ChatGPT as Data Augmentation for Compositional Generalization: A Case Study in Open Intent Detection

Yihao Fang[2,3], **Xianzhi Li**[1,2], **Stephen W. Thomas**[3] **and Xiaodan Zhu**[1,2]

[1]Department of Electrical and Computer Engineering, Queen's University
[2]Ingenuity Labs Research Institute, Queen's University
[3]Smith School of Business, Queen's University
{yihao.fang, 21xl17, stephen.thomas, xiaodan.zhu}@queensu.ca

Abstract

Open intent detection, a crucial aspect of natural language understanding, involves the identification of previously unseen intents in user-generated text. Despite the progress made in this field, challenges persist in handling new combinations of language components, which is essential for compositional generalization. In this paper, we present a case study exploring the use of ChatGPT as a data augmentation technique to enhance compositional generalization in open intent detection tasks. We begin by discussing the limitations of existing benchmarks in evaluating this problem, highlighting the need for constructing datasets for addressing compositional generalization in open intent detection tasks. By incorporating synthetic data generated by ChatGPT into the training process, we demonstrate that our approach can effectively improve model performance. Rigorous evaluation of multiple benchmarks reveals that our method outperforms existing techniques and significantly enhances open intent detection capabilities. Our findings underscore the potential of large language models like ChatGPT for data augmentation in natural language understanding tasks.

1 Introduction

Open intent detection, a key component of natural language understanding, aims to identify previously unseen intents in user-generated text. This task is of paramount importance for a wide range of applications, such as conversational AI systems, where the ability to recognize new intents can substantially improve the user experience. Although the field has made significant strides in recent years, a major challenge remains in addressing compositional generalization, which refers to the capability of models to handle unseen combinations of language components. This capability is essential for the successful deployment of AI systems in real-world scenarios, where users may express intent in unforeseen ways.

In this paper, we present a case study that investigates the potential of ChatGPT, a state-of-the-art large language model, as a data augmentation technique for enhancing compositional generalization in open intent detection tasks. Our study begins by identifying the shortcomings of existing benchmarks in evaluating this problem, which underscores the need for the development of datasets tailored to assess compositional generalization in open intent detection tasks.

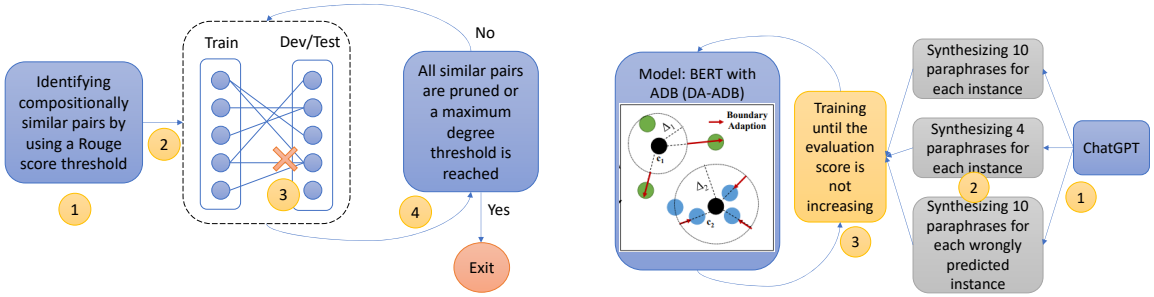
To address this issue, we leverage ChatGPT to generate synthetic data that is then incorporated into the training process. By doing so, we aim to improve the model's ability to recognize new combinations of language components, thereby enhancing its open intent detection capabilities. Through rigorous evaluation of multiple benchmarks, we demonstrate that our proposed method outperforms existing techniques and leads to significant performance improvements.

Our findings highlight the potential of large language models, such as ChatGPT, for data augmentation in natural language understanding tasks. This case study offers valuable insights into the development of more effective dialogue systems capable of handling a wider range of user intents and fostering better human-computer interactions.

Our primary contributions to the literature include:

- Dataset Construction for Compositional Generalization: We construct compositionally diverse subsets derived from existing open intent detection benchmark datasets.
- ChatGPT Data Augmentation: We propose using ChatGPT to generate paraphrases of training dataset instances, thereby enhancing model generalization and performance on unseen compositions.
- We evaluate three different strategies for incorporating ChatGPT-generated paraphrases into

Proceedings of the Joint Workshop of the 5th Financial Technology and Natural Language Processing (FinNLP) and 2nd Multimodal AI For Financial Forecasting (Muffin), Macao, August 20, 2023.



