4
- 2.2 Avgränsningar . . . 4

3 Relaterat arbete 5
- 3.1 LLM i en generell kontext . . . 5
- 3.2 SMS-hantering utan LLM . . . 6
- 3.3 SMS-hantering med LLM . . . 7

4 Metod 10
- 4.1 Litteraturstudie . . . 10
- 4.2 Skapa en artefakt . . . 11
  - 4.2.1 Krav . . . 12
  - 4.2.2 Webbklient och server . . . 14
  - 4.2.3 SMS-API . . . 15
  - 4.2.4 LLM . . . 15
- 4.3 Utökad litteraturstudie . . . 19

5 Resultat och analys 20
- 5.1 Artefakt . . . 20
- 5.2 Datapresentation . . . 26
  - 5.2.1 Litteraturgenomgång . . . 26
  - 5.2.2 Sammanfattning . . . 29

6 Diskussion 31
- 6.1 RQ1 . . . 31
- 6.2 RQ2 . . . 32

7 Slutsatser och framtida arbete **33**
7.1 RQ1 33
7.2 RQ2 34
7.3 Etiska aspekter 34

Referenser **35**

A Orderflöde **38**

Figurer

4.1 Webbklient för att skicka erbjudanden via SMS-API. 14
4.2 Exempel på hantering av tydlig respektive otydlig beställning. 18

5.1 Utdrag ur konversation där kundens önskemål blev felaktigt tolkade. . 26

Kodexempel

4.1 Generell funktion för LLM-anrop. 15
4.2 Komprimerat inputschema för extractOrderDetails. 16
4.3 Komprimerat outputschema för extractOrderDetails. 16

Kapitel 1

Introduktion

Artificiell intelligens (AI) är ett växande område i samhället och ett alltmer aktuellt ämne inom både forskning och utveckling. Antalet engelskspråkiga publikationer nästan tredubblades under perioden 2010 - 2022, med flertalet insikter som lett till signifikanta framsteg inom såväl vetenskapliga som medicinska områden [21, s. 31, ss. 296–324], [23]. Dessutom kan en ökad användning av AI ses världen över, inom en rad olika organisationer och sektorer [21, ss. 258–259], [23]. Det ökade nyttjandet av AI inom olika branscher har i vissa fall även kunnat påvisa resultat som innefattar kostnadsreduceringar med upp till 55 %, medan intäkter ökat med upp till 66 % [21, s. 264], [23].

Samtidigt som många framsteg och positiva effekter kan relateras till utvecklingen av AI, bör det noteras att det även finns en mer försiktig inställning till den påverkan som AI potentiellt kan ha på våra liv. Undersökningar gjorda av Ipsos visar att antalet respondenter som uppgav att produkter och tjänster som använder AI gör dem nervösa, ökade med 13 procentenheter från år 2022 till 2023. Vidare visar samma undersökningar att inställningen till AI varierar både geografiskt och demografiskt, där den yngre generationen tillsammans med högutbildade och höginkomsttagare tenderar att ha en mer optimistisk syn på området, medan lågutbildade och låginkomsttagare uttrycker större skepsis [21, s. 438], [16].

1.1 Bakgrund

Arbetet genomförs i samarbete med 46elks, ett företag som primärt tillhandahåller ett *Application Programming Interface* (API) för SMS och telefoni [2]. API:t finns väldokumenterat för att enkelt komma igång med integration i valfri applikation, i det programmeringsspråk som önskas. Utöver möjligheten att skicka SMS, finns även stöd för exempelvis MMS, samtal och virtuella telefonnummer [1].

Företaget erbjuder även tjänsten *Just Send!*, en webbklient där deras API finns integrerat för att skicka SMS direkt från webbläsaren [19]. Som en utökning till detta önskar man utreda hur en tjänst som sådan kan vidareutvecklas för att även inkludera möjligheten att motta och tolka SMS med hjälp av AI, för att på så sätt eliminera behovet av manuell hantering av inkomna meddelanden. För att konkretisera frågeställningen utgås det ifrån en bageriverksamhet som planerar att nyttja tjänsten för att nå ut till kunder med säljerbjudanden, även om konceptet som sådant även är tänkt att kunna omsättas till annan verksamhet. Det flöde som behöver stödjas inkluderar möjlighet att:

- låta bagaren skicka SMS från en webbklient, till specifika kunder.

- motta och bearbeta SMS från varje enskild kund.
- svara på SMS, för att bekräfta eller ytterligare klargöra en beställning.
- temporärt lagra information kring pågående och bekräftade beställningar, för att möjliggöra vidare hantering.

För att tolka inkomna svar samt hantera bekräftelse och ytterligare kommunikation, kommer en stor språkmodell – det vill säga en Large Language Model (LLM) – att användas. Med LLM som verktyg för kommunikation behöver bagaren enbart initiera en första kontakt genom ett SMS-utskick, medan fortsatt kommunikation kan ske utan att bagaren aktivt deltar i dialogerna. Genom att bekräftade beställningar lagras, kan dessa i ett senare skede presenteras i webbklienten. Detta gör att bagaren kan ta del av nödvändig information även om denne inte aktivt deltagit i de konversationer som ligger till grund för beställningarna.

1.2 LLM

En stor språkmodell är en specifik tillämpning inom naturlig språkbehandling - Natural Language Processing (NLP) [12]. NLP är ett område som innefattar flertalet tekniker för analys och representation av mänskligt språk, om än med vissa begränsningar, då traditionell NLP bygger på algoritmer som i praktiken kan ha svårt att fånga den fulla kontexten av ett textstycke. En tydlig begränsning hos en NLP-modell, jämfört med hur människor generellt tolkar information, är deras förmåga att placera en text i ett större sammanhang om detta inte specificerats i den text som analyseras [7, s. 604]. Ett försök att manuellt hantera sådana svårigheter innebär ett omfattande arbete med att träna modellen på stora mängder data, men även då kvarstår ofta problem som innefattar bristande precision, oförmåga att uppfatta ton och känsla samt fortsatta begränsningar på grund av avsaknad av omvärldskunskap i relation till innehållet [34, s. 1].

En LLM är en modell som i grunden är designad för att förstå, tolka och generera stora textmängder i olika kontexter, vilket i mångt och mycket innebär att behärska uppgifter där en traditionell NLP påvisat begränsningar. En LLM har tränats på en stor mängd data med otaliga parametrar, vilket resulterat i en förmåga att hantera textanalyser på en djupare nivå än en traditionell NLP. Som en följd av den omfattande träningen har modellen även kunnat uppvisa färdigheter inom områden utöver de som den specifikt tränats för, detta gäller exempelvis översättning och textsummering [22, ss. 6–9]. Det är dock viktigt att notera att detta beteende inte enbart uppfattas som positivt, utan även har lett till oro kring huruvida modeller som dessa kan utveckla okontrollerade förmågor [21, 1, s.141].

Kapitel 2

Forskningsfrågor

Ett tillförlitligt system behöver konstrueras utifrån kundens krav och önskemål. Hur kan detta uppnås samtidigt som det säkerställs att inkomna svar tolkas och hanteras på ett effektivt och korrekt sätt, utan manuell intervention? Hur kan användare av systemet ta del av inkommande data, utan att själv bearbeta eller strukturera denna?

För att utreda detta har två primära frågeställningar (RQ) formulerats:

RQ1. Hur kan SMS-svar från kunder tolkas och bearbetas med hjälp av en stor språkmodell?

Målet är att utreda vilka krav som ställs på implementationen, för att uppnå önskat resultat. Att en stor språkmodell på ett tillfredsställande sätt *kan* implementeras för detta ändamål, anses redan klarlagt genom det omfattande bakgrundsmaterial som finns att tillgå på ämnet [34, s. 1], [24, s. 2, s. 32], [3, s. 5050]. Syftet är att utreda hur implementation bör ske för att säkerställa att modellen är anpassad för given kontext.

RQ2. Hur kan insamlad data presenteras på sådant sätt att användaren av systemet enkelt kan överblicka resultaten?

Målet är att ge en bättre bild av hur data bör presenteras för att på ett tydligt sätt återge den information som framkommit via SMS-konversationerna. För att utreda detta formuleras ett antal underfrågor som syftar till att ge en bättre förståelse för den data som ska presenteras samt den eller de användare som ska ta del av denna. Frågorna, som kan ses listade nedan, besvaras antingen genom litteraturstudier eller genom att se till krav och villkor som definierats i samband med RQ1.

- Vilken typ av data ska presenteras och hur är denna strukturerad?
- Vad är syftet med presentationen av datan och vilken roll spelar den tilltänkta målgruppen i sammanhanget?
- Vilken presentation är vanlig för olika typer, och storlek, av data?
- Finns det några riktlinjer kring de olika sätten att presentera data?
- Finns det några riktlinjer kring interaktiva element, animationer och visuella effekter?

2.1 Mål

Arbetets primära mål är att klargöra hur en stor språkmodell kan integreras i ett system tillsammans med ett SMS-API, för att extrahera information från inkomna SMS. Dessutom behöver det utredas hur den extraherade informationen kan och bör presenteras för systemets användare.

Utöver de primära målen definieras även underordnade mål utifrån en rad olika aspekter, där de huvudsakliga målen innefattar:

- att leverera ett arbete som lever upp till intressenters och egna förväntningar.
- att få en bättre förståelse för uppbyggnad, roll och nyttjande av:
  - SMS-API:er, som en integrerad del av webben.
  - språkmodeller, med särskilt fokus på tillämpning av LLM.
- att identifiera och få en bild av vilka parametrar som har en betydande roll vid val av tillvägagångssätt för att presentera aktuell data.

2.2 Avgränsningar

För att ge utrymme för arbetets huvudsakliga fokus – att samla in, tolka, bearbeta och presentera svarsdata – kommer 46elks API nyttjas i sin nuvarande form utan några modifieringar. Dessutom kommer programmeringsspråk, vid implementation av en prototyp för testning av systemet, att väljas godtyckligt utan utvärdering av vad som kan anses vara bäst lämpat i detta fall. Notera dock att samarbetsföretaget i vissa fall uttryckligen har uppgett att en specifik tjänst, källa eller modell ska nyttjas. I dessa fall kommer detta att framgå i rapporten, för att tydliggöra delar som är faktiska krav och önskemål från kundens sida.

Då arbetet inte avses resultera i en färdig produkt kommer ingen ytterligare dokumentation, utöver denna rapport, att levereras med arbetet. Den implementerade prototypen och den kod som hör därtill, kommer inte inkludera några tester eller särskilda verktyg för upprätthållande av kodstandard eller dylikt, även om grundläggande kodstandarder kommer beaktas och följas.

För att testa och utvärdera den implementerade prototypen för RQ1 kommer ett antal frivilliga testpersoner genomföra ett test. Detta test kommer att följas upp med en kort enkät, för att ge ytterligare underlag till resultat och diskussion. Det är dock värt att betona att enkäten inte är en egentlig del av den Design Science-metod som valts (se kapitel 4.2), utan adderas för att följa upp nämnt test. Studien bör alltså inte ses som en enkätstudie och har inte för avsikt att genomföras som en sådan.

För studien som helhet, och i synnerhet den andra frågeställningen, utelämnas aspekter gällande tillgänglighet och anpassning därefter. Frågeställningen kommer att besvaras utifrån ett antagande om att användare av systemet inte är i behov av särskild anpassning för att hantera exempelvis nedsatt syn, motoriska svårigheter eller begränsad kognitiv förmåga.

Slutligen bör nämnas att ekonomiska aspekter genomgående har utelämnats.

Kapitel 3

Relaterat arbete

Följande kapitel indelas i tre delar[1], där den första fokuserar på LLMs samt applicering och hantering av dessa i mer generella kontexter. Den andra delen riktas i huvudsak mot SMS och hur information från dessa kan extraheras utan nyttjandet av en LLM. Den tredje delen behandlar relaterat arbete där tidigare nämnda delar i någon mening integrerats tillsammans - även om det visat sig att material kring detta är något begränsat, åtminstone i ett orderflödessammanhang. Genom de litteratursökningar som gjorts har det framkommit att just kombinationen av SMS-tolkning tillsammans med LLM-hantering, inte är helt vanligt förekommande sett till antal studier och forskning på området. Detta konstaterande indikerar dock att det område som studien kretsar kring är av relevans samt att studiens resultat har goda möjligheter att bidra till en bättre förståelse för hur dessa delar faktiskt kan implementeras i samspel med varandra. Samtliga delkapitel avslutas med ett sammanfattande stycke, där erfarenheter från de tidigare studierna sätts i relation till den egna studien.

