Learning to Adapt Large Language Models to One-Shot In-Context Intent Classification on Unseen Domains

Joongbo Shin* **Youbin Ahn*** **Seungpil Won** **Stanley Jungkyu Choi**
LG AI Research
{jb.shin,youbin.ahn}@lgresearch.ai

Abstract

In this paper, we explore one-shot in-context intent classification using large language models (LLMs) with the goal of minimizing the effort required to adapt models to unseen domains. To enhance the one-shot in-context learning capabilities of LLMs, we employ in-context tuning, leveraging its cross-domain transferability to unseen domains. To this end, we introduce the IC-collection, a compilation of open-source intent classification datasets from diverse domains, which are meticulously divided into held-in and held-out datasets. Our experiments demonstrate the effectiveness of the proposed method, showing that our model, with only 7B parameters, not only outperforms GPT-4 on intent classification but also achieves state-of-the-art in unseen domains with only one-shot demonstrations. Both our benchmark and model will be made publicly available to advance research in the chatbot systems.

1 Introduction

Building accurate intent classifiers remains a significant challenge for chatbot systems in real-world scenarios. The labor-intensive process of labeling utterances for new and evolving intents complicates the development and maintenance of chatbots across diverse domains (Sung et al., 2023; Li and Zhang, 2021). In this study, we aim to minimize the effort required to adapt intent classification (IC) models to unseen domains and intents.

To this end, in-context learning, which leverages large language models (LLMs) to achieve high performance on various tasks with only a few input-output pairs, presents a promising direction (Brown et al., 2020; Loukas et al., 2023). Recent research has demonstrated that prompting LLMs only with few-shot examples of text-label pairs can outperform fine-tuned models (Milios et al., 2023). Despite these advancements, they primarily focus on few-shot settings, where five or more examples are required per intent. We argue that previous approaches do not sufficiently minimize the effort required to deploy intent classifiers rapidly across various domains.

This leads us to a research question: *Can we push the limits of in-context learning ability to perform* ***one-shot in-context intent classification****?* To address this, we propose adopting in-context tuning (Min et al., 2022; Chen et al., 2022) for training on seen domains to enhance in-context learning ability on unseen domains. We leverage the cross-task transferability of in-context tuning to improve cross-domain transferability.

To this end, we first construct IC-collection, a benchmark designed for training a model in intent classification across diverse domains and evaluating performance on unseen domains. The IC-collection is a mixture of open-source intent classification datasets, encompassing 13 held-in and 3 held-out datasets[1]. After that, we present OSIC2-7B, where OSIC2 stands for One-Shot In-Context Intent Classification, by training a 7B language model on our benchmark. Our results demonstrate that OSIC2-7B achieves state-of-the-art (SOTA) in unseen domains with only one-shot demonstration, even surpassing GPT-4 in the same setting.

Our contributions are as follows:

- We introduce IC-collection, a training and evaluation benchmark covering diverse domains, specially designed for intent classification on unseen domains.
- We show that OSIC2-7B achieves SOTA performances and even outperforms GPT-4 on unseen domains, highlighting that the effectiveness of our approach.
- To advance research on chatbot systems, we

* indicates equal contribution.

[1] Held-out datasets are carefully selected to minimize domain overlaps.

Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U), pages 182–197
November 16, 2024

Feature	Held-in	Held-out
# of datasets	13	3
# of domains	53 (48)	3 (3)
# of intents	748 (734)	109 (104)

Table 1: Statistics of IC-collection. Unique number of domains and intents are denoted in parentheses.

make our data collection and model publicly available.

2 Data Collection

We compile the IC-collection from diverse open-source resources to cover various domains. During the dataset collection process, several considerations are made: (i) non-english utterances are excluded from multilingual datasets, (ii) only the initial turns from multi-turn interaction dataset are utilized, (iii) duplicate utterances within the dataset were removed, and (iv) multi-labeled utterances are excluded.

In order to evaluate generalization capabilities on unseen domains, the datasets are divided into held-in and held-out sets[2]. Held-out datasets were selected based on their minimal overlap with the held-in datasets, both in terms of domains and intents[3]. Through preliminary experiments conducted to assess the overlap between the held-out and held-in datasets, we confirmed that the impact of the overlap is negligible (see details in Table 6). Table 1 shows the statistics of IC-collection, and followings are the list of datasets:

Held-in datasets The open-sourced IC datasets utilized for training in this study include ACID (Acharya and Fung, 2020), ATIS (Hemphill et al., 1990), BANKING77 (Casanueva et al., 2020), BITEXT[4], CLINC150 (Larson et al., 2019), GENISYS[5], HWU64 (Liu et al., 2021), MCID (Arora et al., 2020a), PRESTO (Goel et al., 2023), SMALLTALK[6], SNIPS (Coucke et al., 2018), SNIPSBI (Coucke et al., 2018), and TOPv2 (Chen et al., 2020).

[2] The original test splits of all datasets are not used for training or in-context demonstrations.

[3] Only highly common intents like `thanks` show any overlap.

[4] https://www.kaggle.com/datasets/scodepy/customer-support-intent-dataset

[5] https://www.kaggle.com/datasets/elvinagammed/chatbots-intent-recognition-dataset

[6] https://www.kaggle.com/datasets/salmanfaroz/small-talk-intent-classification-data

Held-out datasets The held-out datasets include CUREKART, POWERPLAY11, SOFMATTRESS from HINT3 (Arora et al., 2020b).

More details about IC-collection such as the entire list of domains and intents can be found in Appendix D.

3 Our Approach

Task Definition Given an utterance, one-shot in-context intent classification (OSIC2) identifies the correct intent from a list of provided intents, using only one example per intent as a reference. The following is the formulation of OSIC2:

$$f(g(\bar{x}_d, \mathcal{I}_d)) \rightarrow \bar{y}_d,$$

where $f(\cdot)$ is a function for mapping the target utterance $\bar{x}_d$ to intent label $\bar{y}_d$ using a natural language prompt $g(\cdot)$. $\mathcal{I}_d = \{(x_{d_i}, y_{d_i}) | i = 1, ..., C_d\}$ represents the in-context one-shot demonstrations sampled from the train split of each dataset $d \in \mathcal{D}$, and (x_{d_i}, y_{d_i}) denotes an utterance-label pair of i'th index of total classes C_d.