(a) Dataset construction

(b) ChatGPT data augmentation

Figure 1: a-1) Compositionally-similar utterance instances are identified by a Rouge score threshold. a-2) An undirect graph is created by connecting compositionally-similar instances with edges. a-3) Node (instance) degrees are counted in the undirected graph and highest-degree nodes and their edges are first pruned. a-4) This process iterates until all similar pairs are pruned or a maximum degree threshold is reached. b-1) Paraphrases are generated by ChatGPT for utterances in the training dataset. b-2) Paraphrases are incorporated into the BERT training process through three different strategies. b-3) The training process iterates until the evaluation score (accuracy) is not increasing.

the training process of BERT (Devlin et al., 2019) with ADB (Zhang et al., 2021b) (DA-ADB Zhang et al., 2023).

The rest of the paper is organized as follows: Section 2 provides a background on open intent classification and reviews related work. Section 3 describes our proposed method in detail. Section 4 presents the experimental setup, results, and analysis. Finally, Section 5 concludes the paper and suggests directions for future research.

2 Related Work

Open intent classification is an important problem in natural language understanding and dialogue systems, aiming to identify known intents and detect unseen open intents using only the prior knowledge of known intents. Several recent studies have explored various techniques for addressing this challenging task.

One line of research involves aligning representation learning with scoring functions. For instance, the unified neighbourhood learning framework (UniNL) was proposed to detect OOD intents by designing a KNCL objective for representation learning and introducing a KNN-based scoring function for OOD detection (Mou et al., 2022b). Another study proposed a unified K-nearest neighbour contrastive learning framework for OOD intent discovery, which focuses on inter-class discriminative features and alleviates the in-domain overfitting problem (Mou et al., 2022a).

Another direction focuses on learning discriminative representations and decision boundaries for open intent detection. The Deep Open Intent Classification with Adaptive Decision Boundary (**ADB**) method learns an adaptive spherical decision boundary for each known class, balancing both the empirical risk and the open space risk without requiring open intent samples or modifying the model architecture (Zhang et al., 2021b). Similarly, the **DA-ADB** framework successively learns distance-aware intent representations and adaptive decision boundaries for open intent detection by leveraging distance information and designing a loss function to balance empirical and open space risks (Zhang et al., 2023).

In summary, various methods have been proposed to address the challenges associated with detecting unseen intents. However, none of them have explored compositional generalization in open intent detection tasks. We highlight the need for constructing datasets and leverage ChatGPT to generate synthetic data to address this problem. Our proposed method in detail is given in the following section.

3 Methodology

3.1 Dataset Construction for Compositional Generalization

The construction of the dataset starts with identifying compositionally-similar utterance instances by utilizing a Rouge score threshold (Figure 1a). The Rouge score is a widely-used metric for evaluating the similarity between a pair of text sequences by comparing the number of overlapping n-grams (Lin, 2004). By setting a threshold value, instances

with Rouge scores above this threshold are deemed to be compositionally similar, allowing for the effective detection of instances with a high degree of overlap in content or structure.

Once the compositionally-similar utterance instances are identified, an undirected graph is created by connecting these instances with edges. In this graph, each node represents an instance, and an edge is drawn between two nodes if their corresponding instances are compositionally similar according to the Rouge score threshold. This representation allows for a better understanding of the relationships between the instances, making it easier to discern patterns and outliers in the data. Furthermore, the graph-based approach facilitates the efficient pruning of highly similar instances in subsequent steps.

To refine the dataset and ensure maximum diversity, the highest-degree nodes and their connecting edges are first pruned. In this context, the degree of a node refers to the number of edges connected to it. By pruning the highest-degree nodes, the instances with the most similarities to other instances are removed from the dataset. This process iterates until all similar pairs have been pruned or a maximum degree threshold is reached. The result is a dataset with a high degree of diversity and helps to access the compositional generalizability of the model trained on this dataset.

The aforementioned approach is utilized on three open intent detection benchmark datasets: **Banking** (Casanueva et al., 2020), **OOS** (Larson et al., 2019) and **StackOverflow** (Xu et al., 2015), resulting in three compositionally diverse subsets derived from these datasets, namely **Banking_CG**, **OOS_CG**, and **StackOverflow_CG**. (Refer to Appendix A for dataset construction in detail.)

