Intelligent Channel Navigation in Customer Service using Large Language Models

Constantin Ehmanns, 17.02.2025, Master Thesis Final Presentation

Lehrstuhl für Software Engineering betrieblicher Informationssysteme (sebis)
Fakultät für Informatik
Technische Universität München
wwwmatthes.in.tum.de

Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

How to improve customer service through multichannel navigation?

Insurance company wants to explore ways to improve customer journey through optimal channel usage

Company Environment

Large European insurance company

- **Call** with human operator
- **Chat** with human operator
- Self-service **web** forms
- Self-service **app**

Problems

Issues motivating the thesis

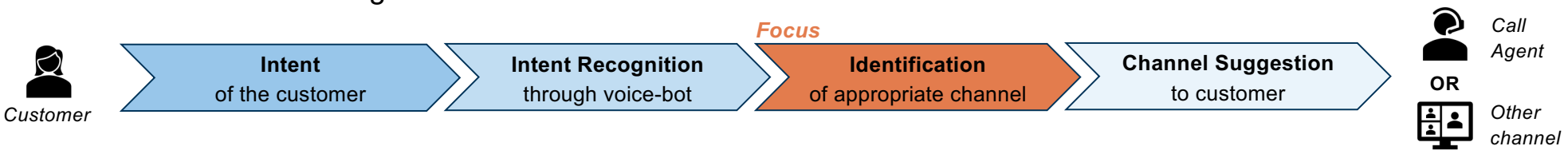
- Human operator-backed channels are expensive and are the **bottleneck**
- No insights into channel switching preferences/acceptance of customers
- No data available for this specific use case (neither with nor without labels)

Thesis Scope

Reduce organizational complexity

- **German branch** of the insurance company
- Content focus: **damage claims only**
- Starting point for potential **channel switch: call channel**

Process of interest: concierge bot in customer service



Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

How to improve customer service through multichannel navigation?

Guiding research questions

RQ1: In customer service centers, what are **relevant factors** for deciding the optimal channel for customer service requests?

- Structured literature review
- Interviews in case study company

→ **List of decision factors**

RQ2: How do different **input factors** and **prompt strategies influence the effectiveness of LLMs** in selecting appropriate communication channels for customer service requests?

- Independently test input factors
- Similarly test prompt strategies

→ **Analyze performance**

RQ3: How well do LLMs **predict the reasons** for choosing a customer service channel?

- Predict reasons for channel decisions with LLM

→ **Compare similarities**

Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

7 Factors identified for comprehending channel choice

Lack of data requires understanding what drives the channel choice from customer side

Structured Literature Review

- Customer channel choice determinants in customer care / services[1]
- 2000 – 2024 (October)
- EBSCO Host, Scopus, Web of Science

47 factors from 19 articles

Workshop and Interviews

- What factors should play a role in deciding the channel for the customer?
- 8 participants (5 in workshop, 3 in interviews)
- Green field

46 factors from 4 interactions

Factor Overview		
Only in Literature	**Mentioned in Both**	**Only in Company**
Gender	**Age**	App registration
Redress seeking	Comfort	Digital channels used before
Usage frequency	Channel affinity	End-customer vs. intermediary
Marital status	Integration quality	Patience
Information need	**Intent**	Acute danger
Engagement level	Venting anger	Customer has all necessary data for process
Trust in security	Social / Personal connection	Customer authenticated
Privacy	**Previously used channels**	Digital process exists
Mobile identity	**Infrastructure**	Customer satisfaction with process
Relationship to company	Channel awareness	Process-transparency
Personal assertiveness	Openness to change	(Process cost)
(Hidden) cost	Perceived necessity (to use other channels)	Prioritization
Region	**Complexity**	Customers' uncertainty
Income per region	Product risk	Potentially not saying the truth
Perceived reliability	Range of channel provision	Time already spent in channel
Current satisfaction	Channel usefulness	How visible is the phone number vs. other channel entries
Perceived media richness	Availability	Pre-filled data available
Culture of customer	Service quality	Customer vs. family/...
Income	Error risk	Multiple contact attempts in same channel
Working hours	**Innovativeness**	Language barrier
Ethical considerations	**Previous claim**	Multiple conctact attempts across channels
Company reputation		(Upselling potential)
Personal control		(Conversion rate)
Recommendation		
Physical constraints		
Cognitive abilities		

1 Extended from: L. Wolf and M. Steul-Fischer (Dec. 2023)

Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

Survey with over 700 people revealed customer preference

Survey with 709 insurance customers produced first data set in the problem space

Survey department offered logistic support for survey

We developed a questionnaire for our use case

709 customers* approached (through the department)

*Each participant worked on three scenarios

Participant is confronted with three scenarios …

„Bitte stellen Sie sich folgende Situation vor: Sie rufen den Kundenservice Ihres Versicherungsunternehmens an, sind aber noch nicht mit einem Mitarbeiter verbunden.