Training Our ultimate goal is to train a model that can effectively leverage one-shot examples at test time, enabling it to adapt proficiently to new domains. To this end, we employ the in-context tuning framework following Min et al. (2022); Chen et al. (2022). While the original concept of in-context tuning is developed for *cross-task* transfer learning, we apply this approach to *cross-domain* transfer learning within the intent classification task.

Prompt Construction As illustrated in Appendix A, we concatenate the task instruction, the in-context examples, and the target utterance into a single input sequence. Also, we use a fixed prompt format for IC-collection to focus on the changes in the intent list along with the corresponding one-shot examples.

Specifically, we first construct the training pool by randomly selecting maximum of ten examples[7] for each intent from the held-in datasets in order to ensure a balance for each intent. From this training pool, we randomly draw a one-shot example for each instance, based on the following assumptions: 1) The dynamic selection of one-shot examples improves the adaptability of the model to changes in the in-context examples. 2) This can partially resolve problems of different granularity

[7] Maximum of ten examples is determined as the optimal number of examples per intent for training as seen in Table 8

Model	CUREKART	POWERPLAY11	SOFMATRESS	Average
LLaMA-2-7B-chat	40.31	47.25	57.71	48.42
LLaMA-3-8B-instruct	67.10	56.96	75.10	64.20
Mistral-7B-instruct-v0.1	51.85	56.31	67.19	58.45
Mistral-7B-instruct-v0.2	73.42	60.84	83.40	72.55
GPT-3.5-turbo	80.39	57.93	84.98	74.43
GPT-4-turbo	**85.62**	65.70	85.77	79.03
OSIC2-7B (avg.)	83.15	**67.31**	**88.54**	**79.67**
OSIC2-7B (best)	**86.27**	**68.93**	**88.54**	**81.25**
previous SOTA (in-domain)	88.05	66.54	78.78	-

Table 2: Accuracies for OSIC2 on unseen domains. **Bolds** denote top-2 results among LLM baselines.

of intents (Huang et al., 2024) and labeling noise in training instances (Ying and Thomas, 2022). When inferring on held-out datasets, we utilize a fixed-representative utterances sourced from the original training datasets as in-context examples. Note that the intent list at the instance level is confined within each dataset.

Preprocessing For all datasets, minimal processing is applied to ensure data quality. The pattern of intent name is standardized with the steps: (i) converting all letters to lowercase, (ii) joining words with underscores, and (iii) writing out abbreviations. Furthermore, intents categorized as out-of-scope (*e.g.*, `no_nodes_detected`) are excluded across all datasets due to their differing levels of granularity compared to other defined intents.

4 Experiments

4.1 Experimental Setup

Implementation Details For our experiments, we utilize Mistral-7b-v0.1 as a backbone. The models are fine-tuned using AdamW with a learning rate of 6×10^{-7}, regulated by a cosine scheduler. We set a batch size of 128.

To confirm the efficacy of the training, the training process is repeated three times using different seed data, which altered the composition of randomly sampled examples. We report both the averaged and best results obtained from these models, evaluated on the same test set.

Baselines We evaluate six SOTA LLMs: LLaMA2-7B-Chat (Touvron et al., 2023), LLaMA3-8B-Instruct[8] (AI@Meta, 2024), Mistral-7B-Instruct-v0.1, and Mistral-7B-Instruct-v0.2 (Jiang et al., 2023) as our open-source LLM baselines, and GPT-3.5-turbo (Ouyang et al., 2022) and GPT-4-turbo (Achiam et al., 2023) as our closed-source LLM baselines. In addition, we also include previous SOTA results taken from Vishwanathan et al. (2022) for comparison.

4.2 Results on Unseen Domains

For the main experiment, we compare our model with baselines on unseen domains, as shown in Table 2. Our model consistently and significantly outperforms open-source LLMs of similar size.We hypothesize that this gap arises because general-purpose LLMs are not specialized for the intent classification task, thereby highlighting the necessity for task-specific fine-tuning. Moreover, our best model outperforms GPT-4 across all unseen domains and even outperforms the previous in-domain SOTA on two out of three datasets, despite the previous SOTA being trained using all available training data for each dataset. This demonstration of generalization to unseen domains using only 7B-scale model underscores the effectiveness of our approach.

4.2.1 Ablation Study

The key finidngs from our empirical ablation studies on generalization capabilites to unseen domains are summarized below. Details of each study can be found in Appendix C.

- Increasing the diversity in terms of intents and domains of held-in datasets effectively enhances generalization (Table 7).
- Increasing the maximum number of training instances per intent does not always prove helpful for generalization (Table 8).
- Pre-processing techniques applied to training data, the standardization of intent formats, the

[8]For LLaMA-3, we follow the choice of prompt in the official homepage: https://llama.meta.com/docs/model-cards-and-prompt-formats/meta-llama-3/.

Method	CURE.	POWER.	SOFM.
OSIC2-7B	86.27	68.93	88.54
OSIC2-7B+	89.98	72.82	87.35
OSIC2-7B (aug)	90.41	73.79	89.72

Table 3: Further analysis on unseen domains. OSIC2-7B+ denotes the upper-bound model and (aug) denotes our remedy for low-recall intents during inference.

Method	C150	H64	B77
OSIC2-7B	95.93	89.78	84.42
7B SOTA (5-shot)	95.35	87.17	85.91
7B SOTA (10-shot)	96.02	90.33	89.48

Table 4: Results on the seen domains, compared with the previous 7B SOTA models, derived from the Llama-2-7B 4K as reported by Milios et al. (2023).

incorporation of various one-shot demonstrations, and the exclusion of out-of-scope intents during training can all contribute to enhanced generalization (Table 9).

4.2.2 Further Analysis

We conduct further analysis on held-out datasets using the best performing OSIC2-7B as shown in Table 3. To establish the upper bound of model performance on IC-collection, we train another model, OSIC2-7B+, by incorporating three benchmarks into the held-in datasets. As expected, OSIC2-7B+ performs better overall than OSIC2-7B, suggesting that additional generalization to unseen domains is possible through training on a wider variety of intents. Meanwhile, OSIC2-7B surpasses OSIC2-7B+ on the SOFMATTRESS dataset, underscoring the robustness and efficacy of our approach.