3.2 ChatGPT Data Augmentation

The training process involves generating paraphrases for utterances in the training dataset using ChatGPT (Figure 1b). This paraphrasing approach aids in enhancing the model's understanding of language by providing alternative compositions of the same meaning. The incorporation of these paraphrases into the training process not only improves the generalizability of the model but also leads to better performance on unseen compositions. (Refer to Appendix C for ChatGPT's paraphrases in detail.)

To effectively integrate paraphrases into the training process of BERT (Devlin et al., 2019) with **ADB** (**DA-ADB**), three different strategies are evaluated. The first strategy involves synthesizing 10 paraphrases for each instance in the dataset (**GPTAUG-F10**), while the second strategy generates 4 paraphrases for each instance (**GPTAUG-F4**). The third strategy, on the other hand, focuses on instances that the model predicts incorrectly at the current iteration and synthesizes 10 paraphrases for each of these instances (**GPTAUG-WP10**). This targeted approach aims to help address specific weaknesses in the model's understanding. The training process iterates through these strategies until the evaluation score, such as accuracy, no longer exhibits any improvement. This iterative process ensures that the model continues to refine its understanding of language by learning from the generated paraphrases, ultimately resulting in a more robust and capable BERT model.

4 Experiments

4.1 Experimental Setup

In our experimental setup, we have extended the TEXTOIR platform (Zhang et al., 2021a), a toolkit that integrates a variety of state-of-the-art algorithms for open intent detection, to conduct our experiments. To ensure fair comparisons across all tests, we employed the pre-trained BERT-base model from Hugging Face (Wolf et al., 2020) as the foundation of our approach. The optimization of the BERT model with ADB (DA-ADB) was carried out using Python and the PyTorch framework (Paszke et al., 2019) and executed on NVIDIA RTX 2080 TI GPUs for computational efficiency.

4.2 Results and Analysis

Experimental results (Table 1) show that ADB (Zhang et al., 2021b) and DA-ADB (Zhang et al., 2023) are not robust and exhibit poor performance in the compositionally diverse subsets: Banking_CG, OOS_CG, and StackOverflow_CG. These subsets are derived from more extensive datasets, namely Banking, OOS, and StackOverflow. This indicates that these models struggle to achieve compositional generalization in more challenging contexts.

Interestingly, ADB is found to be more robust than DA-ADB, particularly in the OOS_CG subset, where the model has to predict a larger number of intents (151 intents) in the test phase. This is about twice the number of intents in Banking_CG and 7.5

Table 1: Performance of our ChatGPT augmentation approaches (GPTAUG-F4, GPTAUG-F10, and GPTAUG-WP10) and the baselines (ADB and DA-ADB). The best results among each setting are bolded. All results are an average of 10 runs using 10 different seed numbers considering that the selection of known intents is a pseudo-random process. (Refer to Appendix D in more detail.)