Sie haben versehentlich das
Handy von einem Freund fallengelassen, dabei wurde es beschädigt.
Diesen simplen Haftpflichtschaden
wollen Sie Ihrem Versicherungsunternehmen melden.
Bisher hatten Sie die Versicherung noch nicht
kontaktiert, um diesen Schaden zu
melden.“

… and asked questions about them

- How do you assess the following communication channels for reporting the previous situation?

  The listed **channels** are: *call/call-back*, *chat*, *website*, *app*.

  Each channel can be given one of three **categories**:
  “*preferred*”,
  “*not ideal but acceptable*”,
  “*would not use*”.

- Please justify why you chose <preference> for channel <channel>?
- *And many more …*

Feature rich data set with 2127 scenario evaluations

Survey with 709 insurance customers produced first data set in the problem space

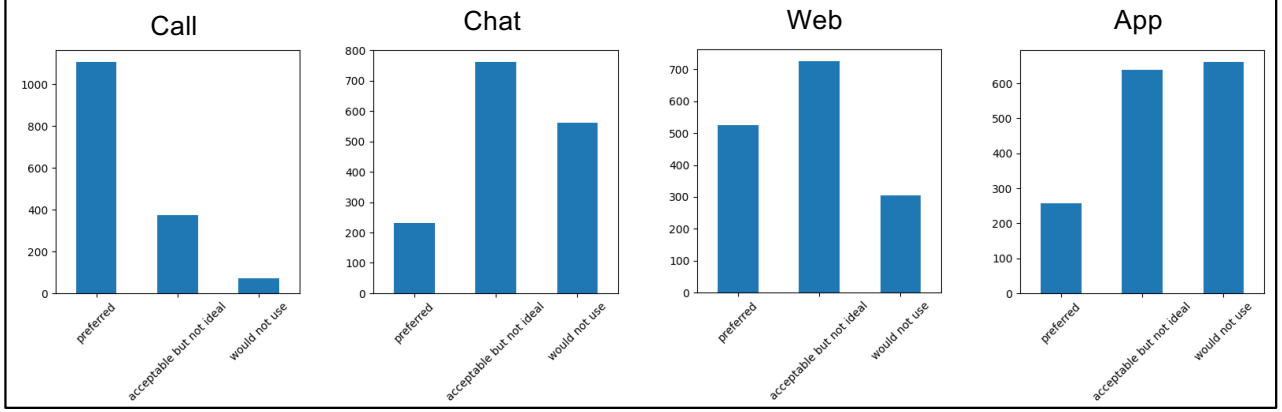
Resulting data set with 2127 data points

Evaluation data set		
Variable	**Type**	**Variable Level**
Age	Numeric	
Previously used channels	Categorical	personal visit at broker, personal visit at home, video-call with broker, call with broker, hotline, e-mail, WhatsApp, chat with human, chatbot, customer portal, app, sms, web-form
Maximum waiting time hotline satisfied	Numeric [min]	
Minimum waiting time hotline dissatisfied	Numeric [min]	
Proneness to use new technology	5-Point Likert Scale	1-5
Interest for AI	5-Point Likert Scale	1-5
Ease to use new software	5-Point Likert Scale	1-5
Registration at customer portal	Binary	True, False
Customer type	Categorical	Proprietary customer types (4)
Available devices	Categorical list	tv, pc, laptop, tablet, e-book, phablet, smartphone, regular mobile, gaming console, voice assistant, media receiver, streaming box, smartwatch, hybrid pc
Intent (type of damage)	Categorical	(liability, lost key, accident, break-in, pipe-damage) case
Complexity of claim	Categorical	Low, High
Existing previous claim	Binary	True, False
Preference for call channel	Categorical	preferred, acceptable but not ideal, would not use
Preference for chat channel	Categorical	preferred, acceptable but not ideal, would not use
Preference for web channel	Categorical	preferred, acceptable but not ideal, would not use
Preference for app channel	Categorical	preferred, acceptable but not ideal, would not use
Reason for choosing preference of channel X	Free text	
Ranking of relevance of given information	Ranking of options	complexity, intent, previous claim
Description of situation in own words	Free text	
Was there necessary information missing to decide the preference?	Binary	
Information that was missing	Free text	
Not used: state, job type, school education, household size and number of children		