Furthermore, while beyond our primary scope, we explore the potential improvement achievable by modifying the number of examples for each intent during inference. To this end, we augment the prompt with up to a maximum of two additional examples for intents identified as having low recall, referred as OSIC2-7B (aug) (prompts can be found in Table 13, 14, and 15). The addition of these extra examples during inference enhances performance, surpassing even that of OSIC2-7B+ which we initially assumed to represent the upper bound. This observation suggests that OSIC2-7B not only enables the development of high-performance intent classifiers for new domains but also facilitates the easy maintenance of existing intent classifiers through simple prompt modifications.

4.3 Results on Seen Domains

For the performance on held-in datasets, OSIC2-7B (best) is compared with the previous SOTA method[9] across three representative benchmarks, CLINC150 (C150), HWU64 (H64), and BANKING77 (B77), as shown in Table 4. The performance of OSIC2-7B is comparable to that of the similarly sized SOTA models, except for BANKING77[10], demonstrating that our approach does not sacrifice performance on held-in domains for the sake of generalization to held-out domains.

[9]The SOTA model utilizes a retriever that selects the most relevant utterance-label pairs from the example pool to design the prompt for in-context learning. This model achieves SOTA performance in both 5-shot and 10-shot.

5 Related Work

Language model prompting, particularly with instruction-following LLMs has proven effective for zero-shot or few-shot intent classification (Wei et al., 2022; Sanh et al., 2022). Recent studies have demonstrated the effectiveness of few-shot in-context learning (ICL) for intent classification (Loukas et al., 2023) and emotion classification (Milios et al., 2023). Our approach similarly employs ICL technique to adapt open-LLMs to unseen domains, differing by extending its application to one-shot classification, an extreme scenario where little labeling effort is required.

Additionally, methodologies such as in-context tuning have proposed a meta-learning approach, where k-shot examples are augmented with contextual information for both training and testing (Min et al., 2022; Chen et al., 2022). While these methods primarily focus on transferring learning to unseen tasks through instruction tuning, our objective is to transfer learning to unseen domains.

Methodologically, our approach is most similar to the study on in-context cross-lingual transfer (Villa-Cueva et al., 2024). While that study focuses on text classification and explores cross-lingual transferability, our research emphasizes domain transferability within intent classification.

6 Conclusion

This paper investigates one-shot intent classification as an extreme case of data scarcity in real-world scenarios. By enhancing the in-context learning capabilities of LLMs, our specialized 7B-scale model achieves state-of-the-art performances on

[10]We leave the exploration of this aspect for future work.

held-out datasets, even in the context of unseen domains. To promote research on the creation of accurate intent classifiers that are easily adaptable to any domain, we release our data collection and model.

Limitations & Future Work

Our work is limited in multiple dimensions. First, one example may not be sufficient to represent a definition of the intent. Future work may explore extending our approach to adaptive k-shot in-context intent classification. Second, providing at least one example for all intents in the prompts requires much longer context, resulting in less efficiency at the inference stage. Reducing intent list corresponding to the given utterance with an off-the-shelf retriever needs to be studied in the future. Third, we only consider 7-billion parameter LLMs. Applying our approach on larger LLMs may achieves better generalization on unseen domains. At last, our IC-collection contains English-only datasets. Cross-lingual transferability of intent classification is in our future work.

Acknowledgments

References

Shailesh Acharya and Glenn Fung. 2020. Using optimal embeddings to learn new intents with few examples: An application in the insurance domain. In *Converse@ KDD*.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

AI@Meta. 2024. Llama 3 model card.

Abhinav Arora, Akshat Shrivastava, Mrinal Mohit, Lorena Sainz-Maza Lecanda, and Ahmed Aly. 2020a. Cross-lingual transfer learning for intent detection of covid-19 utterances.

Gaurav Arora, Chirag Jain, Manas Chaturvedi, and Krupal Modi. 2020b. HINT3: Raising the bar for intent detection in the wild. In *Proceedings of the First Workshop on Insights from Negative Results in NLP*, pages 100–105, Online. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Iñigo Casanueva, Tadas Temčinas, Daniela Gerz, Matthew Henderson, and Ivan Vulić. 2020. Efficient intent detection with dual sentence encoders. *arXiv preprint arXiv:2003.04807*.

Xilun Chen, Asish Ghoshal, Yashar Mehdad, Luke Zettlemoyer, and Sonal Gupta. 2020. Low-resource domain adaptation for compositional task-oriented semantic parsing. *arXiv preprint arXiv:2010.03546*.

Yanda Chen, Ruiqi Zhong, Sheng Zha, George Karypis, and He He. 2022. Meta-learning via language model in-context tuning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 719–730, Dublin, Ireland. Association for Computational Linguistics.

Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, et al. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190*.

Rahul Goel, Waleed Ammar, Aditya Gupta, Siddharth Vashishtha, Motoki Sano, Faiz Surani, Max Chang, HyunJeong Choe, David Greene, Kyle He, et al. 2023. Presto: A multilingual dataset for parsing realistic task-oriented dialogs. *arXiv preprint arXiv:2303.08954*.

Charles T Hemphill, John J Godfrey, and George R Doddington. 1990. The atis spoken language systems pilot corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990*.

Simin Huang, Peijie Huang, Yuhong Xu, Jingzhou Liang, and Jingde Niu. 2024. Exploring label hierarchy in dialogue intent classification. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11511–11515. IEEE.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Stefan Larson, Anish Mahendran, Joseph J Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K Kummerfeld, Kevin Leach, Michael A Laurenzano, Lingjia Tang, et al. 2019. An evaluation dataset for intent classification and out-of-scope prediction. *arXiv preprint arXiv:1909.02027*.

Yue Li and Jiong Zhang. 2021. Semi-supervised meta-learning for cross-domain few-shot intent classification. In *Proceedings of the 1st Workshop on Meta Learning and Its Applications to Natural Language Processing*, pages 67–75, Online. Association for Computational Linguistics.

Xingkun Liu, Arash Eshghi, Pawel Swietojanski, and Verena Rieser. 2021. Benchmarking natural language understanding services for building conversational agents. In *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*, pages 165–183. Springer.

Lefteris Loukas, Ilias Stogiannidis, Prodromos Malakasiotis, and Stavros Vassos. 2023. Breaking the bank with ChatGPT: Few-shot text classification for finance. In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 74–80, Macao. -.

Aristides Milios, Siva Reddy, and Dzmitry Bahdanau. 2023. In-context learning for text classification with many labels. In *Proceedings of the 1st GenBench Workshop on (Benchmarking) Generalisation in NLP*, pages 173–184, Singapore. Association for Computational Linguistics.