	Methods	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	F1-OOD	F1-All	Acc-All	F1-IND	F1-OOD	F1-All	Acc-All	F1-IND	F1-OOD	F1-All	Acc-All
25%	ADB	53.49	81.10	54.87	72.31	49.13	90.65	50.19	83.96	58.37	79.04	61.82	71.35
	DA-ADB	53.33	**86.15**	54.97	**78.43**	38.27	91.70	39.64	85.45	**62.32**	**84.84**	**66.08**	**77.68**
	ADB+GPTAUG-F4	56.73	83.37	58.06	75.51	53.54	91.93	54.52	86.23	60.94	82.62	64.55	75.32
	ADB+GPTAUG-F10	**57.58**	84.04	**58.90**	76.46	**54.26**	**92.07**	**55.23**	**86.48**	58.99	80.34	62.55	72.66
	ADB+GPTAUG-WP10	50.04	70.47	51.06	61.47	48.03	88.57	49.07	80.83	51.62	63.54	53.60	56.17
	DA-ADB+GPTAUG-F4	54.58	84.82	56.09	77.11	43.98	91.85	45.21	85.89	59.97	78.03	62.98	70.79
	DA-ADB+GPTAUG-F10	53.52	84.00	55.04	76.00	44.20	91.74	45.42	85.70	59.82	75.31	62.40	70.10
	DA-ADB+GPTAUG-WP10	54.72	82.89	56.13	74.38	43.18	91.55	44.42	85.33	54.61	64.87	56.32	59.22
50%	ADB	59.93	69.63	60.18	65.38	52.32	83.99	52.73	75.66	71.45	76.14	71.88	73.58
	DA-ADB	54.57	74.45	55.08	67.77	33.66	83.31	34.31	73.37	**75.97**	**81.75**	**76.49**	**79.14**
	ADB+GPTAUG-F4	**62.55**	73.20	**62.83**	69.24	55.36	85.39	55.76	78.07	71.58	77.03	72.08	74.37
	ADB+GPTAUG-F10	62.28	73.23	62.56	69.36	**55.40**	**85.57**	**55.80**	**78.44**	70.97	77.56	71.57	74.52
	ADB+GPTAUG-WP10	59.87	61.11	59.90	60.27	53.25	83.06	53.64	74.61	67.04	64.13	66.78	65.37
	DA-ADB+GPTAUG-F4	57.06	**74.67**	57.52	**69.41**	38.85	84.00	39.45	74.96	72.28	74.07	72.44	73.88
	DA-ADB+GPTAUG-F10	56.52	74.42	56.98	69.09	39.02	83.94	39.61	74.90	70.32	74.78	70.72	73.27
	DA-ADB+GPTAUG-WP10	58.63	70.33	58.93	65.30	40.26	83.99	40.84	74.81	69.91	64.58	69.43	66.65
75%	ADB	64.30	53.36	64.12	62.82	53.87	76.24	54.07	68.33	76.13	61.56	75.22	71.58
	DA-ADB	54.74	52.46	54.70	56.94	29.58	71.76	29.96	59.91	**78.57**	**65.80**	**77.77**	**74.51**
	ADB+GPTAUG-F4	**66.65**	**54.82**	**66.45**	**64.89**	**55.99**	**77.36**	**56.18**	**70.70**	75.72	61.68	74.84	71.08
	ADB+GPTAUG-F10	66.22	54.61	66.02	64.54	55.64	77.04	55.83	70.54	75.35	61.15	74.46	70.61
	ADB+GPTAUG-WP10	65.22	47.98	64.93	62.22	54.91	75.78	55.10	67.97	73.86	49.92	72.37	67.67
	DA-ADB+GPTAUG-F4	55.87	51.59	55.80	57.67	33.50	72.40	33.85	61.81	76.87	60.79	75.86	71.51
	DA-ADB+GPTAUG-F10	54.86	50.66	54.78	56.65	33.81	72.48	34.15	62.04	73.65	54.81	72.48	68.10
	DA-ADB+GPTAUG-WP10	62.31	53.33	62.15	61.87	40.21	74.07	40.51	64.11	77.77	60.62	76.70	72.52

times that of StackOverflow_CG. However, DA-ADB outperforms ADB in the StackOverflow_CG subset, which is more balanced and has far fewer intents to predict.

In the Banking_CG subset, it was observed that the overall F1 scores of ADB with ChatGPT data augmentation were consistently higher (by about 2 to 4%) than those of ADB and DA-ADB. A similar trend was seen in the OOS_CG subset, where the F1 scores of ADB with ChatGPT data augmentation were 2 to 5% better than ADB and DA-ADB. These results demonstrate that data augmentation can indeed help bridge the gap between the training and test sets, even when they exhibit compositional dissimilarity.

ADB with ChatGPT data augmentation outperforms DA-ADB with augmentation in both Banking_CG and OOS_CG. Interestingly, GPTAUG-WP10, a more sophisticated data augmentation method (which paraphrases wrongly predicted instances), underperforms when compared to simply incorporating all ChatGPT paraphrases into the training process (GPTAUG-F4 and GPTAUG-F10).

Finally, DA-ADB performs best in StackOverflow_CG, considering that this subset is relatively more balanced and has fewer intents to predict.

5 Conclusion

In conclusion, this paper addresses the challenge of compositional generalization in open intent detection by leveraging the capabilities of ChatGPT, a state-of-the-art large language model. By constructing compositionally diverse datasets (i.e., **Banking_CG**, **OOS_CG**, and **StackOverflow_CG**) and incorporating ChatGPT-generated paraphrases into the training process, we have demonstrated large improvements in model performance on unseen compositions.

Future research should focus on developing more advanced data augmentation approaches that can generate more diverse compositions. One possible direction involves designing better-instructed prompts for ChatGPT to encourage more diverse paraphrases that can help improve compositional generalization even further. Additionally, exploring alternative strategies for incorporating augmented data and refining the iterative training process may lead to further performance improvements.