Data set manually labeled for useful utterances:
~73% remaining

Set aside 5 datapoints for fewshotting; split the remaining in 80% training and 20% test set

Preference distribution



Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

Two levers: System instructions and input information

API-access only restricted LLM optimizations to prompt and input structure

System Instructions
Description of the problem and the intented outcome

Input
Detailing of the individual customers situation

Output
channel preferences based on the input

API call to Azure OpenAI services (mostly GPT-4o-mini)

You are an assistant for a customer service center of an insurance company. Your goal is to help customers navigate to the appropriate channel based on a variety of factors which are given in the user prompt. The available channels are: Call, Chat, Self-Service Web, Self-Service App. Assign a label to each channel. The labels are: preferred, acceptable, undesired. Multiple channels with the same label are possible. Provide your answer as a JSON string with the channel names as key.

+

Customer said: “Ich habe doch schonmal Bescheid gegeben, wegen dieser Sache mit meinem Schlüssel”; intent: lost key; complexity: low, existing claim: True; age: 32; previously used channels: [“call_hotline”; “visit_broker”; “chat”]; available_devices: [“pc”, “smartphone”, “tablet”]; innovativeness: “low”

=

LLM ouput parsed to JSON object

```
{
    "Call": "preferred",
    "Chat": "preferred",
    "Self-Service Web": "undesired",
    "Self-Service App": "acceptable"
}
```

MCC, F1-Score and custom scoring for determining suitability

Selection of various metrics for comparing model performance

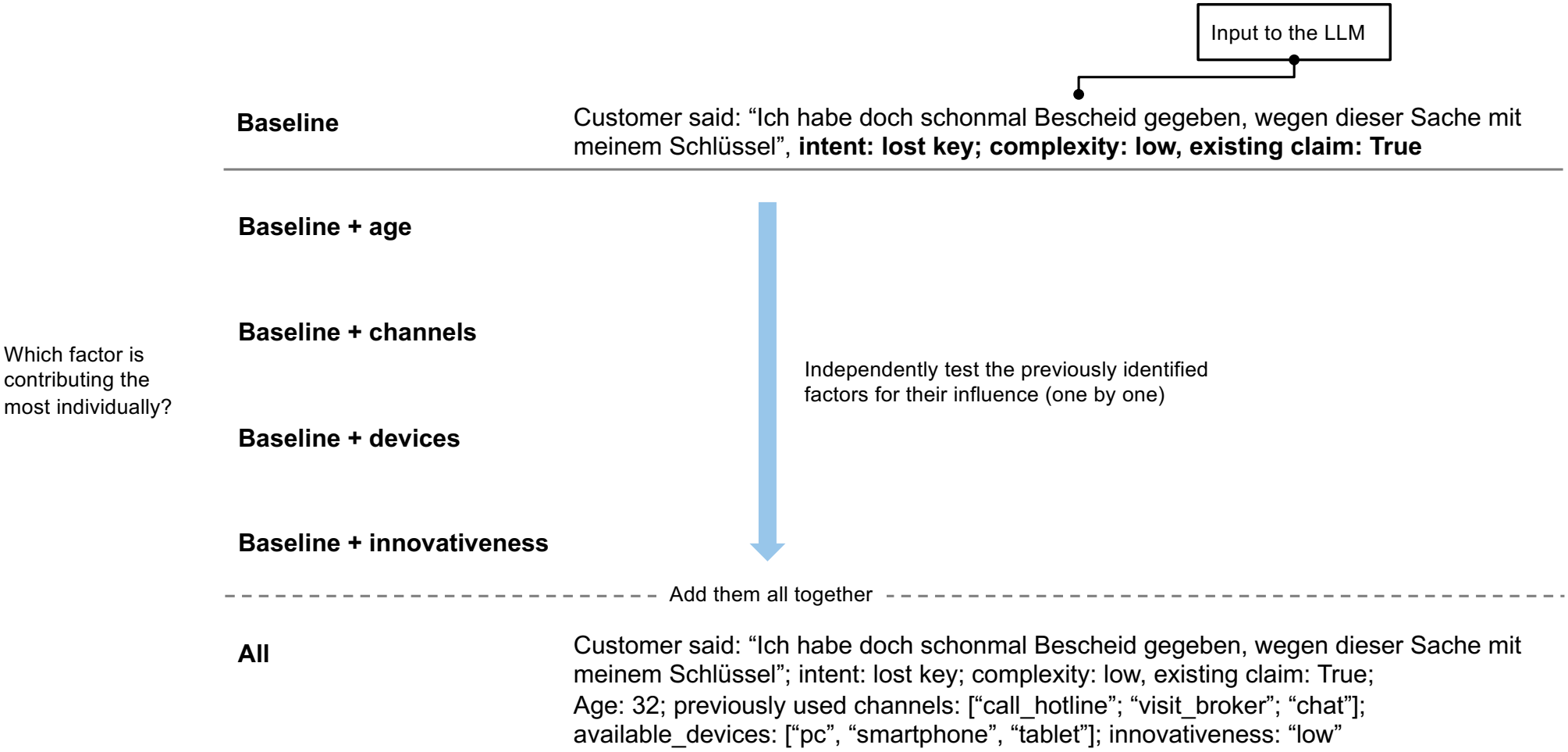
	F1-Score	**Matthews Correlation Coefficient**	**Custom Score[2]**
	Combining precision and recall through harmonic mean	Correlation score useful for unbalanced datasets with multiple classes[1]	Incorporating varying degrees of *true*: preferred vs. acceptable and preferred vs. would not use
Range	[0, 1]	[-1, 1]	[0, 1]
Interpretability	The higher, the better	1 means perfect prediction 0 means random -1 means inverse prediction	The higher, the better