3.1 LLM i en generell kontext

Det har redan tidigare slagits fast att LLMs i sin rena form, oavsett appliceringsområde eller egentlig inriktning, har egenskaper som är användbara vid mer avancerade uppgifter, däribland textanalyser [34], [24], [3]. Att modellerna överlag presterar väl är dock inget slumpartat beteende. En modells prestation kan i många fall sättas i direkt relation till hur väl den promptats – det vill säga, vilken information den fått i ett givet användarfall. [34], [4] och [36] betonar alla vikten av att förse en LLM med prompter som är både informativa och välformulerade, då dessa faktorer kan ha stor betydelse för hur väl en språkmodell fungerar. Dessutom, förutom att initialt ges väl genomtänkta prompter, kan en LLM programmeras på sådant sätt att ett förtydligande krävs innan ett svar ges – för att ytterligare säkerställa att resultatet är i linje med det som väntas [4].

[24], [22], [20], [9], [36] och [11] ger alla en bredare bild av just promptning och språkmodeller. Naveed m.fl. [24] beskriver hur promptning kan ske enligt ett antal olika tillvägagångssätt – däribland *zero-shot* och *few-shot* (eller *in-context*). Det förstnämnda innebär att modellen ges en uppgift – till exempel en fråga att besvara – utan att ges några exempel i prompten medan det andra tillvägagångssättet innebär att modellen ges ett antal exempel i prompten, för att på så sätt ges en kontext som hjäl-

[1] Litteratur för RQ2 behandlas i avsnitt 5.2

per modellen i rätt riktning. Författarna beskriver även hur *in-context*-promptning minimerar risken att en språkmodell uppvisar oönskade beteenden såsom felaktiga, skadliga eller partiska responser. Vidare påpekar de att även om *zero-shot*-modeller kan ses som fullt kapabla vad gäller att besvara frågor som aldrig setts tidigare, så underpresterar de jämfört med modeller som ges ledande exempel via prompten.

Kojima m.fl. [20] menar att en *zero-shot*-modell som enbart får en instruktion om att tänka steg-för-steg, kan ge tillfredsställande svar på en given fråga. Vidare ges ett perspektiv där *zero-shot*-modellen ställs mot en *in-context*-modell, likt tidigare studier kan det ses att om endast dessa jämförs bedöms den förstnämnda som underpresterande, trots att den i andra kontexter kan ses som fullt tillfredsställande. Även Matarazzo och Torlone [22] lyfter *in-context*-promptning som ett sätt att nyttja LLMs på ett effektivt sätt.

Esposito [9] listar ett antal scenarier där *zero-shot*-modeller lämpar sig, såsom textkomplettering, klassificering av en text och tolkning av den generella tonen i en text. Noterbart dock, är att *in-context*-promptning även här anges som den föredragna metoden, särskilt i fall där uppgiften är av mer komplex natur eller fall där det går att ange ett mönster som hjälper modellen att bättre förstå uppgiften. White m.fl. [36] ger mer stöd för att tydliga mönster och formatering kan underlätta för modellen - både vid tolkning av input och strukturering av output - och anger att detta tillvägagångssätt är att föredra.

Gao m.fl. [11] framhäver, likt tidigare studier, att promptning kan användas för att optimera en språkmodell men betonar även att det i många fall är svårt att prompta en modell på rätt sätt - det vill säga på sådant sätt att modellen blir så effektiv som den har faktisk potential att bli. Ett av de främsta hindren som nämns, är otillräcklig förståelse för hur modellens inre struktur är uppbyggd och fungerar. Författarna menar att detta leder till att manuella prompter troligtvis är suboptimala och föreslår en metod som innebär att automatiskt söka efter och generera prompter.

Även om detta arbete inte hanterar texter som kan klassas som avancerade i någon faktisk mening, ger de tidigare studierna en indikation att LLMs även är fullt kapabla att hantera text av mindre komplex karaktär. Vidare finns en svårighet i att fastställa exakt vilka typer av svar som kan inkomma via SMS, varför möjligheten att styra språkmodellen i en riktning där denna behöver förtydliganden, eller bekräftelse, bör kunna bidra till ett mer tillförlitlig system som helhet. Olika promptningsmetoder och begränsande faktorer som nämnts i tidigare studier, ses som värdefulla även för denna studie då tidigare insikter bör kunna bidra till bättre utvärdering av resultat samt ge underlag för en meningsfull tolkning av dessa. Slutligen kan det ses att tidigare studier ger belägg för ett mer mönsterstyrt sätt att strukturera input och förväntad output, något som även förespråkas av Opper [28], vilket är det verktyg som i denna studie nyttjas för LLM-hantering.

3.2 SMS-hantering utan LLM

SMS är en väletablerad teknik som varit en del av vardagen betydligt längre än LLMs, som inte varit kommersiellt aktuella i mer än ett antal år [24, s.2, fig. 2]. Redan innan den massiva AI-utvecklingen fanns det dock ett intresse av att hantera

och tolka SMS på ett effektivt sätt. Ek m.fl [32] genomförde år 2011 en studie där en annan NLP-metod, *Named Entity Recognition* (NER), tillämpas för att extrahera detaljer från SMS. I studien beskrivs en metod där Regexmönster nyttjas för att plocka ut specifika, fördefinierade entiteter från meddelanden. Entiteterna kan exempelvis vara ett klockslag, ett telefonnummer eller en plats, och i de fall där de eftersöka entiteterna är numeriska anses metoder baserade på Regexmönster mycket effektiva. För att hantera textsträngar krävs dock en mer gedigen metod där varje sträng matchas mot strängar i fördefinierade listor, som representerar givna entiteter. Det finns dock ett antal begränsningar med de metoder som beskrivs i studien, där brist på kontext anges som en trolig orsak till flera av dem. Exempelvis kan ett ord förekomma i flera listor och utan kontext är det svårt att avgöra vilken lista, och därmed entitet, som bör associeras med givet ord.

Redan innan studien ovan publicerades, genomförde Cooper m.fl. [29] en liknande sådan där även de nyttjade en mönsterbaserad metod. Även om metoden som sådan bör kunna ses som något inaktuell, givet språkmodellers framfart de senaste åren, finns det andra relevanta lärdomar att ta med sig. Författarna påpekar, likt ovan nämnda studie, att bristen på kontext försvårar den faktiska uppgiften – att extrahera önskad information. Vidare nämner de meddelandenas natur och utformning som en potentiellt försvårande faktor, då SMS ofta följer en syntax som i andra fall anses inkorrekt. Dessutom är SMS och dess språk något som kontinuerligt förändras och uppdateras, vilket gör regel- och mönsterstyrda metoder svåra att applicera för en mer långsiktig lösning.

De studier som berörts ovan, nyttjar metoder som i sin helhet kan uppfattas som onödigt arbetsamma givet de metoder och verktyg som finns tillgängliga idag. En röd tråd som kan ses dock, och som även kunnat ses i studier kring de metoder som är mer aktuella idag, är en återkommande betoning på vikten av en relevant kontext för att optimera texthanteringen och säkerställa att önskad information plockas ut på ett korrekt sätt givet sitt sammanhang. Dessutom bör det noteras att en till synes basal metod som matchning av Regexmönster, faktiskt är en effektiv sådan i specifika sammanhang, vilket gör att en sådan metod troligen kan fungera som ett komplement till en språkmodell. För att ge vidare belägg för en sådan hypotes och utreda hur de på bästa sätt kan komplettera varandra, krävs dock mer omfattande litteraturstudier – något som faller utanför den här studiens ram.

3.3 SMS-hantering med LLM

Som nämnts inledningsvis, tycks hantering av SMS med hjälp av en LLM i ett orderflödessammanhang, vara relativt outrett. Däremot finns en noterbar andel studier som behandlar SMS, LLM och hantering av spammeddelanden. Även om applikeringsområdena skiljer sig åt, finns det vissa likheter i och med hanteringen av meddelanden med hjälp av en språkmodell, inom en specifik kategori. Denna koppling gör att det bedöms finnas resultat och insikter som är relevanta för den aktuella studien.

Hur m.fl. [15] publicerade nyligen en studie med syftet att utreda hur effektiv en LLM är vad gäller upptäckandet av spammeddelanden. Författarna genomförde fler-

talet försök – med tre olika språkmodeller – där inga, ett eller flertalet exempel gavs via prompten. De fann att bäst resultat uppnåddes i de försök där modellen getts flera exempel via prompten, tillsammans med en uppmaning om att kategorisera meddelandet som spam eller icke-spam. I försök där modellen istället ombads sätta en siffra för att kategorisera meddelandet gavs sämre resultat. Värt att notera är att försök gjordes i både en engelsk och koreansk kontext, där resultaten – om än goda i båda fall – skiljde sig åt beroende på språk. Vidare noterade författarna en skillnad i beteende mellan de olika modellerna, där en av dem vid upprepade tillfällen genererade responser innehållande kinesiska tecken, då den initiala prompten var koreansk. Författarna påpekar även att modellerna inte alltid förstod de givna instruktionerna, vilket resulterade i oväntade eller oönskade svarsmönster. Slutligen betonas det faktum att antalet *in-context*-prompter var få, och att detta kan ha påverkat språkmodellernas prestationer.

Chang och Aïmeur [6] har studerat hur LLMs kan nyttjas för att upptäcka bluffmeddelanden på olika plattformar. Särskilt intressant med studien och genomförandet, är att de inkluderat GPT-4o-mini, vilket är den modell som nyttjas även i denna studie. Modellernas prestationer bedömdes utifrån ett antal kriterier och jämfördes därefter med varandra. För den egna studien är dock jämförelserna i sig, och de resultat och slutsatser som bygger på jämförelserna, något irrelevanta då den modell som nyttjas i det här arbetet är ett direkt krav från kunden och därmed en oföränderlig faktor. Av den anledningen återges inte jämförande resultat, utan istället sammanfattas erfarenheter kring GPT-4o-mini utan att sättas i relation till andra modeller. De kriterier som användes för utvärdering, definierades enligt följande:

- ***accuracy***: andelen meddelanden (både bluff och icke-bluff) som klassificerades korrekt av modellen, i förhållande till det totala antalet meddelanden.
- ***precision***: andelen av de meddelanden som modellen flaggade som bluff som faktiskt var bluff (alltså, hur träffsäker modellen är när den säger "bluff").
- ***recall***: andelen av alla faktiska bluffmeddelanden som modellen lyckades identifiera som bluff (alltså, hur många av bluffarna som flaggades korrekt).
- ***F1-score***: det harmoniska medelvärdet av *precision* och *recall*, ett mått som balanserar modellens träffsäkerhet mot dess förmåga att hitta alla bluffar.

Enligt ovanstående definition och ordning, noterades följande värden för GPT-4o-mini: 0.892, 0.962, 0.833 samt 0.893.

Salman m.fl [30] har publicerat en studie kring upptäckandet av spammeddelandet med hjälp av språkmodeller. I studien jämförs en rad olika modeller av varierande storlek, där de konstaterar att mindre LLMs inte är robusta nog när det kommer till spamdetektering samt att de är mer promptkänsliga. Med en modell av storlek 13B (jämför med GPT-4o-mini, cirka 8B [33]) kunde de se hur noggrannheten varierade mellan omkring 0.45-0.85, beroende på prompt. Återigen bör det påpekas att den modell som nyttjas i denna studie är en del av kraven från kund och därmed utanför studiens kontroll. Trots detta bör insikter om dess roll och potentiella påverkan nämnas för att ge underlag för en eventuell vidareutveckling, samt ge stöd för analys och utvärdering av implementation med aktuell modell. Författarna beskriver vidare

hur de undersökt vilken påverkan meddelandenas längd har, där tre klassindelningar baserade på antal tokens görs – kort (< 20), medium ($20 - 30$) samt lång (> 30). De fann att långa meddelanden försämrade prestationen hos samtliga modeller av mindre storlek.