Sewon Min, Mike Lewis, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2022. MetaICL: Learning to learn in context. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2791–2809, Seattle, United States. Association for Computational Linguistics.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Victor Sanh, Albert Webson, Colin Raffel, Stephen Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Teven Le Scao, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M Rush. 2022. Multitask prompted training enables zero-shot task generalization. In *International Conference on Learning Representations*.

Mujeen Sung, James Gung, Elman Mansimov, Nikolaos Pappas, Raphael Shu, Salvatore Romeo, Yi Zhang, and Vittorio Castelli. 2023. Pre-training intent-aware encoders for zero-and few-shot intent classification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10433–10442.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Emilio Villa-Cueva, A Pastor López-Monroy, Fernando Sánchez-Vega, and Thamar Solorio. 2024.

Adaptive cross-lingual text classification through in-context one-shot demonstrations. *arXiv preprint arXiv:2404.02452*.

Asha Vishwanathan, Rajeev Warrier, Gautham Vadakkekara Suresh, and Chandra Shekhar Kandpal. 2022. Multi-tenant optimization for few-shot task-oriented FAQ retrieval. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 188–197, Abu Dhabi, UAE. Association for Computational Linguistics.

Jason Wei, Maarten Bosma, Vincent Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V Le. 2022. Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*.

Cecilia Ying and Stephen Thomas. 2022. Label errors in BANKING77. In *Proceedings of the Third Workshop on Insights from Negative Results in NLP*, pages 139–143, Dublin, Ireland. Association for Computational Linguistics.

A Appendix A: Prompt Format

[Instruction]
You are an AI assistant for intent classification. For the user query, you should select the correponding intent name from the intent list. Do not write anything else other than the intent name. The intent list is given below and example queries can be followed by each intent name.

[Intent List]
- **current_location**: "what's the precise coordinates of this place"
- **directions**: "where is starbucks"
- **distance**: "how far is the grand canyon from my current location in phoenix, az"
- **gas_amount**: "whats in my gas tank"
- **gas_type**: "what gas does the car need"
...

[Conversation]
User: i am needing you to tell me how to get to dallas, i am needing you to tell me how to get to dallas, texas, by bus
Assistant: **directions**

Table 5: Example of natural language prompt in IC-collection. *Italics* denotes the *fixed* instruction template, "*" denotes example utterance, and **bolds** denotes **intent_name**.

B Appendix B: Preliminary Experiments

Train	Test	CURE.	POWER.	SOFM.
0-shot	0-shot	68.70	59.76	77.60
0-shot	1-shot	80.90	66.45	83.93
1-shot	1-shot	83.15	67.31	88.54

Table 6: Results on preliminary experiments for assessing the impact of dataset overlap. "1-shot Training & 1-shot Test" denotes OSIC2-7B (avg.).

Table 6 shows accuracies of models trained with and without one-shot demonstrations. From this table, we can conclude that the impact of the overlap between held-in and held-out datasets is negligible. If the overlap had been significant, the performance difference between "0-shot Train & 0-shot Test" versus "1-shot Train & 1-shot Test" would not have been substantial. This indicates that domain adaptation through one-shot learning plays a more critical role than the overlap.

C Appendix C: Ablation Study

# Train sets	CURE.	POWER.	SOFM.
3	76.83	65.37	85.91
6	79.88	65.37	85.37
9	80.01	65.80	87.09
13: OSIC2-7B (avg.)	83.15	67.31	88.54
16: OSIC2-7B+ (avg.)	88.45	71.63	88.14

Table 7: Accuracies depending on the number of open-source IC datasets utilized for training. 3 includes CLINC150, BANKING77 and HWU64 datasets; 6 additionally incorporates datasets ATIS, SNIPS, and TOPv2; 9 further adds datasets ACID, MCID and PRESTO; and 13 indicates that all held-in datasets used, with the addition of four datasets, BITEXT, GENISYS, SMALLTALK, and SNIPSBI; 16 includes all held-out datasets, CUREKART, POWERPLAY11, SOFMATTRESS. Note that as the number of datasets increases, they encompass all previously metnioned datasets.

# Examples	CURE.	POWER.	SOFM.
5	82.14	65.91	86.56
10	83.15	67.31	88.54
15	79.96	67.42	87.48
20	78.94	66.67	87.75

Table 8: Results across varying numbers used to construct the training pool referenced in Section 3's prompt construction. All models are tested using the same evaluation set up and test set, which utilizes the fixed one-shot demonstrations.

Intent List	CURE.	POWER.	SOFM.
OSIC2-7B	83.15	67.31	88.54
- intent stand.	83.37	64.51	84.98
- dynamic dem.	82.79	67.75	87.62
+ OOS	80.32	65.69	86.96

Table 9: An ablation study to evaluate the impact of excluding each pre-processing technique applied to the training data. Specifically, the model is trained without intent standardization and subsequently tested on intent names that have not undergone under any preprocessing

D Appendix D: Details of Datasets

Dataset	#Intents	#Domains	OOS Intents	License
ACID	175	1	st_general_request	N/A
ATIS	17	1	-	N/A
BANKING77	77	1	-	CC-BY-4.0
BITEXT	27	11	-	N/A
CLINC150	150	10	-	CC-BY-3.0 Legal Code
GENISYS	19	1	-	N/A
HWU64	64	18	general_quirky	CC BY-SA 3.0
MCID	16	1	-	CC BY-NC-SA 4.0
PRESTO	33	1	other	CC-BY-4.0
SMALLTALK	84	1	-	CC0: Public Domain
SNIPS	7	1	-	CC0-1.0 license
SNIPSBI	10	1	-	CC0-1.0 license
TOPv2	68	8	unsupported_*	CC BY-NC 4.0

Table 10: Statistics of training datasets in IC-collection. unsupported_* denotes in-domain OOS intents for each domain excluding the reminder domain of the TOPv2 dataset.

Dataset	#Intents	Seen Intents	OOS Intents	License
CUREKART	29	cancel_order, order_status	no_nodes_detected	Open Database License
POWERPLAY11	58	thanks	no_nodes_detected	Open Database License
SOFMATTRESS	22	cancel_order, order_status	no_nodes_detected	Open Database License

Table 11: Statistics of heldout datasets in IC-collection.