1 Jurman et al. 2012. A Comparison of MCC and CEN Error Measures in Multi- Class Prediction

2 Modififed from: Rao et al. 2023. Evaluating ChatGPT as an Adjunct for Radiologic Decision-Making

Which input factor makes a difference?

API-access only restricted LLM optimizations to prompt and inputs

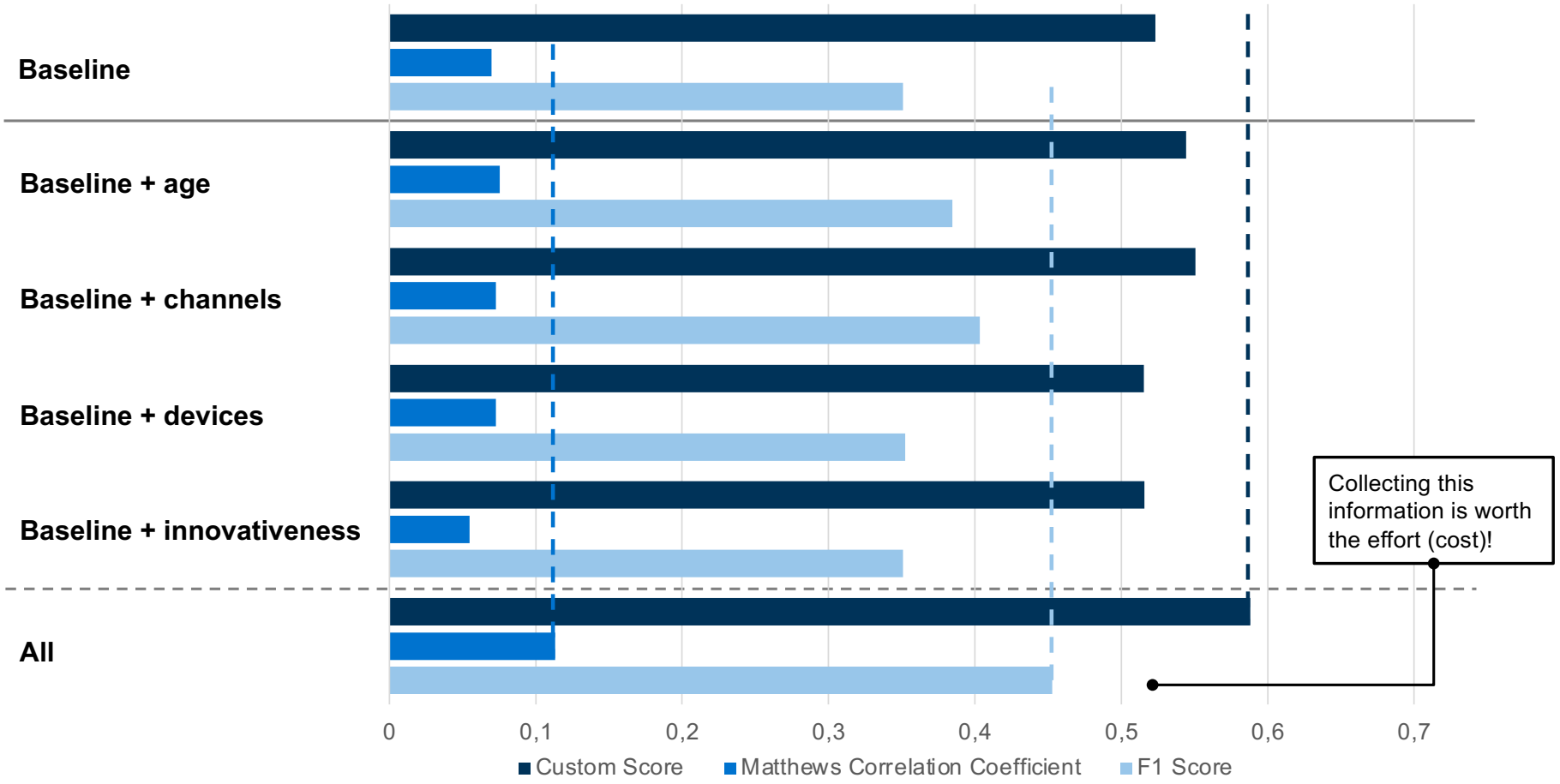


The more the merrier … adding all input factors improves performance

Innovativeness seems to be the least strong contributor as standalone factor

Which factor is contributing the most individually?

Innovativeness seems to contribute the least



Note: All metrics are averaged over the different channels

Which prompt strategy works the best?

Independently test few-shots, channel descriptions and prompt types against each other

Baseline (BL) Hand-crafted prompt based on prompt library from Anthropic for classifier