References

Iñigo Casanueva, Tadas Temčinas, Daniela Gerz, Matthew Henderson, and Ivan Vulić. 2020. Efficient intent detection with dual sentence encoders. In *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, pages 38–45, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Stefan Larson, Anish Mahendran, Joseph J. Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K. Kummerfeld, Kevin Leach, Michael A. Laurenzano, Lingjia Tang, and Jason Mars. 2019. An evaluation dataset for intent classification and out-of-scope prediction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1311–1316, Hong Kong, China. Association for Computational Linguistics.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Yutao Mou, Keqing He, Pei Wang, Yanan Wu, Jingang Wang, Wei Wu, and Weiran Xu. 2022a. Watch the neighbors: A unified k-nearest neighbor contrastive learning framework for OOD intent discovery. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1517–1529, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yutao Mou, Pei Wang, Keqing He, Yanan Wu, Jingang Wang, Wei Wu, and Weiran Xu. 2022b. UniNL: Aligning representation learning with scoring function for OOD detection via unified neighborhood learning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7317–7325, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Jiaming Xu, Peng Wang, Guanhua Tian, Bo Xu, Jun Zhao, Fangyuan Wang, and Hongwei Hao. 2015. Short text clustering via convolutional neural networks. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 62–69, Denver, Colorado. Association for Computational Linguistics.

Hanlei Zhang, Xiaoteng Li, Hua Xu, Panpan Zhang, Kang Zhao, and Kai Gao. 2021a. TEXTOIR: An integrated and visualized platform for text open intent recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 167–174.

Hanlei Zhang, Hua Xu, and Ting-En Lin. 2021b. Deep open intent classification with adaptive decision boundary. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14374–14382.

Hanlei Zhang, Hua Xu, Shaojie Zhao, and Qianrui Zhou. 2023. Learning discriminative representations and decision boundaries for open intent detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

A Dataset Construction in Detail

Banking_CG, OOS_CG, and StackOverflow_CG are subsets derived from Banking, OOS, and StackOverflow by pruning compositionally-similar pairs of utterance instances between their training and test/development sets. Rouge-L score is adopted to identify overlap common subsequences in a pair of utterance instances. The larger Rouge-L score usually indicates that more common compositions (n-grams) are shared among the pair of utterances.

In Banking_CG and OOS_CG, we used a Rouge-L threshold of 0.3 to detect a similar pair of utterance instances between their training and test/development sets, while in StackOverflow_CG, a threshold of 0.2 is adopted. (Refer to Appendix B for more about Rouge score and the corresponding utterance pairs.)

Once the compositionally-similar pair of utterance instances are identified, a graph is created by connecting these instances with edges, then highest-degree nodes (instances) and their edges are pruned iteratively. Considering that training and test/development sets had significantly different numbers of instances, the node degree is multiplied by the weight (the number of remaining instances in the set) to readjust if the node to prune should be from the training or test/development set.

The pruning process iterates until a certain condition is met. In Banking_CG and OOS_CG, we stopped the process when the maximum node degree of the test/development sets reached a number of 5, while in StackOverflow_CG, the process wasn't stopped until all similar pairs were pruned, considering that StackOverflow_CG is relatively more balanced and has fewer intents to predict.

In Banking_CG, 6231, 183, and 1184 utterance instances were pruned from the corresponding training, development and test sets, while in OOS_CG, 11317, 1306, and 2068 were pruned, and in StackOverflow_CG, 9209, 578, and 3095 instances were removed from their training, development and test sets, respectively. Detailed statistics of Banking_CG, OOS_CG, and StackOverflow_CG can be found in Tables 2 to 10.

B Examples of Rouge Scores and Corresponding Utterance Pairs

The compositional similarity of a pair of utterances can be told by Rouge-L score. In the second row of Table 11, given that the Rouge-L score is greater than 0.3, a long span (4-gram) "be using my card" is shared by both the training and test utterance instances. When the Rouge-L score is not greater than 0.3, the first row and the third row of Table 11, literally those pairs are compositionally dissimilar and only a short span (bigram) "my card" is found common between training and test instances.