Domain list of training datasets
account, airline_travel_information_system_(atis), alarm, audio, automobile_and_commute, banking, calendar, cancellation_fee, contact, cooking, covid-19, credit_cards, datetime, delivery, email, event, feedback, food_takeaway, general, insurance, internet_of_things_(iot), invoice, kitchen_and_dining, lists, messaging, meta, music, navigation, news, newsletter, order, payment, play, question_answering_(qa), recommendation, refund, reminder, shipping_address, smalltalk, smarthome, social, timer, transport, travel, utility, virtual_assistant, weather, work
Domain list of held-out datasets
fitness_supplements_retail, online_gaming, mattress_products_retail

Table 12: Domain list of IC-collection.

Intent list of ACID

<domain: insurance> info_add_house, info_add_remove_insuranceured, info_add_remove_vehicle, info_add_vehicle_property_paperless_billing, info_agent_not_responding, info_agent_wrong, info_all_terrain_vehicle_(atv)_insurance_explain, info_american_star, info_amount_due, info_ask_purchase, info_ask_quote, info_automatic_payment_cancel, info_automatic_payment_min_balance, info_automatic_payment_schedule, info_automobile_coverage_question, info_automobile_insurance_canada, info_automobile_policy_cannot_see_in_account, info_bill_due_date, info_billing_account_name_edit, info_billing_account_number, info_billing_department_contact, info_boat_coverage_explain, info_business_policy_cannot_see, info_business_private_policy_(bpp)_question_general, info_cancel_confirm, info_cancel_fee, info_cancel_insurance_policy, info_cannot_see_farm_ranch_policy, info_cannot_see_policy, info_careers, info_change_agent, info_change_autopay_date, info_change_bank_account, info_change_userid, info_claim_adjuster_information, info_claim_check_status, info_claim_complaint, info_claim_direct_repair_program_(drp)_assign, info_claim_direct_repair_program_(drp)_join, info_claim_documents_email, info_claim_documents_fax, info_claim_documents_mail, info_claim_documents_send, info_claim_file_claim, info_claim_filed, info_claim_first_notice_of_loss_(fnol), info_claim_first_notice_of_loss_(fnol)_automobile_hail, info_claim_glass_safelite, info_claim_home_repair_program_(hrp)_join, info_claim_rental, info_claim_shop_add_work, info_claim_shop_send_estimate, info_claim_status, info_claim_update_information, info_collections, info_collision_coverage_explain, info_combine_payments, info_comprehensive_coverage_explain, info_confirm_coverage, info_credit_card_change_number, info_credit_card_fee, info_customer_service_hours, info_declaration_page_needed, info_deductible, info_deductible_explain, info_delete_duplicate_payment, info_different_amounts, info_discounts, info_do_not_contact, info_dreamkeep_rewards, info_dreamkeep_rewards_errors, info_dreams_foundation, info_emergency_roadside_service_(ers), info_emergency_roadside_service_(ers)_contact, info_emergency_roadside_service_(ers)_reimburse, info_employment_verify, info_financial_responsibility_filling_(sr22), info_find_agent, info_flood_insurance_explain, info_forgot_email, info_forgot_password, info_forgot_userid, info_general_policy_coverage_question, info_get_a_quote_auto, info_get_a_quote_automobile_non_owner, info_get_a_quote_business_private_policy_(bpp), info_get_a_quote_other, info_get_a_quote_renters, info_get_a_quote_renters_purchase, info_glass_coverage, info_guaranteed_auto_protection_(gap)_coverage, info_handling_fee_remove, info_health_insurance_quote, info_homesite_contact, info_insurance_card_print, info_insurance_card_proof, info_insurance_card_send, info_insurance_not_available, info_knowyourdrive, info_knowyourdrive_device_activate, info_knowyourdrive_device_return, info_knowyourdrive_errors, info_letter_of_experience, info_liability_explain, info_life_beneficiary_change, info_life_cash_out, info_life_increase_coverage, info_life_policy_amount_due, info_life_policy_automatic_payment, info_life_policy_cancel, info_life_policy_cannot_see, info_life_question_general, info_life_refund, info_life_update_contact_information, info_log_in_error, info_log_out, info_mail_payment_address, info_make_payment, info_mexico_automobile_insurance, info_mortgage_co_proof_of_insurance_(poi), info_name_change, info_new_vehicle_grace_period, info_one_time_payment, info_operating_area, info_operating_company, info_paperless_documents_setup, info_paperless_documents_stop, info_paperless_mail, info_pay_life_insurance, info_payment_confirm, info_payment_due_date_change, info_payment_error, info_payment_history, info_payment_not_ontime, info_payment_process_change, info_payment_setup_automatic_payment, info_payment_time, info_phone_number, info_phone_number_international, info_policy_document_needed, info_policy_number, info_policy_transfer_to_rental, info_premium_breakdown, info_prepaid_card_payment, info_profile_section, info_proof_of_insurance_(poi)_old, info_recreational_vehicle_(rv)_insurance_explain, info_refund_check, info_reinstate_insurance_policy, info_renters_coverage_explain, info_rideshare_coverage, info_salvage_vehicle, info_set_up_account, info_speak_to_representative, info_teen_safe_driver_signup, info_the_general_contact, info_transfer_account_balance, info_travel_insurance_explain, info_university_of_washington_(uw)_alumni_discount, info_update_contact_information, info_update_email, info_update_lienholder, info_update_phone_number, info_who_is_my_agent, info_why_was_policy_cancelled, no, st_general_request, st_hello, st_how_is_chatbot_(abby), st_how_old_is_chatbot_(abby), st_is_chatbot_(abby)_real, st_thank_you, st_what_can_chatbot_(abby)_do, st_where_does_chatbot_(abby)_live, yes

Intent list of ATIS

<domain: airline_travel_information_system_(atis)> abbreviation, aircraft, airfare, airline, airport, capacity, city, day_name, distance, flight, flight_no, flight_time, ground_fare, ground_service, meal, quantity, restriction