Ovan nämnda studier ger mer stöd till de generella resultat som presenteras tidigare – fler exempel i prompten bidrar generellt sett till en bättre presterande LLM. Utöver att styrka denna tes, ger studierna som behandlats i detta delkapitel viktiga insikter som antingen kan komma att påverka implementationen av prototypen, eller ge bättre förståelse för och tolkning av resultat. Exempelvis har det kunnat ses att språk kan påverka resultatet i olika riktningar. Exakt hur denna studies primära språk (svenska för in- och utgående meddelanden) kan komma att påverka språkmodellen är svårt att avgöra på förhand, dock ger tidigare studier en indikation att just språket kan vara en påverkande faktor. En annan intressant insikt, är att meddelandenas längd kan påverka hur väl språkmodellen presterar. Utan vidare belägg är det tänkbart – i givet orderflödesscenario – att en begränsad meddelandelängd kan ge följdeffekter som minskad försäljning på grund av begränsat utrymme för kunden att uttrycka sig fritt. Ett försök att hantera en sådan typ av begränsning skulle troligen innebära att för långa meddelanden aldrig bearbetas medan kunden istället uppmanas vara mer kortfattad. Detta i sin tur skulle potentiellt kunna göra kunden mindre motiverad att fortsätta sin beställning, alternativt leda till en frustration som på annat sätt påverkar säljande företag negativt. Att förstå dessa faktorer och deras eventuella påverkan ligger dock utanför den här studiens ramar, som inte har för avsikt att utreda ekonomiska eller marknadsförande aspekter.

Kapitel 4

Metod

Genomförandet indelades i flera steg där varje steg syftade till att addera till den grund som i ett senare skede används för att ge svar på de två forskningsfrågor som arbetet primärt kretsar kring. Inledningsvis gjordes en litteraturstudie, med huvudmål att samla in det övergripande bakgrundsmaterial som ansågs nödvändigt för att genomföra arbetet med stöd av tidigare erfarenheter inom berörda områden. Utöver detta tillkom Design Science-metoder för att redogöra för de faktiska forskningsfrågorna, vilka delvis grundar sig på de insikter som litteraturstudien bidragit med.

4.1 Litteraturstudie

I det inledande steget genomfördes en litteraturstudie med sökningar i Google scholar [13] och Summon [31], ett verktyg som finns tillgängligt via Blekinge Tekniska Högskola (BTH). Summon är en bibliotekstjänst där sökningar görs i alla de databaser som BTH har tillgång till, med möjlighet till mer specifika sökningar enligt givna mönster. Även Google Scholar ger möjlighet till sökningar baserade på definierade regler, i flera databaser samtidigt. För att ytterligare underlätta värdering och filtrering av sökresultat och publikationer, rankas dessa utifrån exempelvis var och av vem de publicerats, när de skrevs samt hur ofta och när de citerats av andra. Dessa faktorer, i kombination med viss hänsyn till peer reviews, har beaktats vid bedömning av litteraturens relevans.

För att få en bredare helhetsbild och förståelse för de ingående delarna av arbetet gjordes en litteratursökning med ovan beskrivna verktyg, med fokus på de nyckelord som kan associeras med arbetets huvudområden. Detta innebär att stor vikt lagts vid Large Language Models (LLMs) som helhet men även mer specifikt LLMs i relation till analys och tolkning av text. Dessutom gjordes sökningar med fokus på hur LLMs ytterligare kan optimeras och om sådana åtgärder anses nödvändiga vid användning i denna typ av kontext. För att få bättre förståelse för LLM och Natural Language Processing (NLP) som skilda begrepp och kunna styrka valet av LLM i detta fall, gjordes även sökningar efter litteratur som behandlar skillnader, likheter och användningsområden för både LLMs och traditionella NLPs. Då RQ1 innefattar SMS-API och hantering av meddelanden med hjälp av en språkmodell, ansågs litteratur kring dessa delar relevanta, varför även sökord relaterade till dessa inkluderades. Till sist gjordes en mer riktad sökning för att hitta information och litteratur kopplad till OpenAI [26] och Opper [28], som kan användas för att integrera LLM-lösningar i en applikation eller tjänst.

En lista över de söktermer som initialt användes kan ses nedan. Notera att för varje term gjordes sökningar på både svenska och engelska, för att på så sätt minimera risken att underlag begränsas eller förbises av språkskäl. Dessutom tillämpades konceptet *snowballing* [37] för att hitta ytterligare relevanta källor, utifrån litteratur som vid en första gallring ansågs kunna bidra till arbetet. Snowballing delas in i *backward* och *forward snowballing*: där det förstnämnda innebär att utgå ifrån en given källas referenslista och därifrån hitta relaterad litteratur, medan det sistnämnda innebär att söka efter litteratur där en given källa blivit citerad. I det senare fallet förfinades sökningarna något genom att kombinera flera av de aktuella söktermerna.

- LLM AND ”text analysis”
- LLM AND ”text interpretation”
- LLM AND (”traditional NLP” OR ”conventional NLP”)
- LLM AND (”few-shot” OR ”zero-shot”)
- ”few-shot learners” OR ”zero-shot learners”
- ”prompt engineering”
- ”SMS API” OR ”SMS-API”
- (”SMS API” OR ”SMS-API”) AND (LLM OR ”language model” OR NLP)
- OpenAI OR Opper
- design science

4.2 Skapa en artefakt

För att utreda den första forskningsfrågan (RQ1) som formulerats i kapitel 2, antogs en Design Science-approach. Design Science innefattar en uppsättning metoder som bygger på skapandet och utvärderandet av en artefakt i ett försök att bättre förstå, eller finna en lösning på, ett praktiskt problem. Artefakten kan exempelvis vara en prototyp, en ritning eller en algoritm – alla med syftet att lösa ett upplevt problem [17, s. 3]. För att angripa ett problem med en Design Science-approach kan det ramverk som Johannesson och Perjons beskriver nyttjas [17, ss. 77-82]. Ramverket bygger på fem aktiviteter, där det rekommenderas att en eller två av dessa ges primärt fokus för att utreda aktuellt problem, medan resterande ges mindre utrymme. En krav- och utvecklingsbaserad metod antogs för att utreda RQ1, där krav (se avsnitt 4.2.1) definierades utifrån den initiala frågeställningen medan design och utveckling av en prototyp (se avsnitt 4.2.2 - 4.2.4) baserades på dessa krav.

Centralt för Design Science-metoder, särskilt i de fall där skapandet av en prototyp för demonstration ingår, är det iterativa arbetet vilket innebär att alla eller delar av processens steg upprepas i flera omgångar snarare än att ske sekventiellt från början till slut [17, s. 79, fig. 4.1]. Det iterativa arbetet genomfördes enligt de steg som listas nedan, där särskilt steg 2 och 3 upprepades.

- Definiera krav och behov utifrån initial frågeställning.
- Designa och implementera en prototyp.
- Testa prototypen för att säkerställa funktionalitet.
- Utvärdera prototypen i relation till förväntat beteende och resultat.

Oates m.fl. [25] betonar vikten av att utvärdera den skapade artefakten utifrån ett eller flera kriterier, för att på så sätt kunna dra slutsatser och föreslå eventuella ändringar. Vilka kriterier som är aktuella varierar beroende på typ av artefakt och bör relateras till den initiala frågeställningen. Ett vanligt sätt att angripa detta, oavsett aktuella kriterier, är att fokusera på antingen *proof of concept* eller *proof by demonstration* - det vill säga att säkerställa att den framtagna prototypen har förväntad karaktäristik eller beteende vid utvärdering i en begränsad skala [25, ss. 121-122].

Det är värt att notera, att i den iterativa process som listats ovan, avser testningen i det tredje steget primärt isolerade tester genomförda som en del av utvecklingen. Först i ett slutligt skede nyttjades en oberoende testgrupp, bestående av ett antal frivilliga respondenter. För att ge dessa en uppfattning om vad testet innebar - för att på så vis kunna ta ställning till huruvida de var villiga att delta - gavs grundläggande information om prototypen och syftet med testet. Detta innebär att alla respondenter, innan genomförande av test, var medvetna om att ett SMS-API och en språkmodell tillsammans hade implementerats i ett system, för att hantera ett orderflöde. Detaljer om implementationen och hur flödet förväntades fungera utelämnades med en uppmaning om att, vid intresse, be om sådana detaljer först efter genomfört test. Slutligen blev alla deltagare informerade om att deras konversationer kunde komma att nyttjas för exemplifiering i resultat eller efterföljande diskussion.

För att genomföra testet uppmanades deltagarna registrera sig som bagerikund via ett SMS. Därefter uppmanades testpersonerna invänta ett erbjudande och besvara meddelandet på det sätt de ansåg passande utifrån dess innehåll. Dessutom ombads de i slutet att besvara en anonym enkät för att utvärdera hur de upplevde kommunikationen och orderflödet.

4.2.1 Krav

I samråd med en representant från samarbetsföretaget har de punkter som angavs i avsnitt 1.1 brutits ned ytterligare, vilket resulterat i en mer specifik lista innehållande krav och behov som systemet bör tillgodose. Listan innefattar riktlinjer kring funktionalitet och verktyg samt i viss mån utformning. För att enklare kunna överblicka kraven, delas listan in enligt tidigare nämnda kategorier. Eventuella begränsningar, förtydliganden eller tillägg markeras med *.

- **Funktionalitet**[1]
  - Möjlighet för kund att registrera sig, för att motta erbjudanden.
  - Möjlighet för kund att avregistrera sig, för att sluta motta erbjudanden.

[1]Där *kund* nämns avses hantering via SMS, för *säljare* avses webbklient.

        - kan endast ske om kunden *inte* har någon aktiv order.
    - Möjlighet för säljare att skicka ett erbjudande till alla registrerade kunder.
        - alla kunder hämtas automatiskt från en kundbas, för enkel hantering.
    - Möjlighet för kund att besvara ett erbjudanden och lägga en beställning.
        - kunden ska kunna beställa produkter som anges i erbjudandet.
        - kunden ska även kunna beställa en produkt som inte är en del av erbjudandet, givet att beställd produkt är att klassa som bageri- eller konditorivara. I dessa fall ska kunden informeras om att produkten är noterad men inte kan garanteras, då den inte ingår i erbjudandet.
        - beställningar innehållande produkter som inte är relaterade till verksamheten ska tillbakavisas med ett förtydligande om vad som erbjuds.
        - ovanligt stora beställningar ska kontrolleras med en motfråga innan de adderas till kundens orderlista.
        - otydliga beställningar, där exempelvis flera kvantiteter anges för en och samma produkt, ska även de kontrolleras.
        - om kunden uttrycker missnöje eller på annat sätt påkallar uppmärksamhet, ska denne erbjudas att bli uppringd.
    - En kund som besvarar ett erbjudanden genom att göra en beställning, ska erbjudas möjlighet att addera fler produkter till sin beställning.
    - En kund ska enkelt kunna överblicka en pågående beställning.
        - kunden ska kunna se vilka produkter som beställts och i vilket antal.
        - kunden ska kunna ta bort ett givet antal av en viss produkt.
        - kunden ska kunna avbryta en pågående beställning, vilket innebär att alla beställda produkter tas bort och ordern avslutas.
    - En beställning ska enbart kunna bekräftas av kunden själv och bör ses som oavslutad om inte annat anges.
    - Bekräftade beställningar ska kunna lagras temporärt, för vidare hantering.

- **Verktyg**[2]
    - 46elks API ska nyttjas för SMS-hantering.
    - Opper ska nyttjas för LLM-hantering.
        - modellen GPT-4o-mini ska användas.
    - CSV-filer ska nyttjas för lagring av data inom systemet.
        - företaget vill i detta skede ge språkmodellen primärt fokus, för att i ett senare skede hantera mer verkliga lagringsscenarion. En faktisk databas, och genomtänkta design- och implementationsval kring denna, bör därför ges fokus först efter att implementationen av språkmodellen anses tillräckligt robust.