Few-shots

3-shot *vs.* **5-shot**

better than BL *better than BL*

Channel descriptions

Bullets *vs.* **JSON** *vs.* **Plain**

worse than BL *worse than BL* *worse than BL*

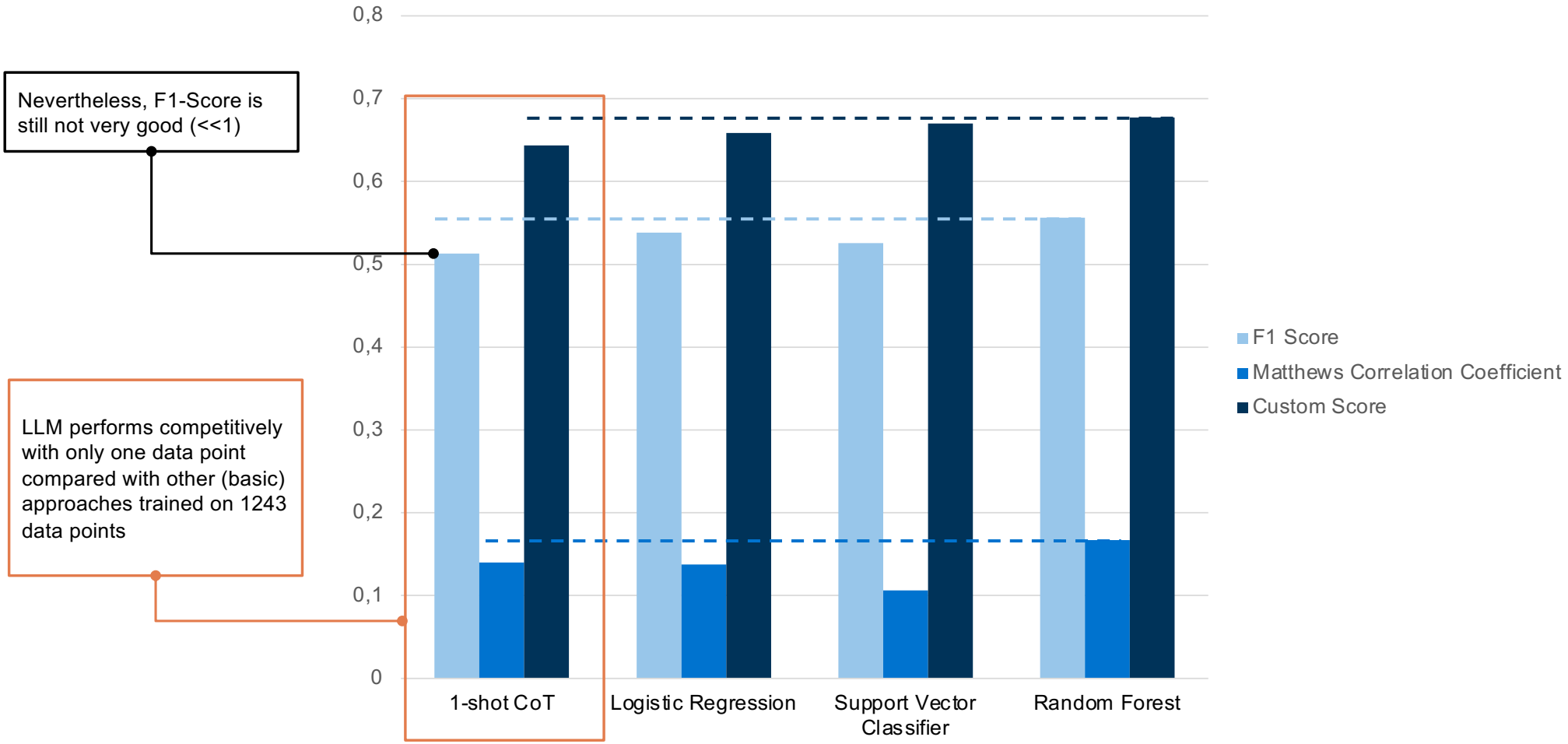
Prompt Type

LLM-generated *vs.* **(Short) Zero-shot Chain-of-Thought** *vs.* **Zero-shot Chain-of-Thought** *vs.* **Tree-of-Thought**

better than BL *better than BL* *better than BL* *worse than BL*

Exploratory search showed (1-shot) CoT as strong prompt candidate

1-shot Chain-of-Thought performs similar to basic data-driven ML techniques on the test set

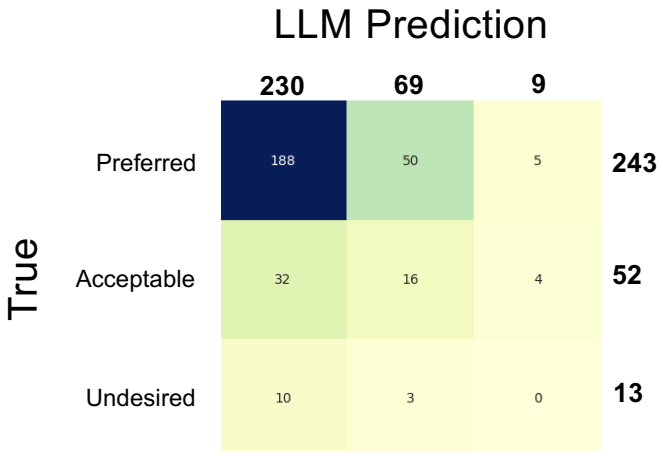


Note: All metrics are averaged over the different channels

Varying performance per channel