Intent list of BANKING77
<domain: banking> activate_my_card, age_limit, apple_pay_or_google_pay, atm_support, automatic_top_up, balance_not_updated_after_bank_transfer, balance_not_updated_after_cheque_or_cash_deposit, beneficiary_not_allowed, cancel_transfer, card_about_to_expire, card_acceptance, card_arrival, card_delivery_estimate, card_linking, card_not_working, card_payment_fee_charged, card_payment_not_recognised, card_payment_wrong_exchange_rate, card_swallowed, cash_withdrawal_charge, cash_withdrawal_not_recognised, change_pin, compromised_card, contactless_not_working, country_support, declined_card_payment, declined_cash_withdrawal, declined_transfer, direct_debit_payment_not_recognised, disposable_card_limits, edit_personal_details, exchange_charge, exchange_rate, exchange_via_app, extra_charge_on_statement, failed_transfer, fiat_currency_support, get_disposable_virtual_card, get_physical_card, getting_spare_card, getting_virtual_card, lost_or_stolen_card, lost_or_stolen_phone, order_physical_card, passcode_forgotten, pending_card_payment, pending_cash_withdrawal, pending_top_up, pending_transfer, pin_blocked, receiving_money, refund_not_showing_up, request_refund, reverted_card_payment, supported_cards_and_currencies, terminate_account, top_up_by_bank_transfer_charge, top_up_by_card_charge, top_up_by_cash_or_cheque, top_up_failed, top_up_limits, top_up_reverted, topping_up_by_card, transaction_charged_twice, transfer_fee_charged, transfer_into_account, transfer_not_received_by_recipient, transfer_timing, unable_to_verify_identity, verify_my_identity, verify_source_of_funds, verify_top_up, virtual_card_not_working, visa_or_mastercard, why_verify_identity, wrong_amount_of_cash_received, wrong_exchange_rate_for_cash_withdrawal

Intent list of BITEXT
<domain: account> create_account, delete_account, edit_account, recover_password, registration_problems, switch_account
<domain: cancellation_fee> check_cancellation_fee
<domain: contact> contact_customer_service, contact_human_agent
<domain: delivery> delivery_options, delivery_period
<domain: feedback> complaint, review
<domain: invoice> check_invoice, get_invoice
<domain: newsletter> newsletter_subscription
<domain: order> cancel_order, change_order, place_order, track_order
<domain: payment> check_payment_methods, payment_issue
<domain: refund> check_refund_policy, get_refund, track_refund
<domain: shipping_address> change_shipping_address, set_up_shipping_address

Intent list of CLINC150

<domain: automobile_and_commute> current_location, directions, distance, gas_amount, gas_type, jump_start, last_maintenance, miles_per_gallon_(mpg), oil_change_how, oil_change_when, schedule_maintenance, tire_change, tire_pressure, traffic, uber

<domain: banking> account_blocked, balance, bill_balance, bill_due, freeze_account, interest_rate, min_payment, order_checks, pay_bill, pin_change, report_fraud, routing, spending_history, transactions, transfer

<domain: credit_cards> annual_percentage_rate_(apr), application_status, card_declined, credit_limit, credit_limit_change, credit_score, damaged_card, expiration_date, improve_credit_score, international_fees, new_card, redeem_rewards, replacement_card_duration, report_lost_card, rewards_balance

<domain: home> calendar_status, calendar_update, next_song, order_status, order_update, play_music, reminder_status, reminder_update, shopping_list_status, shopping_list_update, smart_home_devices, todo_list_status, todo_list_update, update_playlist, what_song

<domain: kitchen_and_dining> accept_reservations, calories, cancel_reservation, confirm_reservation, cook_time, food_last, how_busy, ingredient_substitution, ingredients_list, meal_suggestion, nutrition_info, recipe, restaurant_reservation, restaurant_reviews, restaurant_suggestion

<domain: meta> cancel, change_accent, change_ai_name, change_language, change_speed, change_user_name, change_volume, maybe, no, repeat, reset_settings, sync_device, user_name, whisper_mode, yes

<domain: small_talk> are_you_a_bot, do_you_have_pets, fun_fact, goodbye, greeting, how_old_are_you, meaning_of_life, tell_joke, thank_you, what_are_your_hobbies, what_can_i_ask_you, what_is_your_name, where_are_you_from, who_do_you_work_for, who_made_you

<domain: travel> book_flight, book_hotel, car_rental, carry_on, exchange_rate, flight_status, international_visa, lost_luggage, plug_type, timezone, translate, travel_alert, travel_notification, travel_suggestion, vaccines

<domain: utility> alarm, calculator, date, definition, find_phone, flip_coin, make_call, measurement_conversion, roll_dice, share_location, spelling, text, time, timer, weather

<domain: work> direct_deposit, income, insurance, insurance_change, meeting_schedule, next_holiday, paid_time_off_(pto)_balance, paid_time_off_(pto)_request, paid_time_off_(pto)_status, paid_time_off_(pto)_used, payday, rollover_retirement_savings_plan_(401k), schedule_meeting, taxes, wage_and_tax_statement_(w2)

Intent list of GENISYS

<domain: ai_assistant> clever, courtesy_good_bye, courtesy_greeting, current_human_query, good_bye, gossip, greeting, jokes, name_query, not_talking_to_you, pod_bay_door, real_name_query, self_aware, shutup, swearing, thanks, time_query, understand_query, who_am_i

Intent list of HWU64

<domain: alarm> alarm_query, alarm_remove, alarm_set

<domain: audio> audio_volume_down, audio_volume_mute, audio_volume_up

<domain: calendar> calendar_query, calendar_remove, calendar_set

<domain: cooking> cooking_recipe

<domain: datetime> datetime_convert, datetime_query

<domain: email> email_add_contact, email_query, email_query_contact, email_sendemail

<domain: food_takeaway> takeaway_order, takeaway_query

<domain: general> general_affirm, general_command_stop, general_confirm, general_dontcare, general_explain, general_joke, general_negate, general_praise, general_quirky, general_repeat

<domain: internet_of_things_(iot)> iot_cleaning, iot_coffee, iot_hue_light_change, iot_hue_light_dim, iot_hue_light_off, iot_hue_light_on, iot_hue_light_up, iot_wemo_plug_off, iot_wemo_plug_on

<domain: lists> lists_create_or_add, lists_query, lists_remove

<domain: music> music_likeness, music_query, music_settings

<domain: news> news_query

<domain: play> play_audiobook, play_game, play_music, play_podcasts, play_radio

<domain: question_answering_(qa)> qa_currency, qa_definition, qa_factoid, qa_maths, qa_stock

<domain: recommendation> recommendation_events, recommendation_locations, recommendation_movies

<domain: social> social_post, social_query

<domain: transport> transport_query, transport_taxi, transport_ticket, transport_traffic