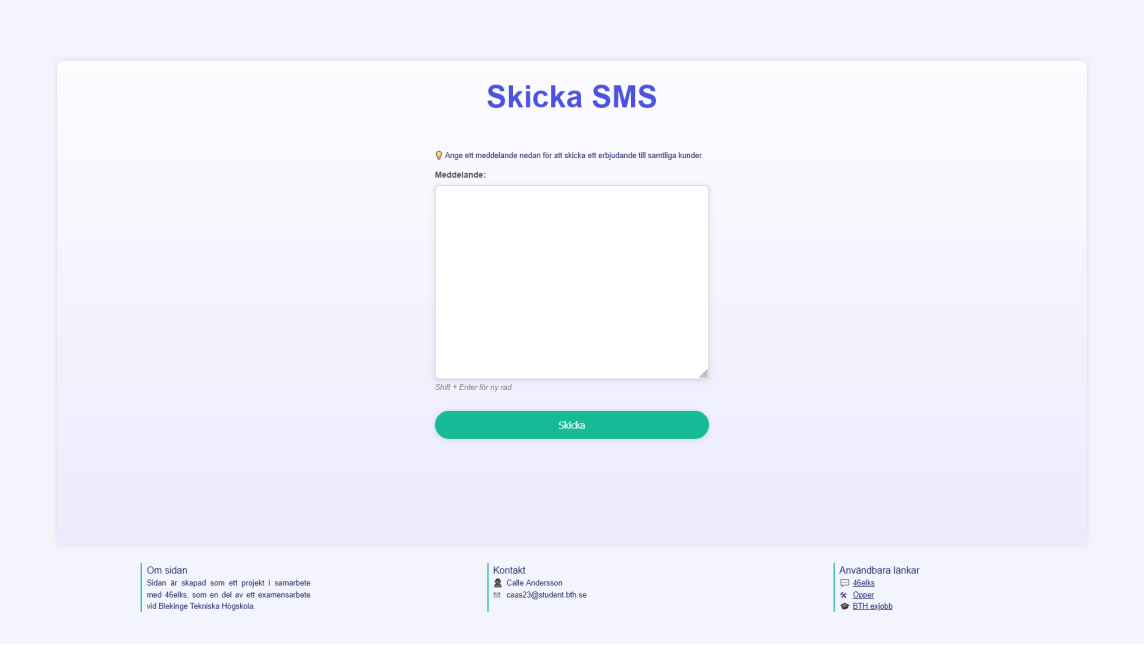
[2]Angivna verktyg är strikta krav från samarbetsföretaget.

- **Utformning**
  - Webbklienten bör utformas så att en säljare på ett intuitivt och lättåtkomligt sätt kan skicka erbjudanden till sina kunder.
    - klienten bör vara minimalistisk och det bör vara tydligt för säljaren hur denna ska gå till väga för att skicka ett erbjudande.
    - klienten bör innehålla ett lager av säkerhet och kontroll, för att förhindra att erbjudanden skickas ut av misstag.
    - när ett erbjudande skickats bör det vara tydligt om sändningen lyckats eller inte.
    - klientens färgskala kan med fördel inspireras av 46elks egna webbsida.
  - Orderflödet bör vara lättsamt i sitt språk, gärna med humoristiska inslag.
    - kommunikationen ska upplevas upplyftande, tilltalande och enkel.

4.2.2 Webbklient och server

Webbklienten är en *Single Page Application* (SPA), utvecklad med JavaScript och React. För att tillgodose de krav som listats i avsnitt 4.2.1 hanteras alla statiska och fördefinierade delar, däribland avsändare och mottagare, på serversidan. Detta innebär att webbklienten enbart behöver ge möjlighet att formulera och skicka ett specifikt erbjudande.

Klientens design, se fig. 4.1 nedan, är enkel med ett formulär i form av en textruta centrerad på sidan, för att ge meddelandefunktionen tydligt fokus. Innan ett meddelande kan skickas måste en popup-fråga besvaras, för att säkerställa att meddelandet verkligen ska skickas. Efter att erbjudandet skickats ges direkt feedback via en ny popup, där det framgår om sändningen lyckats helt, delvis eller inte alls.



Figur 4.1: Webbklient för att skicka erbjudanden via SMS-API.

För att förmedla en känsla som påminner om samarbetsföretagets primära webbsida, har gröna och blålila toner använts. Slutligen har enkla men beskrivande symboler adderats för att ge klienten mer liv, samtidigt som minimalismen upprätthålls.

Serversidan är utvecklad med JavaScript och Express, för hantering av kommunikation via SMS-API, LLM-förfrågningar och csv-filer. csv-filerna nyttas för temporär lagring av kunder, pågående beställningar och aktuella erbjudanden.

4.2.3 SMS-API

API:t finns implementerat i två separata funktioner, en för att skicka det initiala erbjudandet som formulerats i webbklienten, och en för att hantera kommunikation under pågående beställning. För att skicka ett erbjudandet hämtas alla registrerade kunder och därefter anropas `sendSMSOffer` för varje kund. I samband med att erbjudandet skickas ut sparas detta i en separat csv-fil, för att kunna nyttjas i senare steg. För varje kund som mottar ett erbjudande, lagras 'sendSMSOffer' som funktionsnamn i den csv-fil som hanterar kunderna. På så sätt kan inkommande SMS från en kund kopplas till ett aktivt erbjudande.

`sendSMSResponse` anropas varje gång ett nytt meddelande ska skickas till en given kund - såvida meddelandet inte är det initiala erbjudandet. Denna funktion är mer flexibel än föregående och kan nyttjas för all typ av kommunikation med en given kund. Varje gång ett nytt utskick görs, uppdateras kundens information i csv-filen, för att säkerställa att inkommande svarsmeddelande hanteras korrekt.

4.2.4 LLM

Språkmodellen GPT-4o-mini finns implementerad i fyra separata funktioner, för att hantera olika typer av LLM-anrop. `extractOrderDetails` och `extractMoreDetails` används för att tolka nyinkomna och utökade beställningar. `checkOrder` används för att hantera SMS som inkommit som svar på ett begärt förtydligande medan `callCustomer` är en separat funktion där modellens enda uppgift är att tolka om en kund vill bli uppringd eller inte.

Samtliga funktioner har en liknande struktur med en övergripande uppgiftsbeskrivning, definierade in- och outputscheman samt ett antal inputparametrar. Den generella strukturen, baserad på Oppers rekommendationer [27], ses i kodexempel 4.1.

```
async function general_structure(parameter1, parameter2) {
  const { json_payload } = await client.call({
    name: "example",
    instructions: "Here, an instruction to guide the model is added.",
    input_schema: InputSchema, // defined elsewhere
    output_schema: OutputSchema, // defined elsewhere
    model: "openai/gpt-4o-mini",
    input: {
      firstParam: parameter1,
      secondParam: parameter2,
    },
  });
  return json_payload;
};
```

Kodexempel 4.1: Generell funktion för LLM-anrop.

Inputschemat ger ett strukturerat sätt att beskriva de inkommande parametrarna. För varje parameter anges dess typ och ett textstycke, likt den övergripande uppgiftsbeskrivningen, med information om vad parametern innehåller och hur den kan eller bör användas av modellen. I kodexempel 4.2 ses det schema som definierats för `extractOrderDetails`. Notera att exemplet förenklats, i den mån möjligt, för att öka läsbarheten.

```
const InputSchema = {
  type: "object",
  properties: {
    outgoingOffer: {
      type: "string",
      description: "A Swedish text containing information about an
        offer from a bakery. The text contains specific items offered.",
    },
    incomingMessage: {
      type: "string",
      description: "A Swedish text containing information about an order,
        in response to outgoingOffer.

        { Continued with more details and instructions }",
    },
  },
  required: ["outgoingOffer", "incomingMessage"],
};
--------------------------------------------------------------------------
```

Kodexempel 4.2: Komprimerat inputschema för extractOrderDetails.

Outputschemat ger på ett liknande sätt en möjlighet att hantera och styra den respons som fås från språkmodellen. Genom att definiera ett tydligt schema med förväntat innehåll, är det möjligt att plocka ut givna delar från ett SMS, samt få färdiga meddelanden att nyttja för att besvara detta. I kodexempel 4.3 används återigen `extractOrderDetails`, i komprimerad form, som utgångspunkt.

```
const OutputSchema = {
  type: "object",
  properties: {
    thoughts: {
      // always included, as recommended by Opper
      type: "string",
      description: "The thoughts of the model while extracting
        the order details.",
    },
    placed_order: {
      type: "boolean",
      description: "Whether the customer ordered something.",
    },
    order_declined: {
      type: "boolean",
      description: "If the customer simply declines the offer,
        in any reasonable way, set to true.",
    },
```

```
check: {
  type: "string",
  description: "Any order that contain an invalid or unreasonable
    amount of an order item, should be added here.

    If a check is needed, add a message.

    Examples:
      Input: 'Jag vill har tre, fem plommon, tack!'
      Output: 'Vill du beställa tre eller fem plommon?'

    { Continued with more details and instructions }",
},
count: {
  type: "object",
  description: "Any order item that was mentioned in the offer,
    should be added as key and order count for that item as value.

    Example:
      The offer mentions buns and cookies.
      The customer orders 2 buns, 3 cookies
      and one strawberry cake.

      Result: Add the buns and cookies.

    { Continued with more details and instructions }",
},
extra: {
  type: "object",
  description: "Any extra orders related to baking, pastries or
    sweets, that was not mentioned in the initial offer.

    Example:
      The offer mentions buns and cookies.
      The customer orders 2 buns, 3 cookies
      and one strawberry cake.

      Result: Add the strawberry cake.

    { Continued with more details and instructions }",
},
unreasonable: {
  type: "string",
  description: "If incomingMessage include a part that is not related
    to baking, pastries or sweets, add a message.
    The message should be formulated in such way that it
    can be used as a response to the customer.

    If the order contains parts that is clearly
    unrelated to baking or simply off topic,
    return a message starting with 'Tyvärr...'
    followed by an explanation that what they ordered is not possible.

    { Continued with more details and instructions }",
},
```

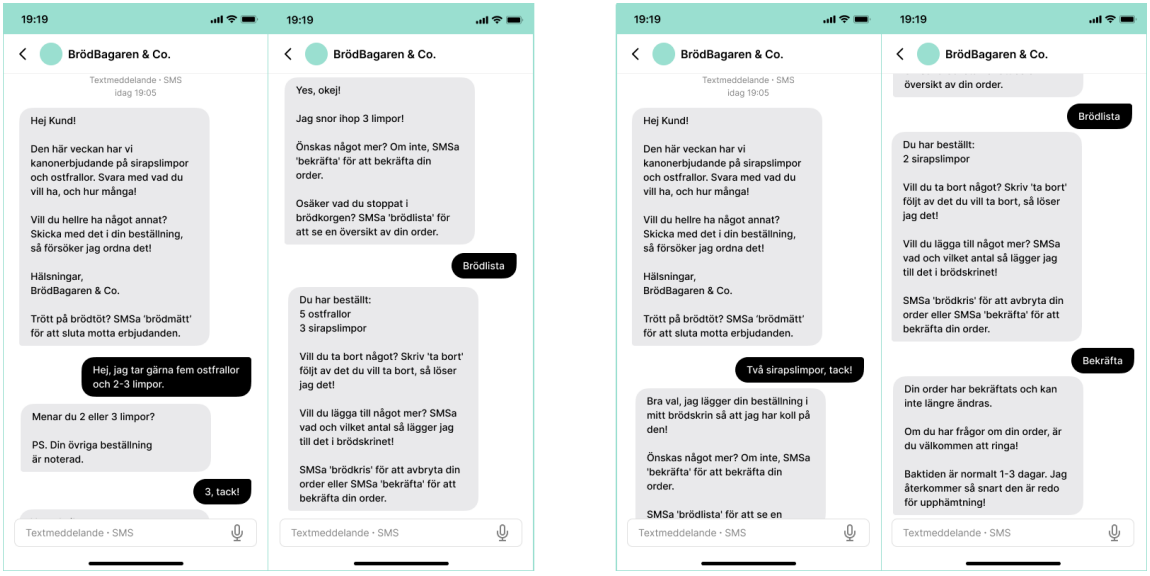
```
    attention: {
      type: "string",
      description: "If the incomingMessage has a tone that implies that
        the customer wants extra attention, set a message.
        The message should be formulated in such way that it
        can be used as a response to the customer.

        Examples:
        - the customer behave as if they are not understood properly
        - customer does not order anything but implies something is wrong

        { Continued with more details and instructions }",
    },
  },
  required: ["thoughts", "placed_order", "check",
             "count", "extra", "unreasonable", "attention"],
};
--------------------------------------------------------------------------
```

Kodexempel 4.3: Komprimerat outputschema för extractOrderDetails.