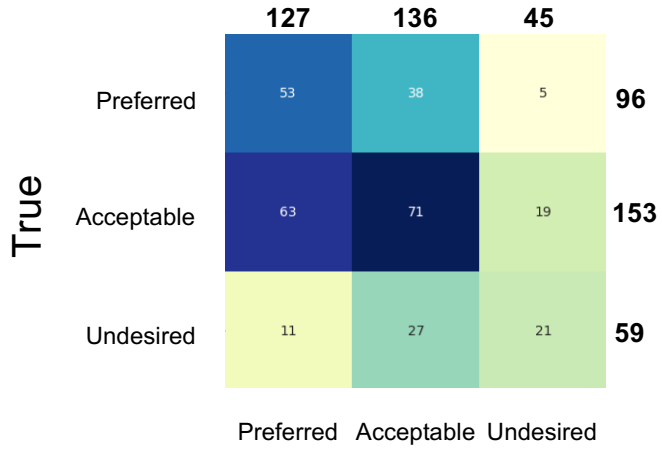
For the web channel, the LLM-classifier works better than for the call channel

Call



- Of the cases deemed “preferred” by the LLM the majority are actually “preferred”
- Almost all cases where the customer labeled “undesired” were wrongly classified as “preferred”
- Not one case the LLM labeled “undesired” was correct

Web



- For the customer labels “preferred” and “acceptable”, the LLM delivered a correct prediction in most cases – the second most frequent is the respective other (“preferred”, “acceptable”)
- Albeit still missclassifying more often than not - the “undesired” discrimination works significantly better than for the call channel