<domain: weather> weather_query

Intent list of MCID

<domain: covid-19> can_i_get_from_feces_animal_pets, can_i_get_from_packages_surfaces, donate, hi, how_does_corona_spread, latest_numbers, myths, news_and_press, okay_thanks, protect_yourself, share, travel, what_are_symptoms, what_are_treatment_options, what_if_i_visited_high_risk_area, what_is_corona

Intent list of PRESTO

<domain: virtual_assistant> add_contact, add_item_to_list, buy_event_tickets, cancel_ride, check_order_status, create_list, create_note, find_parking, get_bill, get_generic_business_type, get_health_stats, get_list, get_message_content, get_note, get_product, get_security_price, initiate_call, log_exercise, log_nutrition, open_app, order_menu_item, order_ride, other, pause_exercise, pay_bill, play_game, post_message, record_video, resume_exercise, send_digital_object, start_exercise, stop_exercise, take_photo

Intent list of SMALLTALK

<domain: smalltalk> agent_acquaintance, agent_age, agent_annoying, agent_answer_my_question, agent_bad, agent_be_clever, agent_beautiful, agent_birth_date, agent_boring, agent_boss, agent_busy, agent_chatbot, agent_clever, agent_crazy, agent_fired, agent_funny, agent_good, agent_happy, agent_hungry, agent_marry_user, agent_my_friend, agent_occupation, agent_origin, agent_ready, agent_real, agent_residence, agent_right, agent_sure, agent_talk_to_me, agent_there, appraisal_bad, appraisal_good, appraisal_no_problem, appraisal_thank_you, appraisal_welcome, appraisal_well_done, confirmation_cancel, confirmation_no, confirmation_yes, dialog_hold_on, dialog_hug, dialog_i_do_not_care, dialog_sorry, dialog_what_do_you_mean, dialog_wrong, emotions_ha_ha, emotions_wow, greetings_bye, greetings_goodevening, greetings_goodmorning, greetings_goodnight, greetings_hello, greetings_how_are_you, greetings_nice_to_meet_you, greetings_nice_to_see_you, greetings_nice_to_talk_to_you, greetings_whatsup, user_angry, user_back, user_bored, user_busy, user_can_not_sleep, user_does_not_want_to_talk, user_excited, user_going_to_bed, user_good, user_happy, user_has_birthday, user_here, user_joking, user_likes_agent, user_lonely, user_looks_like, user_loves_agent, user_misses_agent, user_needs_advice, user_sad, user_sleepy, user_testing_agent, user_tired, user_waits, user_wants_to_see_agent_again, user_wants_to_talk, user_will_be_back

Intent list of SNIPS

<domain: smart_home> add_to_playlist, book_restaurant, get_weather, play_music, rate_book, search_creative_work, search_screening_event

Intent list of SNIPSBI
<domain: smarthome> book_restaurant, compare_places, get_directions, get_place_details, get_traffic_information, get_weather, request_ride, search_place, share_current_location, share_estimated_time_of_arrival_(eta)

Intent list of TOPv2
<domain: alarm> create_alarm, delete_alarm, get_alarm, silence_alarm, snooze_alarm, unsupported_alarm, update_alarm
<domain: event> get_event, get_event_attendee, get_event_attendee_amount, get_event_organizer, unsupported_event
<domain: messaging> cancel_message, get_message, ignore_message, react_message, send_message, unsupported_messaging
<domain: music> add_to_playlist_music, create_playlist_music, dislike_music, like_music, loop_music, pause_music, play_music, previous_track_music, remove_from_playlist_music, replay_music, set_default_provider_music, skip_track_music, start_shuffle_music, stop_music, unsupported_music
<domain: navigation> get_directions, get_distance, get_estimated_arrival, get_estimated_departure, get_estimated_duration, get_info_road_condition, get_info_route, get_info_traffic, get_location, unsupported_navigation, update_directions
<domain: reminder> create_reminder, delete_reminder, get_reminder, get_reminder_amount, get_reminder_date_time, get_reminder_location, help_reminder, update_reminder, update_reminder_date_time, update_reminder_todo
<domain: timer> add_time_timer, create_timer, delete_timer, get_timer, pause_timer, restart_timer, resume_timer, subtract_time_timer, unsupported_timer, update_timer
<domain: weather> get_sunrise, get_sunset, get_weather, unsupported_weather

Intent list of CUREKART
<domain: fitness_supplements_retail> call_center, cancel_order, chat_with_agent, check_pincode, consult_start, delay_in_parcel, expiry_date, franchise, immunity, international_shipping, modes_of_payments, modify_address, no_nodes_detected, order_query, order_status, order_taking, original_product, payment_and_bill, portal_issue, recommend_product, refer_earn, refunds_returns_replacements, resume_delivery, side_effect, sign_up, start_over, store_information, user_goal_form, work_from_home

Intent list of POWERPLAY11
<domain: online_gaming> account_balance_deducted, account_not_verified, account_reset, appreciation, bank_verification_details, cannot_see_joined_contests, capabilities, cash_bonus, cash_bonus_expiry, change_bank_account, change_mobile_number, change_profile_team_details, chat_with_an_agent, check_deposit_status, check_wallet_balance, contact_number, criticism, deducted_amount_not_received, delete_pan_card, download_powerplay11, fairplay_violations, fake_teams, feedback, greetings_day, how_points_calculated, how_to_play, instant_withdrawal, join_contest, less_winnings_amount, match_abandoned, new_team_pattern, no_email_confirmation, no_nodes_detected, offers_and_referrals, pan_verification_failed, points_not_updated, presence, refund_of_added_cash, refund_of_wrong_amount, signup_bonus, taxes_on_winnings, team_deadline, thanks, types_bonus, types_contests, unutilized_money, update_app, verify_email, verify_mobile, verify_pan, what_if_theres_a_tie, why_verify, winnings, withdraw_cash_bonus, withdrawal_intro, withdrawal_status, withdrawal_time, wrong_scores

Intent list of SOFMATTRESS
<domain: mattress_products_retail> 100_night_trial_offer, about_sof_mattress, cancel_order, cash_on_delivery_(cod), check_pincode, comparison, delay_in_delivery, distributors, equated_monthly_instalment_(emi), ergonomic_features, lead_generation, mattress_cost, no_nodes_detected, offers, order_status, orthopedic_features, pillows, product_variants, return_exchange, size_customization, warranty, what_size_to_order