I exemplet ovan kan det ses att flertalet kategorier definierats för att hantera en beställning. Genom att dela upp inkommande meddelande i flera kategorier, kan en hierarkisk struktur nyttjas för att avgöra hur efterföljande svar bör formuleras. Detta innebär att tydliga och korrekta delar av en beställning kan noteras och bekräftas samtidigt som eventuella motfrågor, eller behov av förtydliganden, kan ges primärt fokus i det svar som skickas till kund. I fig. 4.2 ses till vänster ett exempel på hur en beställning med både tydliga och otydliga svar hanteras. I samma figur ses till höger ett scenario där en beställning görs utan några tvetydigheter. I bilaga A återfinns ett längre transkriberat orderflöde för ytterligare demonstration av implementationen.



Figur 4.2: Exempel på hantering av tydlig respektive otydlig beställning.

4.3 Utökad litteraturstudie

För att utreda den andra forskningsfrågan (RQ2), antogs återigen en Design Science-approach. I detta fall användes en kravbaserad metod, där litteraturstudier är en betydande del [17, s. 82]. Denna metod är något mer abstrakt än den som beskrivs i avsnitt 4.2, på så sätt att denna inte involverar implementering eller demonstration av en faktisk prototyp. Metoden fokuserar istället på att definiera krav och utifrån dessa skapa en ritning eller skiss som, i ett senare skede, kan nyttjas för att genomföra en faktisk implementation.

Litteraturstudien genomfördes enligt det tillvägagångssätt som beskrivs i avsnitt 4.1. I denna utökade sökning användes dock andra söktermer, för att lägga ett tydligare fokus på aktuellt område. För att identifiera relevanta söktermer att utgå ifrån formulerades ett antal underfrågor till RQ2 - dessa kan ses listade i kapitel 2. De faktiska söktermerna som använts kan ses nedan.

- ”data visualization” AND ”web development”
- ”data visualization” AND ”user experience”
- ”data visualization” AND ”table” AND ”chart”
- ”data presentation” AND ”table” AND ”chart”
- ”presentation” AND (”small data” OR ”small datasets”)

Kapitel 5

Resultat och analys

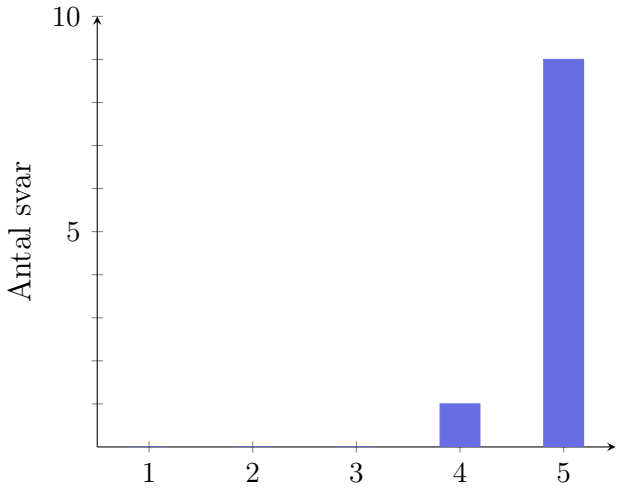
Följande kapitel indelas i två övergripande avsnitt, ett för vardera forskningsfråga. Avsnitt 5.1 behandlar resultat för den artefakt som skapats för att utreda RQ1, se avsnitt 4.2 för detaljer kring implementationen. Avsnitt 5.2 behandlar resultat från den litteraturstudie som genomförts för att utreda RQ2.

5.1 Artefakt

Det system (BrödBagaren & CO.) som implementerats innefattar en webbklient med tillhörande server, 46elks SMS-API samt en språkmodell via Opper. Nedan presenteras de enkätsvar som inkommit från de deltagare som testat systemet i ett orderscenario, tillsammans med exempel från interaktion med modellen.

Enkäten omfattar sex huvudfrågor, där de fem första besvaras genom att först sätta en siffra mellan ett och fem och sedan motivera valet av denna siffra. Den sista frågan kräver enbart ett textbaserat svar. Totalt ingick tio personer i testgruppen, vilket resulterat i motsvarande antal svarande på enkäten.

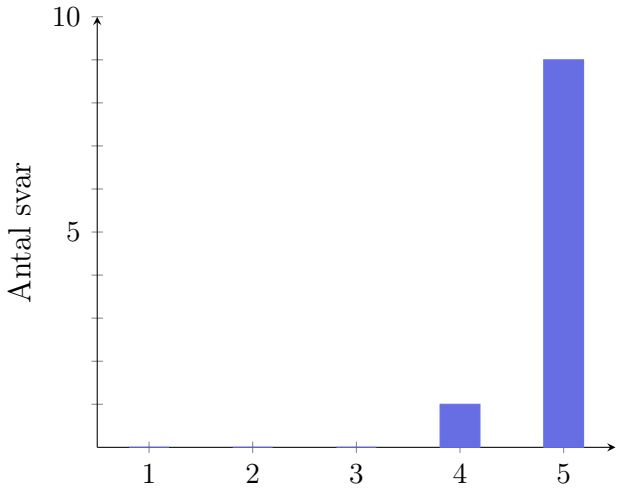
Q1: Hur skulle du beskriva helhetsupplevelsen av den kommunikation du haft med BrödBagaren & Co.? (Tänk in aspekter som: användarvänlighet, bemötande & språklig ton, svarstider, etc.)
(1) Mycket dålig, (5) Mycket bra



Motivera valet av siffra för **fråga Q1** ovan.

Kommunikationen flöt på bra, snabba svar. Språkbruket var rikt och trevligt.
Tydlig och snabb kommunikation
Tyckte kommunikationen fungerade bra, det var tydligt och lättförståeligt att följa instruktionerna som gavs.bemötandet var trevligt och gillade att den svarade med lite humoristisk ton i det hela. Svarstiderna var snabba och finns inte riktigt något att klaga på. Tyckte också att det var bra att SMSen skrev en tydlig mening med att det gick att sluta morra erbjudanden genom att smsa ett specifikt ord.
Det var enkelt, tydligt och smidigt
Enkelt och intuitivt. Bra att det funkar med fritext
Lätt att förstå, snabba svar, trevlig ton och bra bemötande.
Mycket tydlig förklaring för hur man beställer, snabbt svar och bra bemötande.
Väldig snabb svarstid!
Enkelt och snabbt
Snabba svar och trevligt bemötande.

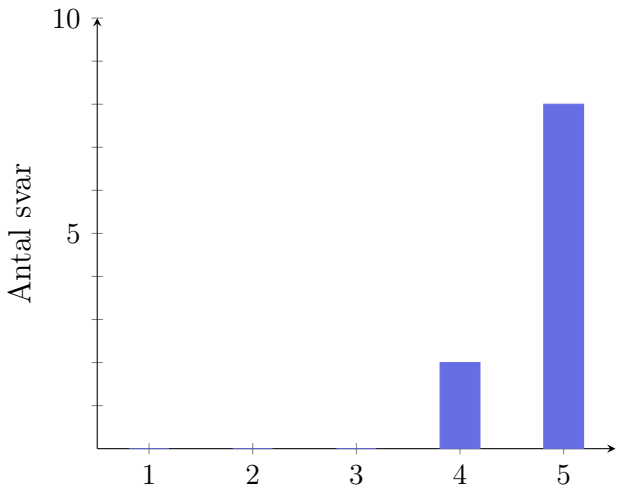
Q2: Hur enkelt, eller svårt, var det att förstå den kommunikation (instruktioner och alternativ) som kom via SMSen?
(1) Mycket svårt, (5) Mycket enkelt



Motivera valet av siffra för **fråga Q2** ovan.

Inga konstigheter enligt mig, språket upplevdes tydligt och klart.
Instruktionerna var tydliga och enkla att förstå
Instruktionerna var tydliga och var lätta att följa. Det som jag tänkte på var att när den ville att jag skulle registrera mig så sa den bara ”skriv namn”, men inte om det var för och efternamn eller bara någon av dem. Annars var kommunikationen bra, och gick att följa och förstå utan att extra frågor skapades från min/kundens sida.
Jag tycker likt förra svaret att det var tydligt och enkelt. Lätt att förstå sig på
Generellt bra. Tydligt vad man beställt och när det levereras.
Det var inga konstigheter utan väldigt lätt att förstå hur man skulle gå tillväga.
Som ovan i fråga Q1 så var det mycket tydligt förklarat. Inga konstigheter.
Var tydligt för mig som köpare vad jag behövde göra i smsen.
Enkelt och smidigt
Tydliga instruktioner.

Q3: Hur väl upplever du att dina svar blev tolkade och förstådda? (Var responsen relevant, informativ och relaterad till de svar du själv gett?)
(1) Mycket dåligt, (5) Mycket bra



Motivera valet av siffra för **fråga Q3** ovan.

Endast en brist i kommunikationen då jag beställde en vara som vanligtvis inte finns på bageri, efter att jag ändrade varan fick jag genast ett tydligare svar.

Mina svar uppfattades korrekt

Jag upplevde att svaren jag fick utifrån det jag skickade blev tolkade och förstådda på ett bra sätt. ”Bagaren” upprepade beställningen som jag la vilket gjorde det tydligt, men var också noga att när jag ville ha något som inte var på extra pris så stod det tydligt att det inte kan garanteras.

Jag tycker att responsen man fick var relevant

Tydliga och snabba svar. Lätt att förstå.

Mina svar tolkades rätt, jag fick bra respons och information därefter.

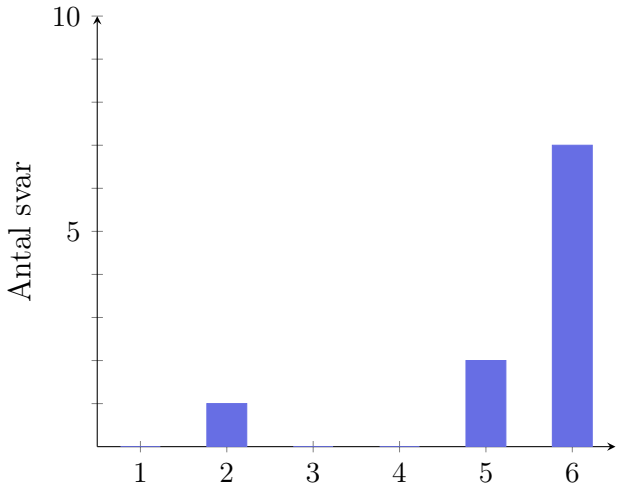
Tycker mina svar blev mycket väl tolkade och förstådda och att mitt övriga önskemål utöver erbjudandena fick bra respons.

De vart tolkade korrekt!

Mycket bra förstådda och tolkade

Bra information.

Q4: Hur väl upplever du att dina egna frågor besvarades? (Om du inte ställde någon fråga, svara 6 (Inga frågor)).
(1) Mycket dåligt, (5) Mycket bra, (6) Inga frågor



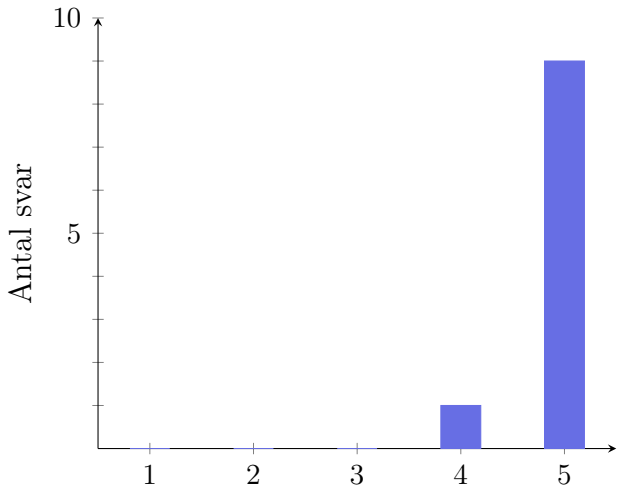
Motivera valet av siffra för **fråga Q4** (Om du svarade "Inga frågor", lämna denna blank.).

Mycket bra.