Note: an ideal classifier would have a diagonal matrix for a confusion matrix

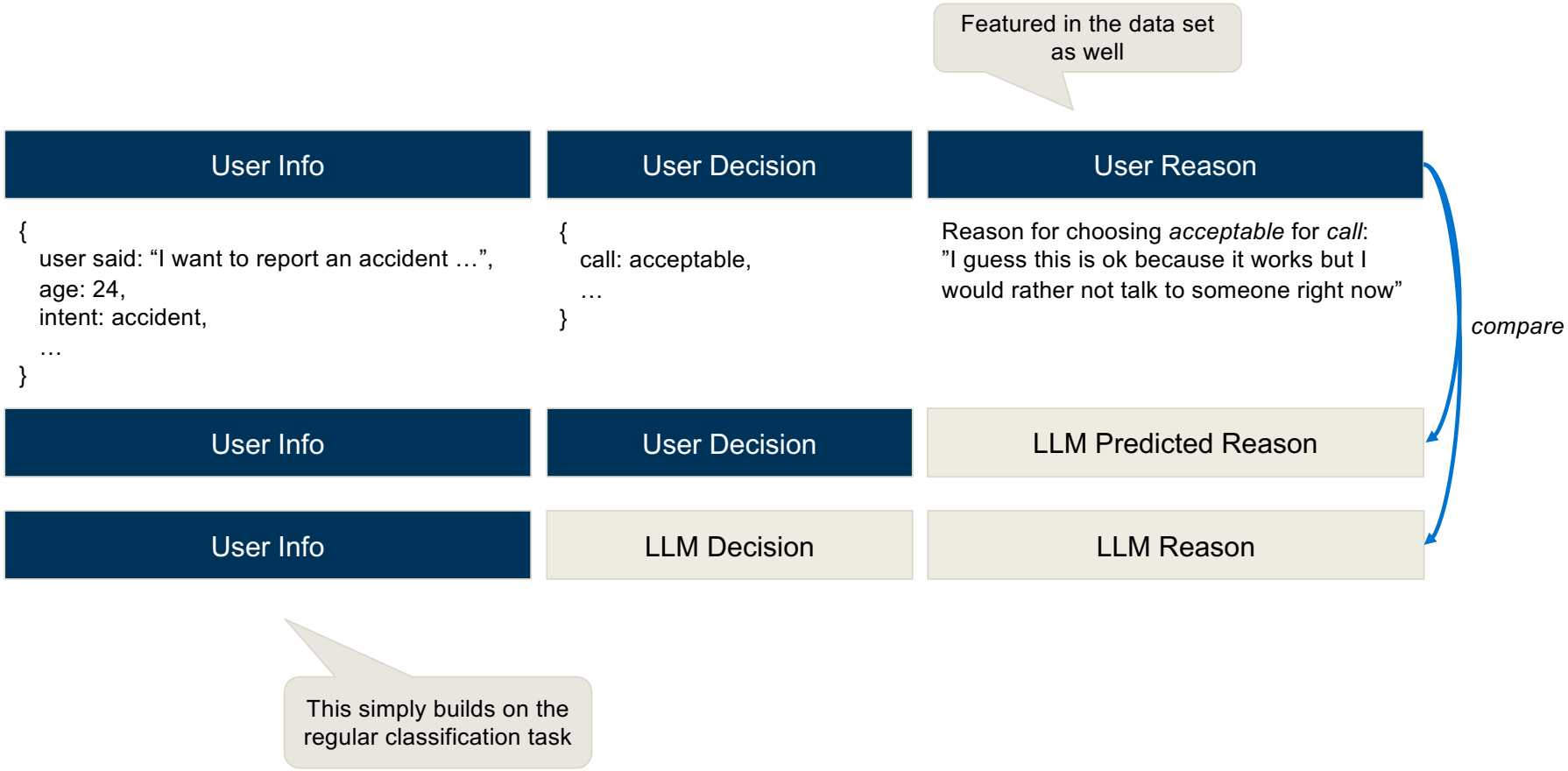
Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