- **call_center**: "what is the time when call center is working"
- **cancel_order**: "I want to place cancellation"
- **chat_with_agent**: "How to complaint"
- **check_pincode**: "Are you shipping to my pincode"
- **consult_start**: "Get Diet & Fitness Advice"
- **delay_in_parcel**: "I am not received my order yet"
- **expiry_date**: "I have received an Expired product"
- **franchise**: "I would like to get dealership"
- **immunity**: "Increase Immunity"
- **international_shipping**: "Delivery out of India"
- **modes_of_payments**: "ways of paymets"
- **modify_address**: "Edit shipping address"
- **order_query**: "Help required on order"
- **order_status**: "How much more time do I have to wait for my parcel"
- **order_taking**: "I want to book shipment on cash on delivery"
- **original_product**: "Show me your authenticity"
- **payment_and_bill**: "my money has been deducted but it's not been place order"
- **portal_issue**: "My Cart Is empty"
- **recommend_product**: "i need supplements", "energy boost in body", "I want home gym product"
- **refer_earn**: "I have referral promo code"
- **refunds_returns_replacements**: "I want my money back"
- **resume_delivery**: "If i order today in how many days it will be delivered"
- **side_effect**: "Side Effect"
- **sign_up**: "I am a new user"
- **start_over**: "Restart the flow"
- **store_information**: "Are your offline stores open?"
- **user_goal_form**: "Re-assess my profile"
- **work_from_home**: "I hope you are also working from home during this time"

Table 13: Augmented version of intent list on CUREKART evaluation.

- **account_balance_deducted**: "What is cycle of account balance deduction"
- **account_not_verified**: "Account Verification"
- **account_reset**: "How to reset account"
- **appreciation**: "Great App"
- **bank_verification_details**: "What details I need to provide for bank account"
- **cannot_see_joined_contests**: "I joined a league but now it's not showing"
- **capabilities**: "Help me"
- **cash_bonus**: "Cash Bonus"
- **cash_bonus_expiry**: "Cash Bonus Expiry"
- **change_bank_account**: "Change My Bank Account"
- **change_mobile_number**: "I want to change my number"
- **change_profile_team_details**: "Edit team name"
- **chat_with_an_agent**: "Need to connect with an agent", "I can't see my withdrawal", "My bonus is incorrect"
- **check_deposit_status**: "Show my transaction"
- **check_wallet_balance**: "Money left in my wallet"
- **contact_number**: "Call me back"
- **criticism**: "Waste app", "You are dumb"
- **deducted_amount_not_received**: "My money was deducted from my account but not showing the amount added. What should I do?"
- **delete_pan_card**: "Pan card remove"
- **download_powerplay11**: "Download app"
- **fairplay_violations**: "How my play will be consider as fair"
- **fake_teams**: "You have your own team in the leagues", "Fake players"
- **feedback**: "Feedback"
- **greetings_day**: "Yes"
- **how_points_calculated**: "How Are Points Calculated on PowerPlay11"
- **how_to_play**: "I need help to play"
- **instant_withdrawal**: "Fast withdrawal available"
- **join_contest**: "Contest joining"
- **less_winnings_amount**: "My winnings are incorrect"
- **match_abandoned**: "If match get abandoned will I get refund"
- **new_team_pattern**: "How many all-rounder I can select"
- **no_email_confirmation**: "When will I receive email confirmation"
- **offers_and_referrals**: "Any promotions available"
- **pan_verification_failed**: "Getting error while verifying PAN Card"
- **points_not_updated**: "Points are not getting updated", "When will scores be updated"
- **presence**: "Are You Online"
- **refund_of_added_cash**: "Is my addded cash is refundable", "Please refund money"
- **refund_of_wrong_amount**: "I added amount by mistake"
- **signup_bonus**: "Signup Bonus"
- **taxes_on_winnings**: "How much tax will be deducted"
- **team_deadline**: "What is Safe Play & Regular Play"
- **thanks**: "Tysm"
- **types_bonus**: "What is difference between Cash Bonus, signup bonus, surprise bonus, winnings"
- **types_contests**: "Types of Contests"
- **unutilized_money**: "Unutilized Amount"
- **update_app**: "How to update the app"
- **verify_email**: "Email verification"
- **verify_mobile**: "Mobile number verification"
- **verify_pan**: "Pan card verification", "How do I verify my PAN"
- **what_if_theres_a_tie**: "Same score between two players"
- **why_verify**: "What is the use of account verification"
- **winnings**: "When will I get the winning amount", "Winnings amount not credited"
- **withdraw_cash_bonus**: "Withdraw Cash Bonus"
- **withdrawal_intro**: "Withdrawal steps"
- **withdrawal_status**: "Status of withdrawal"
- **withdrawal_time**: "When can I expect my withdrawal amount"
- **wrong_scores**: "What if a game is completed with wrong scores?"

Table 14: Augmented version of intent list on POWERPLAY11 evaluation.

- **100_night_trial_offer**: "100 free Nights"
- **about_sof_mattress**: "How is SOF different from other mattress brands"
- **cancel_order**: "I want to cancel my order"
- **cash_on_delivery_(cod)**: "Can pay later on delivery"
- **check_pincode**: "Can you deliver on my pincode", "Will you be able to deliver here"
- **comparison**: "What is the difference between the Ergo & Ortho variants"
- **delay_in_delivery**: "It's been a month", "I did not receive my order yet"
- **distributors**: "Do you have any showrooms in Delhi state", "Need dealership"
- **equated_monthly_instalment_(emi)**: "You guys provide EMI option?"
- **ergonomic_features**: "What are the key features of the SOF Ergo mattress"
- **lead_generation**: "Get in Touch"
- **mattress_cost**: "Price of mattress", "How Much Cost"
- **offers**: "Any discounts"
- **order_status**: "Order Status", "When will the order be delivered to me?"
- **orthopedic_features**: "Features of Ortho mattress"
- **pillows**: "Do you have cushions"
- **product_variants**: "What are the product variants", "Show more mattress"
- **return_exchange**: "Need my money back"
- **size_customization**: "Can mattress size be customised?"
- **warranty**: "Does mattress cover is included in warranty"
- **what_size_to_order**: "Can you help with the size?"

Table 15: Augmented version of intent list on SOFMATTRESS evaluation.