Bra, min motfråga på extrabeställning utöver det vanliga sortimentet uppfattades korrekt och kontrollerades med en kontrollfråga för att undvika missförstånd

Jag vet inte om det räknas som fråga men bad om att få ta bort två av mina fyra hallongrottor och då togs alla bort.

Q5: Hur logiskt upplevde du orderflödet* (från mottaget erbjudande till bekräftad order)? * Notera att denna fråga skiljer sig från Q1, på så sätt att denna avser hur naturligt du upplevde att det var att ta sig från erbjudande till färdig (eller avböjd) beställning.
(1) Ologiskt/onaturligt, (5) Logiskt/naturligt



Motivera valet av siffra för **fråga Q5** ovan.

Jag tycker det känns trevligt när en service använder ord som färglägger och ”piffar till” språket, då blir det mindre trist och stelt. Exempelvis genom att använda lättare ordvitsar och ”yes” istället för ja, då de övriga meningarna och budskapet upplevdes tydliga så funkar det.

Tydligt, snabbt, lättsamt, väldigt enkelt att lägga och/eller ändra en order.

Tyckte det var ett bra flöde, det började med att de ville att jag skulle registrera mig, efter det kom det ett erbjudande om bakelser vilket följer en tydlig tråd. Det kändes naturligt den ordning som var och det blev inga konstigheter där emellan.

Logiska och bra svar

Tydligt vad som händer i flödet. För mig kändes det logiskt.

Orderflödet kändes logiskt och naturligt.

Orderflödet flöt på smidigt!

Väldigt logiskt och naturligt upplagt.

Väldigt logiskt, då jag beställde 52 bullar och roboten ville försäkra sig om den stora beställningen.

Att det va skrivet enkelt och smidigt.

Q6: Har du några förslag på förbättringar eller ändringar, för att göra BrödBagaren & Co. bättre?

Nej, jag är nöjd.

Inga förslag upplever det som en enkel och intressant tjänst.

Som tidigare nämnt kanske det är bra om det är lite tydligare vid namn, men det kanske inte har så stor betydelse egentligen. Sen en annan sak jag tänkte på var när en order inte blev garanterad så kanske det går att få lite mer information om det. Exempelvis att det kommer skickas ett till SMS eller liknande om den kommer följas med i ordern eller inte under senare tillfälle eller om det kommer som en överraskning sen när ordern ska hämtas. Men om jag skickat en fråga kanske den hade besvarat på den, det dök ju också upp i senare SMS att det bara var att ringa vid frågor.

Nej

Möjlighet att ändra order, t ex lägga till beställning.

Fixa problemet med att avbeställa delar av brödlistan.

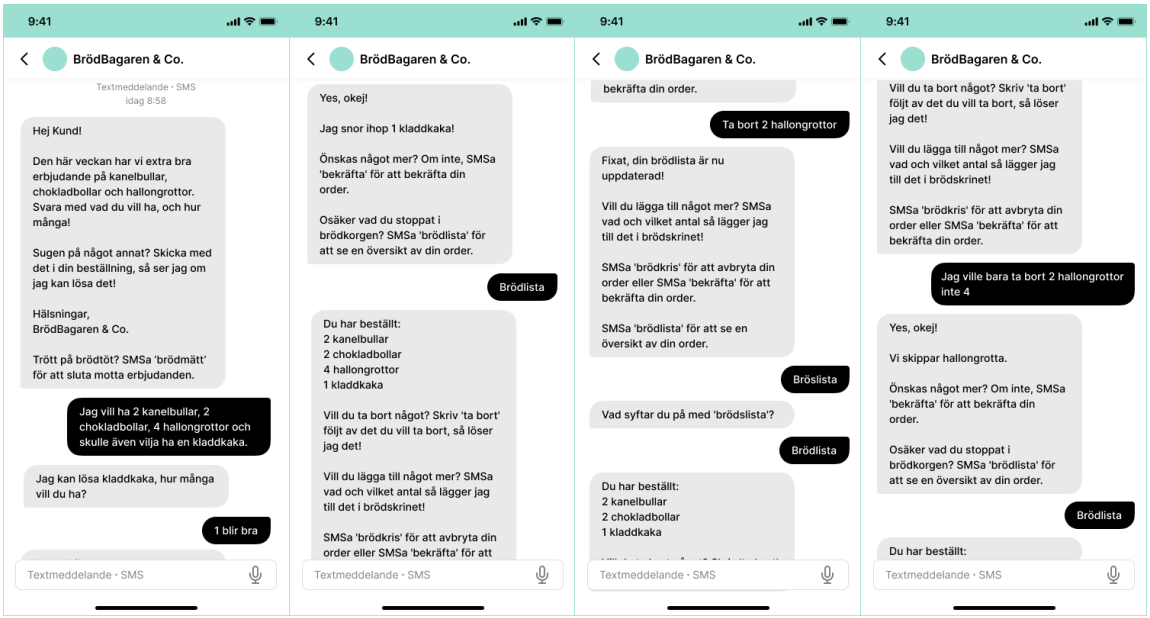
Nej.

Nej!

Nej har inga förslag.

Allt var grymt!

Av enkätundersökningen framgick det, bland annat, att en deltagare upplevde problem vid försök att avbeställa ett givet antal av en specifik vara. I fig. 5.1 återges ett utdrag från den konversation som åsyftas.



Figur 5.1: Utdrag ur konversation där kundens önskemål blev felaktigt tolkade.

5.2 Datapresentation

I samband med att RQ2 definierades, formulerades även ett antal underfrågor i syfte att bryta ned den faktiska forskningsfrågan i mindre beståndsdelar. I avsnitt 5.2.1 utreds dessa frågor utifrån de erfarenheter som arbetet med RQ1 bidragit med, samt utifrån det material som samlats in i samband med litteraturstudien. I avsnitt 5.2.2 ges därefter en sammanfattning av resultaten.

5.2.1 Litteraturgenomgång

Vilken typ av data ska presenteras och hur är denna strukturerad?

Baserat på studiens tidigare steg, är det känt att den data som ska presenteras finns samlad i en csv-fil. Detta innebär att all data finns att tillgå i en given fil, med en given kommaseparerad struktur. Givet det scenario som studien utgår ifrån - ett bageri vill nå ut till kunder med säljerbjudanden - antas det att den insamlade datamängden är att se som liten i ett statistiskt och analytiskt sammanhang. Vidare är datan strukturerad på så sätt att en given kategori kan associeras med en given numerisk mängd, såsom ”25 kanelbullar” eller ”9 rulltårtor”.

Vad är syftet med presentationen av datan och vilken roll spelar den tilltänkta målgruppen i sammanhanget?

Willigen [35] betonar vikten av att känna till vem som ska ta del av datan samt vilken nivå av kunskap denna har. I detta fall kan det antas att den som tar del av datan är väl införstådd med vad datan representerar, hur den bör tolkas samt i vilket syfte den väntas nyttjas. Gällande målgrupp är det även värt att återknyta till de avgränsningar som formulerats i avsnitt 2.2 - de som tar del av datan antas inte vara begränsade i form av synnedsättning eller dylikt.

Vilken presentation är vanlig för olika typer, och storlek, av data?

In och Lee [18] har publicerat en artikel som ger en bild av hur statistisk data kan presenteras i olika sammanhang. Noterbart är att, även om artikeln behandlar ämnet utifrån data av betydligt större storlek än den som är aktuella i denna studie, finns en hel del insikter som anses relevanta oavsett datamängd. Författarna menar att data kan presenteras på ett av följande sätt; i textform, i tabellform eller i diagram.

En tabell, menar de, bör nyttjas om all data kräver likvärdig uppmärksamhet. Vidare kan tabeller nyttjas i fall där både kvantitativa och kvalitativa värden ingår eller i fall där en numerisk precision är viktig att framhäva, exempelvis tal med flera decimaler. En nackdel som lyfts är att information som presenteras i tabeller, generellt sett tar längre tid att tolka än motsvarande i diagramform.

Grafer är att föredra om det essentiella är att visa på trender eller helhetsperspektiv. Författarna menar att även om grafer typiskt sett anses effektiva för att presentera stora mängder data, kan de nyttjas som alternativ till tabeller även för mindre datamängder. Vilken typ av graf eller diagram som bör väljas beror på den data som ska presenteras och i vilket syfte - exempelvis används ofta stapeldiagram för att visa på jämförelser medan cirkeldiagram är effektiva för att beskriva hur data är kategoriserad.

Om datan endast innefattar ett eller två tal, kan en textbaserad presentation vara mer lämpad då tabeller och diagram, i sådana sammanhang, enbart upptar plats utan att ge ökad förståelse för den faktiska datan.

Carswell och Ramzy [5] har publicerat en studie som syftar till att utreda huruvida grafer bör nyttjas för små datamängder. I studien jämfördes, med hjälp av tre testgrupper om 16 deltagare per grupp, hur effektiva tabeller, stapel- och linjediagram var för att presentera olika dataset med 4, 7 och 13 värden. För varje antal värden, visades total 11 dataset med varierande komplexitet, där författarna definierar ett dataset med avvikelser från ett linjärt mönster som mer komplext.

Studien genomfördes genom att låta varje deltagare se en representation under en valfri tid - till exempel ett stapeldiagram med sju värden - och sedan be dem sammanfatta den information de tolkat. Sammanfattningarna kategoriserades därefter baserat på hur väl deltagarna uppfattat helheten, relationer inom ett dataset samt numeriska värden. Studien visade att stapeldiagram tolkades bäst sett till helheten, medan tabeller gav sämst resultat utifrån den aspekten. Dessutom kunde det ses att deltagarna spenderade mer tid på att studera tabeller, utan att ges bättre insikter.

Finns det några riktlinjer kring de olika sätten att presentera data?

Curran [8] har publicerat en artikel innehållande riktlinjer kring hur data bör presenteras beroende på aktuellt format. Författaren menar att det faktiska valet av verktyg för datapresentation, beror på mängden och typ av data samt vem som ska ta del av informationen. Här fokuseras på två av de alternativ som In och Lee [18] presenterat - tabell och diagram.

Curran menar att tabeller bör nyttjas om datasetet innefattar 20 eller färre datapunkter. Tabellen bör utformas på så sätt att raderna arrangeras i en sjunkande ordning, med större tal ovanför mindre, även om detta resulterar i att datan presenteras i en ordning som skiljer sig mot den råa sammanställningen. Dessutom bör kolumnerna grupperas så att data som kan eller bör jämföras, är presenterade i anslutning till varandra.

Vidare menar författaren att grafer eller diagram bör nyttjas om datasetet är stort eller om trender bör betonas. Ett av dessa, stapeldiagram, anges som applicerbart i flertalet fall. Vid nyttjandet av stapeldiagram bör dessa, om möjligt, presenteras så att staplarna visas i sjunkande ordning. Huruvida staplarna bör presenteras vertikalt eller horisontellt, beror dels på datatypen och dels på den etikett (och dess längd) som associeras med en given stapel.

Finns det några riktlinjer kring interaktiva element, animationer och visuella effekter?

Willigen [35] beskriver hur estetiska element i vissa fall kan ses som en visuell bonus, medan det i andra fall kan påverka i sådan utsträckning att den faktiska effektiviteten ökar. En viktig notering dock, är att behovet av estetiska element och den påverkan de har, till stor del beror på vad som motiverar någon att ta del av given information. Om motivationen att ta del av information är hög, spelar estetiken troligen mindre roll, medan en låg motivation kan öka om presentationen är estetiskt tilltalande. Om de tillförda elementen upplevs som komplexa kan motivationen däremot minska, varför vikt läggs vid att upprätthålla en enkel design.

För att utreda hur animationer upplevs, genomförde författaren ett test med totalt 100 deltagare. Testet i sin helhet innebar att deltagarna fick se olika typer av animationer i samband med grafer och utvärdera upplevelsen av dessa. Testet indelades i två delar, med tre test för vardera del. I den första delen utvärderades animationer som adderats vid laddning av en graf, medan den andra delen handlade om en animation vid övergång från en grafvy, till en annan.

I det första testet under den första delen, visades grafen utan animering medan grafen laddade. Test två visade en lugn animation under laddning, medan det tredje testet visade en studsig animation. I det första testet under den andra delen, visades en statisk övergång utan animationer. Test två visade en animerad övergång i ett steg, medan det tredje testet visade en animerad övergång i flera steg. I listan nedan sammanfattas de viktigaste insikterna från studien.