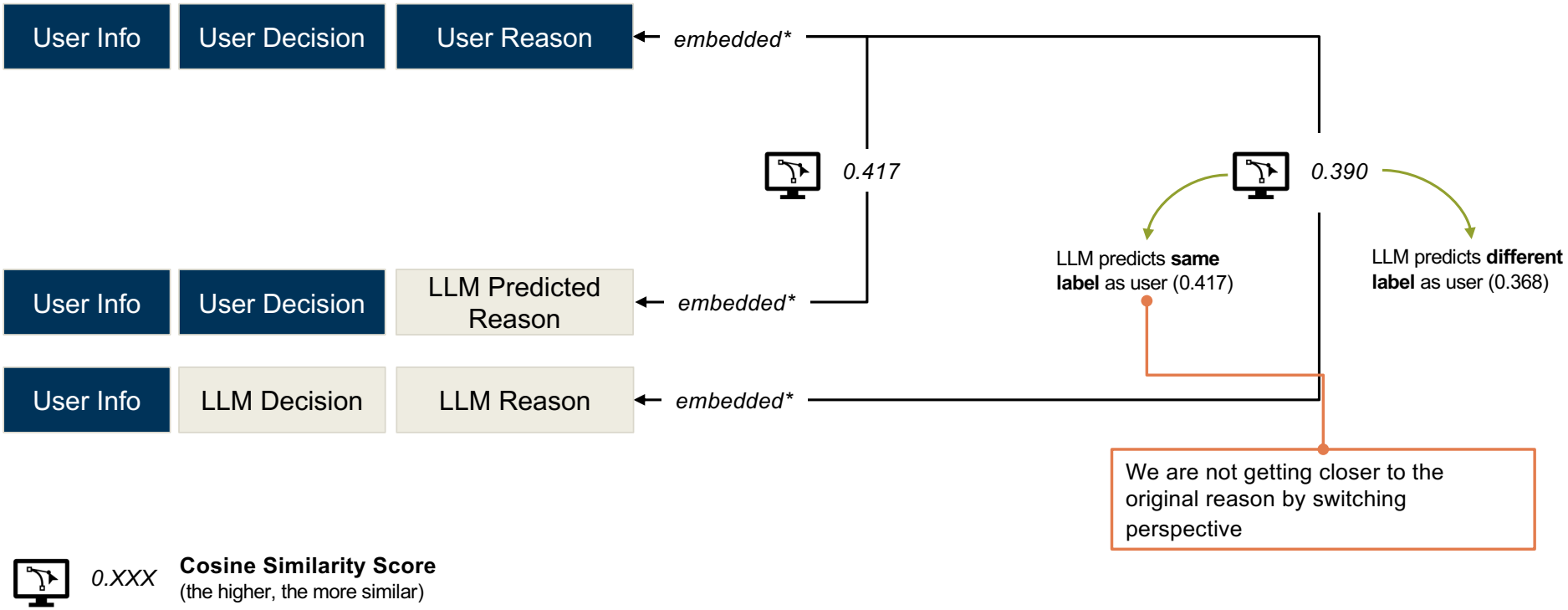
Can the LLM predict the reasons behind the decision?

Masking different information to see how well the LLM predicts the reasons behind a channel preference



When making same decision, LLM for itself and LLM relating with user match

Cosine based similarity measure on the embedded reasons



*Embedding model used is *text-embedding-3-large* via Azure OpenAI

Can the LLM predict the reasons behind the decision?

Masking different information to see how well the LLM predicts the reasons behind a channel preference

Top 10

entries based on similarity –
Exemplary shared concepts

- “Don’t work well with apps”
- Call takes too much time
- Prefer direct contact / direct issue resolution (via call)
- Prefer calling in general

Bottom 10

entries based on similarity –
Exemplary differences (empty fields excluded)

- User sentences very short (ok, no, no idea) while LLM produced longer sentences
- LLM provided more detailed reasons (“I like chat because there is no queue…”)

Intelligent Channel Navigation in Customer Service using LLMs

Outline

- Motivation
- Research Questions
- Channel Choice Determinants
- Data Collection
- LLM-based Channel Navigation
- Prediction of Reasons
- Conclusion

Summary

Reflection on the initial research questions

RQ1: In customer service centers, what are **relevant factors** for deciding the optimal channel for customer service requests?

- ~70 factors identified in literature and interviews
- **Age, Intent, Previous Channels, Infrastructure, Complexity, Innovativeness, Previous Claim** deemed most appropriate for this project

RQ2: How do different **input factors** and **prompt strategies influence the effectiveness of LLMs** in selecting appropriate communication channels for customer service requests?

- The more input factors, the better
- Few-shots and Chain-of-Thought boosts performance
- LLM can compete with data-based approaches
- Classifier performance not excellent

RQ3: How well do LLMs **predict the reasons** for choosing a customer service channel?

- Varying degree of similarities of predicted reasons
- High overlap possible for relatable reasons