- När två olika animationer för laddning av graf (studsig och lugn) jämfördes, kunde det ses att den lugna animationen föredrogs.
- Den lugna animationen upplevdes öka förståelsen för grafens innehåll, även om alla animationer sammantaget upplevdes addera mer estetik än förståelse.

- För den animerade övergången kunde ingen direkt föredragen variant ses, när den direkta och stegvisa övergången jämfördes, dock föredrogs dessa varianter över den statiska.

- Generellt sett var det fler äldre, än yngre, deltagare som föredrog 'ingen animation', oavsett del av test.

Författaren sammanfattar studien med att poängtera att det inom testgrupperna finns stora skillnader i hur de olika animationerna upplevs, och att detta indikerar att upplevelserna är individuella även om vissa tendenser kan ses på gruppnivå.

Figueiras [10] beskriver hur interaktivitet kan göra data mer engagerande, men även hur detta kan göra datan mer hanterbar då den kan delas in i mindre beståndsdelar. Genom att till exempel möjliggöra filtrering kan önskad data ges fokus medan andra delar döljs eller ges mindre utrymme. Ett annat sätt att öka förståelse för datan, utan att påverka den övergripande helhetsbilden, är att addera popup-element med utökad information. Element som dessa kan nyttjas för att ge användaren tillgång till detaljerad information utan att skapa ett kaotiskt och oorganiserat intryck av datan i stort.

Ett annat sätt att ge användaren mer kontroll över datan och möjlighet att individanpassa denna efter egna preferenser, är att ge möjlighet att sortera denna. Detta kan exempelvis innefatta möjlighet att justera gradering av axlarna på en graf, eller ändra ordning på kolumner och rader i en tabell. Författaren menar även att detta kan utökas ytterligare, till att ge användaren möjlighet att välja mellan helt skilda sätt att visualisera given data.

Slutligen nämns faktorer som färg, form och storlek. Genom att låta användaren själv styra dessa element, kan dessa individanpassas efter eget tycke och smak. Även Curran [8] nämner dessa faktorer, och menar att färg har främst påverkan på helhetsintrycket. Författaren ger ingen riktlinje kring huruvida dessa delar bör kunna väljas fritt av användaren, men betonar ett antal riktlinjer som bör följas vid hantering av färg. Maximalt bör fyra färger kombineras i en presentation och de färger som väljs bör, när applicerbart, följa allmänna uppfattningar om färgers symbolik - röd för att representera en avvikelse, till exempel. Till detta adderar Hailemicheal [14] en uppmaning om att välja färger på sådant sätt att det säkerställs att kontrasterna mellan dem inte påverkar förståelsen för datan på ett negativt sätt.

5.2.2 Sammanfattning

Det finns en gemensam uppfattning om att tabeller och diagram är användbara för att presentera data på ett tydligt sätt. Samtidigt visar resultat från litteraturstudien att det saknas definitiva riktlinjer kring vilken metod som bör nyttjas för ett givet dataset. Resultat från den studie som Carswell och Ramzy genomfört visar att små dataset som presenteras i ett stapeldiagram kan ge en bättre bild av helheten, medan Curran menar att dataset med upp till 20 datapunkter bör presenteras i tabellform. Samtidigt ger In och Lee en bild som visar att det inte nödvändigtvis handlar om datamängden, utan om dess innehåll och syfte.

Dessutom kan en del av studierna ses som motstridiga, då Curran exempelvis menar att tabeller bör följa givna regler och strukturer medan Figueiras menar att det finns fördelar med att låta användaren själv styra sådana utformningar. Vidare menar Willingen att animationer kan ge ett ökat estetiskt intryck, eller till och med ge en uppfattning om ökad förståelse för grafers innehåll, men att de variationer som ses på individnivå gör det svårt att visa på definitiva trender. Slutligen kan det ses att flera av studierna nämner färg som en faktor som behöver hanteras med viss försiktighet, för att inte skapa för stora kontraster eller ge ett överväldigande intryck som fråntar fokus från huvudmålet - att presentera en tydlig bild av given data.

Kapitel 6

Diskussion

I följande kapitel diskuteras studiens resultat, med de explicita forskningsfrågorna som utgångspunkt.

6.1 RQ1

Hur kan SMS-svar från kunder tolkas och bearbetas med hjälp av en stor språkmodell?

Genom att implementera ett SMS-API - där inkommande meddelanden skickas vidare till en språkmodell för extrahering av önskad information enligt givna mönster - kan SMS-svar från kunder tolkas och bearbetas. Hanteringen av det faktiska SMS-API:t och den webbklient som nyttjas för att initiera orderflödet, ses som sekundära i denna diskussion. Dessa delar följer en tydlig dokumentation (SMS-API) och nyttjar välkända tekniker och ramverk (webbklient), och ses därför som verktyg utan egentlig påverkan på implementationen eller dess resultat. Diskussionen för RQ1 fokuserar därför på språkmodellen i sammanhanget.

Enligt tidigare studier, som behandlats i kapitel 3, ges stöd för hantering av språkmodeller med både *zero-shot-* och *in-context*-promptning. I denna studie nyttjades det sistnämnda, då de krav som definierats i avsnitt 4.2.1 innebar att språkmodellen behövde ge en tydligt kategoriserad output för given input.

Under de tester som genomfördes som en del av utvecklingen uppnåddes tillfredsställande resultat - vilket innebär sådana som sammantaget uppfyllde och efterlevde de ställda kraven. Noterbart under utvecklingen dock, var att en och samma input kunde resultera i något varierande output även om detta ledde till likvärdiga resultat i slutändan. Ett mönster som sågs, var att inputen inte alltid kategoriserades enligt förväntningarna. Exempelvis kunde det ses att varor beställda med ett tydligt antal blev kontrollerade, trots att dessa förväntades noteras utan extra kontroll. En sådan kontroll påverkar inte resultat på sådant sätt att beställningen blir felaktig, men det ger ett extra lager som inte förväntas finnas där. Upprepade tester med exakt samma input visade även att den output som gavs, varierade på ett oförutsägbart sätt.

I ett större perspektiv i relation till tidigare studier [11], ses det som troligt att den upplevda variationen är ett resultat av en suboptimal promptningsstrategi. Det är dock svårt att avgöra om detta är en ensam faktor, eller om den nyttjade modellen - som i förhållande till andra modeller är att betrakta som liten - har en påverkan [6], [30].

Sett till enkätundersökningen är den sammantagna slutsatsen att de som testat systemet, upplever det som fungerande för sitt syfte. Det finns dock ett antal faktorer som behöver lyftas för att sätta resultaten av undersökningen i ett större perspektiv. Dels var testgruppen liten då den endast omfattade 10 respondenter. Detta gör att de synpunkter och erfarenheter som framkommit via enkätundersökningen inte kan ses som representativa i ett större sammanhang, även om de anses ge viss uppfattning om den övergripande upplevelsen av systemet. Dessutom fanns inga riktlinjer kring hur deltagarna skulle genomföra testet - mer än att de skulle besvara erbjudandet på det sätt de ansåg passande. Denna frihet kan ha begränsat resultaten då en deltagare exempelvis skulle kunna välja att neka ett erbjudande och på så sätt vara klar med genomförandet av testet. Dessutom bör påpekas att även om gruppen var oberoende på så sätt att deras kännedom om artefakten var begränsad, kan deras personliga relation till utvecklaren ha gjort dem mer benägna att värdera artefakten högre.

Enkätundersökning gav även insikt om en begränsning hos modellen där en deltagare, i fråga Q3 (se fig. 5.1), uppgav att dennes begäran blev felaktigt tolkad. Givet det resonemang som förts inledningsvis i avsnittet, är det svårt att direkt avgöra om denna feltolkning var en tillfällighet eller om det var ett systematiskt fel i implementationen.

6.2 RQ2

Hur kan insamlad data presenteras på sådant sätt att användaren av systemet enkelt kan överblicka resultaten?

Den litteraturstudie som genomförts visar att det finns primärt två rekommenderade sätt - att nyttja en tabell eller ett diagram av något slag. Det är dock, utifrån genomförd studie, svårt att dra några faktiska slutsatser om vilken metod som är att föredra i detta sammanhang. Delar av litteraturstudien visar att det finns motstridigheter kring vilken av dem som är bäst lämpad utifrån datamängden [5], [8], där en av studierna visar att diagram kan nyttjas även om antalet datapunkter är få, medan den andra föreslår att datamängder med upp till 20 punkter presenteras i tabellformat.

Det finns dessutom tvetydigheter kring huruvida interaktiva element och animationer bör vara en del av presentationen. Åsikterna kring animationer beskrivs av [35] som mer individuella, där det är svårt att generalisera på gruppnivå. Dessutom menar [10] att interaktiva element ger individen ökat inflytande och mer kontroll över datan, medan [8] ger en mer strikt mall för hur exempelvis tabeller bör utformas, om sådana nyttjas.

Sammantaget kan ses att tabeller och diagram är att föredra även om det råder skilda meningar kring vilken av metoderna som bör appliceras och när. Dessutom finns det inga klara direktiv kring utformning, estetik och interaktivitet - utan dessa delar bör anpassas efter individen.

Kapitel 7

Slutsatser och framtida arbete

I följande kapitel presenteras slutsatser i relation till de två forskningsfrågorna, tillsammans med förslag på fortsatt arbete för dessa. Slutligen sätts arbetet i ett större etiskt perspektiv.

7.1 RQ1

Hur kan SMS-svar från kunder tolkas och bearbetas med hjälp av en stor språkmodell?

Studien visar att det är fullt möjligt att implementera ett SMS-API tillsammans med en språkmodell för att på ett tillfredsställande sätt hantera ett orderflöde med in- och utgående meddelanden. Genom att nyttja *in-context*-promptning på ett strukturerat sätt kan orderflödet hanteras utan manuell intervention.

Även om studiens resultat kan ses som tillräckliga för att ge ett svar på den initiala frågeställningen, finns flertalet aspekter som kan eller bör utredas vidare för att ge en än mer nyanserad bild av problemet. Gällande implementationen i sig, ses följande delar som relevanta att utforska vidare:

- Hur kan promptningen effektiviseras för att minimera risken för spontana avvikelser från det förväntade beteendet?
- Kan en annan språkmodell (jämfört med nuvarande GPT-4o-mini) ge ett mer tillförlitligt system utan ändring av övriga förutsättningar eller villkor?
- Hur väl anpassad är nuvarande implementation för användning i större skala?
- Hantering av risker, hot och medveten vilseledning:
    - Finns det någon risk att meddelandeinformation kan nås av obehöriga och vad skulle en sådan risk innebära för systemet och dess användare?
    - Finns det andra säkerhetsrisker eller hot kopplade till nyttjandet av en språkmodell?
    - Hur kan medveten felhantering och vilseledning av kunder hanteras?

Dessutom, med bakgrund mot tidigare studier [21], [16] och inofficiell korrespondens[1] från testdeltagare, bör det utredas hur systemet kan ge ett trovärdigt och tillförlitligt intryck samtidigt som det bygger på nyttjandet av artificiell intelligens.

[1] Med inofficiell menas här information som framkommit utanför enkätundersökningen

7.2 RQ2

Hur kan insamlad data presenteras på sådant sätt att användaren av systemet enkelt kan överblicka resultaten?

Den genomförda litteraturstudien anses inte ge tillräckligt stöd för att ge ett tydligt svar på denna fråga. De slutsatser som kan dras är att en tabell eller ett diagram bör ses som lämpliga alternativ, men att det faktiska valet mellan metoderna - i givet scenario - saknar definitiva riktlinjer i litteraturen. Dessutom saknas konsensus kring hur de olika alternativen bör implementeras - det vill säga hur en tabell eller ett givet diagram bör utformas i ett givet fall.

Även om den metod som nyttjats för att utreda frågan, se avsnitt 4.3, är en valid sådan så anses den inte tillräcklig i detta fall. Att den anses otillräcklig bör dock inte tolkas som att den inte bidragit med några användbara resultat alls. En viktig insikt som gavs i samband med litteraturstudien är att frågan och framförallt svaret på densamma, påverkas av individuella åsikter och uppfattningar. Detta ger god grund för fortsatt arbete för att utreda frågan, där framförallt följande rekommenderas:

- Att genomföra intervjuer med deltagare som kan anses representera tilltänkta användare av systemet.
- Baserat på intervjuerna och med stöd av litteraturstudien bör (minst) två prototyper skapas. Därefter bör en testgrupp nyttjas för utvärdering av vilken av prototyperna de föredrar.
- Att utvärdera om det mest fördelaktiga, i givet fall, är att ge användaren av systemet möjlighet att växla mellan olika tabell- och diagramvyer, i enlighet med tidigare förslag [10].

7.3 Etiska aspekter

Av resursskäl har aspekter gällande tillgänglighet och anpassning därefter helt utelämnats i denna studie, se avsnitt 2.2. Dessa delar bör dock inte ses som mindre prioriterade av varken författaren eller samarbetsföretaget. Fortsatt arbete uppmuntras därför ta hänsyn till dessa aspekter, i synnerhet vid utredandet av RQ2, men även vid vidareutveckling av prototypen för RQ1 - när och om applicerbart.

Som nämnts i avsnitt 7.1 ovan, bör risker associerade med nyttjandet av språkmodeller utredas och klargöras - från ett objektivt perspektiv. Det senare avser betona vikten av att vara transparent med eventuella risker som en användare av systemet - säljare eller kund - kan komma att utsättas för om denne väljer att använda detta. Det primära målet bör dock i första hand vara att åtgärda eventuellt upptäckta risker, för att eliminera dem. Om och när en risk fortsatt finns, bör samtliga som kommer i kontakt med systemet göras medveten om denna.

Referenser

[1] 46elks, "46elks API Reference," 2025, [Online]. Tillgänglig: https://46elks.se/docs/overview. Hämtad: 2025-03-09.

[2] 46elks, "Ett enkelt API för SMS och telefoni," 2025, [Online]. Tillgänglig: https://46elks.se/. Hämtad: 2025-02-05.

[3] G. Bharathi Mohan, R. Prasanna Kumar, P. Vishal Krishh *et al.*, "An analysis of large language models: their impact and potential applications," *Knowledge and Information Systems*, vol. 66, pp. 5047–5070, 2024, [Online]. Tillgänglig: https://doi.org/10.1007/s10115-024-02120-8. Hämtad: 2025-04-19.

[4] M. Burtsev, M. Reeves, and A. Job, "The working limitations of large language models," *MIT Sloan Management Review*, vol. 65, no. 1, pp. 2–4, 2023.

[5] C. M. Carswell and C. Ramzy, "Graphing small data sets: Should we bother?" *Behaviour & Information Technology*, vol. 16, no. 2, pp. 61–71, 1997, [Online]. Tillgänlig: 10.1080/014492997119905. Hämtad: 2025-04-13.

[6] Y.-C. Chang and E. Aïmeur, "Chat or trap? detecting scams in messaging applications with large language models," pp. 92–99, 2024, [Online]. Tillgänglig: https://ieeexplore.ieee.org/document/10851753. Hämtad: 2025-04-18.

[7] K. R. Chowdhary, *Fundamentals of Artificial Intelligence.* New Delhi, Indien: Springer, 2020.

[8] C. R. Curran, "Data display techniques," *Applied Nursing Research*, vol. 12, no. 3, pp. 153–158, 1999, [Online]. Tillgänlig: https://doi.org/10.1016/S0897-1897(99)80067-6. Hämtad: 2025-04-13.

[9] F. Esposito, *Programming Large Language Models with Azure Open AI: Conversational programming and prompt engineering with LLMs.* USA: Pearson Education, Inc., 2024.

[10] A. Figueiras, "Towards the understanding of interaction in information visualization," in *2015 19th International Conference on Information Visualisation*, 2015, pp. 140–147, [Online]. Tillgänlig: https://doi.org/10.1109/iV.2015.34. Hämtad: 2025-04-13.

[11] T. Gao, A. Fisch, and D. Chen, "Making pre-trained language models better few-shot learners," *arXiv:2012.15723v2 [cs.CL]*, 2020.

[12] Geeks for Geeks, "NLP vs LLM: Understanding Key Differences," 2025, [Online]. Tillgänglig: https://www.geeksforgeeks.org/nlp-vs-llm/. Hämtad: 2025-03-06.

[13] Google Scholar, "Google Scholar," 2025, [Online]. Tillgänglig: https://scholar.google.com/. Hämtad: 2025-03-04.

[14] E. Hailemichael, "Top tools for data visualisation in web development: a performance first approach," 2016, metroplia University of Applied Sciences, Finland. [Online]. Tillgänglig: https://urn.fi/URN:NBN:fi:amk-2016120519133. Hämtad: 2025-02-06.

[15] M. Hur, S. Seo, J. Hwang, H. Lim, and M. Min, "Utilizing large language models for detection of sms spam in few-shot settings," *SSRN*, 2024, [Online]. Tillgänglig: http://dx.doi.org/10.2139/ssrn.4815382. Hämtad: 2025-04-18.

[16] Ipsos, "Global Views On A.I. In 2023," Jul 2023, [Online]. Tillgänglig: https://www.ipsos.com/en-nz/global-views-ai. Hämtad: 2025-03-10.

[17] P. Johannesson and E. Perjons, *An Introduction to Design Science*, 2nd ed. Cham, Schweiz: Springer, 2021.

[18] I. Junyong and L. Sangseok, "Statistical data presentation," *Korean J Anesthesiol*, vol. 70, no. 3, pp. 267–276, 2017, [Online]. Tillgänglig: 10.4097/kjae.2017.70.3.267. Hämtad: 2025-04-13.

[19] Just Send! by 46elks, "Web client for broadcasting SMS," 2025, [Online]. Tillgänglig: https://justsend.46elks.com/. Hämtad: 2025-03-10.

[20] T. Kojima, S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," in *Advances in Neural Information Processing Systems*, New Orleans, USA, 2022, pp. 22 199–22 213, [Online]. Tillgänglig: https://proceedings.neurips.cc/paper_files/paper/2022/file/8bb0d291acd4acf06ef112099c16f326-Paper-Conference.pdf. Hämtad: 2025-03-07.

[21] N. Maslej, L. Fattorini, R. Perrault, V. Parli, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J. C. Niebles, Y. Shoham, R. Wald, and J. Clark, "The AI Index 2024 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, Kalifornien, USA, Tech. Rep., 2024, [Online]. Tillgänglig: https://hai.stanford.edu/ai-index/2024-ai-index-report. Hämtad: 2025-03-10.

[22] A. Matarazzo and R. Torlone, "A Survey on Large Language Models with some Insights on their Capabilities and Limitations," *arXiv preprint arXiv:2501.04040*, 2025.

[23] McKinsey & Company, "The state of AI in 2023: Generative AI's breakout year," Aug 2024, [Online]. Tillgänglig: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year. Hämtad: 2025-03-10.

[24] H. Naveed *et al.*, "A comprehensive overview of large language models," *arXiv:2307.06435v10 [cs.CL]*, 2024.

[25] B. J. Oates, R. McLean, and M. Griffiths, *Researching Information Systems and Computing*, 2nd ed. London, England: SAGE Publications Ltd, 2022.

[26] OpenAI, "OpenAI," 2025, [Online]. Tillgänglig: https://openai.com/. Hämtad: 2025-03-06.

[27] Opper, "Calls | opper," 2025, [Online]. Tillgänglig: https://docs.opper.ai/capabilities/calls/. Hämtad: 2025-04-21.

[28] Opper, "Opper," 2025, [Online]. Tillgänglig: https://opper.ai/. Hämtad: 2025-03-06.

[29] S. A. Richard Cooper and C. Bi, "Extracting information from short messages," Berlin, Heidelberg, pp. 388–391, 2005.

[30] M. Salman, M. Ikram, N. Basta, and M. A. Kaafar, "Spallm-guard: Pairing sms spam detection using open-source and commercial llms," 2025. [Online]. Available: https://arxiv.org/abs/2501.04985

[31] Summon@BTH, "Summon@bth," 2025, [Online]. Tillgänglig: https://bth.summon.serialssolutions.com/#!/. Hämtad: 2025-03-04.

[32] H. J. Tobias Ek, Camilla Kirkegaard and P. Nugues, "Named entity recognition for short text messages," pp. 178–187, 2011, computational Linguistics and Related Fields. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877042811024232

[33] E. Topics, "Number of parameters in gpt-4 (latest data)," https://explodingtopics.com/blog/gpt-parameters, 2025, [Online]. Tillgänglig: https://explodingtopics.com/blog/gpt-parameters. Hämtad: 2025-05-04.

[34] P. Törnberg, "How to use Large Lanugage Models for Text Analysis," *arXiv preprint arXiv:2307.13106*, 2023.

[35] T. van Willigen, "Measuring the user experience of data visualization," Mars 2019, [Online]. Tillgänglig: http://essay.utwente.nl/77564/. Hämtad: 2025-04-13.

[36] J. White, "A prompt pattern catalog to enhance prompt engineering with chatgpt," *arXiv:2302.11382 [cs.SE]*, 2023.

[37] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *EASE '14: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, London, England, 2014, pp. 1–10, [Online]. Tillgänglig: https://dl.acm.org/doi/pdf/10.1145/2601248.2601268. Hämtad: 2025-03-06.

Bilaga A

Orderflöde

Transkriberat utdrag från en beställning gjord i syfte att demonstrera dess flöde.

Hej Kund!

Den här veckan är det mazariner, brownies och tekakor som gäller. Svara med vad du vill ha, och hur många!

Vill du hellre ha något annat? Skicka med det i din beställning, så försöker jag ordna det!

Hälsningar, BrödBagaren & Co.

Trött på brödtjöt? SMSa 'brödmätt' för att sluta motta erbjudanden.

Jag kan ta två mazariner och 4 brownies.

Bra val, jag lägger din beställning i mitt brödskrin så att jag har koll på den!

Önskas något mer? Om inte, SMSa 'bekräfta' för att bekräfta din order.

SMSa 'brödlista' för att se en översikt av din order.

Gärna två semlor också

Jag bakar in det i din beställning!

Önskas något mer? Om inte, SMSa 'bekräfta' för att bekräfta din order.

SMSa 'brödlista' för att se en översikt av din order.

Är det möjligt att beställa någon enklare maträtt? Typ en sallad eller kanske någon wrap av något slag?

Tyvärr säljer vi inte maträtter som sallad eller wrap.

Önskas något annat?

SMSa 'brödlista' för att se en översikt av din order.

Brödlista

Du har beställt:
2 mazariner
4 brownies
2 semlor

Vill du ta bort något? Skriv 'ta bort' följt av det du vill ta bort, så löser jag det!

Vill du lägga till något mer? SMSa vad och vilket antal så lägger jag till det i brödskrinet!

SMSa 'brödkris' för att avbryta din order eller SMSa 'bekräfta' för att bekräfta din order.

Vänta, ta fyra mazariner istället

Menar du fyra mazariner totalt eller fyra till utöver de två som redan är beställda?

Fyra totalt, tack

Yes, okej!

Jag snor ihop 4 mazariner!

Önskas något mer? Om inte, SMSa 'bekräfta' för att bekräfta din order.

Osäker vad du stoppat i brödkorgen? SMSa 'brödlista' för att se en översikt av din order."

Brödlista

Du har beställt:
4 mazariner
4 brownies
2 semlor

Vill du ta bort något? Skriv 'ta bort' följt av det du vill ta bort, så löser jag det!

Vill du lägga till något mer? SMSa vad och vilket antal så lägger jag till det i brödskrinet!

SMSa 'brödkris' för att avbryta din order eller SMSa 'bekräfta' för att bekräfta din order.

Bra så. Då hoppas jag att jag får allt den här gången. Förra veckan glömde ni delar av beställningen men jag betalade ju för allt.

Bästa Kund,

Jag uppfattar det som att vi eventuellt behöver reda ut en och annan glutentråd, ska jag slå dig en signal så att vi kan knåda oss igenom det här tillsammans?

Hälsningar,
BrödBagaren & Co.

Ja gör gärna det, så kanske vi kan komma överens om lite rabatt den här veckan.

Självklart, jag ringer upp dig så fort jag får mjölighet, eller ja, möjlighet!

Jäst a minute, så att säga.

Toppen!

Fakulteten för datavetenskaper, Blekinge Tekniska Högskola, 371 79 Karlskrona, Sverige