C Examples of ChatGPT's Paraphrases

Table 12 demonstrates that ChatGPT's paraphrases introduce diverse compositions from the original utterances. For example, in the first row of Table 12b, the bigram "equivalent of" is replaced with a trigram "corresponding phrase for" of the same meaning. In the first row of Table12a, the original sentence "i have a pending top-up" is put into its passive voice structure. The diversities brought by ChatGPT's paraphrases eventually bridge the gap between compositionally dissimilar training and test sets.

D Experimental Results in Detail

For a fair comparison, all settings are evaluated using the seed numbers 0 to 9 for known intent sampling. All settings are built on the BERT-base backbone and are optimized using the ADAM gradient descent algorithm. Full experimental results are shown in Tables 13 to 20.

Table 2: **Banking_CG** training dataset statistics

Intent	#Instance	Intent	#Instance
Refund_not_showing_up	54	get_physical_card	18
activate_my_card	49	getting_spare_card	29
age_limit	31	getting_virtual_card	16
apple_pay_or_google_pay	28	lost_or_stolen_card	22
atm_support	25	lost_or_stolen_phone	24
automatic_top_up	31	order_physical_card	29
balance_not_updated_after_bank_transfer	65	passcode_forgotten	26
balance_not_updated_after_cheque_or_cash_deposit	69	pending_card_payment	51
beneficiary_not_allowed	37	pending_cash_withdrawal	41
cancel_transfer	49	pending_top_up	44
card_about_to_expire	29	pending_transfer	45
card_acceptance	13	pin_blocked	33
card_arrival	51	receiving_money	22
card_delivery_estimate	32	request_refund	49
card_linking	34	reverted_card_payment?	45
card_not_working	33	supported_cards_and_currencies	38
card_payment_fee_charged	85	terminate_account	34
card_payment_not_recognised	64	top_up_by_bank_transfer_charge	20
card_payment_wrong_exchange_rate	68	top_up_by_card_charge	20
card_swallowed	10	top_up_by_cash_or_cheque	21
cash_withdrawal_charge	60	top_up_failed	35
cash_withdrawal_not_recognised	43	top_up_limits	21
change_pin	26	top_up_reverted	39
compromised_card	14	topping_up_by_card	20
contactless_not_working	15	transaction_charged_twice	57
country_support	34	transfer_fee_charged	55
declined_card_payment	40	transfer_into_account	21
declined_cash_withdrawal	54	transfer_not_received_by_recipient	63
declined_transfer	44	transfer_timing	26
direct_debit_payment_not_recognised	89	unable_to_verify_identity	15
disposable_card_limits	25	verify_my_identity	27
edit_personal_details	24	verify_source_of_funds	22
exchange_charge	25	verify_top_up	24
exchange_rate	21	virtual_card_not_working	5
exchange_via_app	24	visa_or_mastercard	37
extra_charge_on_statement	51	why_verify_identity	30
failed_transfer	33	wrong_amount_of_cash_received	67
fiat_currency_support	37	wrong_exchange_rate_for_cash_withdrawal	53
get_disposable_virtual_card	12		

Table 3: **Banking_CG** development dataset statistics

Intent	#Instance	Intent	#Instance
Refund_not_showing_up	13	get_physical_card	8
activate_my_card	12	getting_spare_card	10
age_limit	9	getting_virtual_card	6
apple_pay_or_google_pay	9	lost_or_stolen_card	6
atm_support	9	lost_or_stolen_phone	10
automatic_top_up	11	order_physical_card	9
balance_not_updated_after_bank_transfer	14	passcode_forgotten	5
balance_not_updated_after_cheque_or_cash_deposit	15	pending_card_payment	15
beneficiary_not_allowed	15	pending_cash_withdrawal	11
cancel_transfer	12	pending_top_up	13
card_about_to_expire	10	pending_transfer	14
card_acceptance	4	pin_blocked	7
card_arrival	11	receiving_money	8
card_delivery_estimate	8	request_refund	17
card_linking	10	reverted_card_payment?	14
card_not_working	7	supported_cards_and_currencies	10
card_payment_fee_charged	18	terminate_account	9
card_payment_not_recognised	17	top_up_by_bank_transfer_charge	9
card_payment_wrong_exchange_rate	12	top_up_by_card_charge	8
card_swallowed	4	top_up_by_cash_or_cheque	10
cash_withdrawal_charge	16	top_up_failed	13
cash_withdrawal_not_recognised	16	top_up_limits	8
change_pin	11	top_up_reverted	14
compromised_card	9	topping_up_by_card	10
contactless_not_working	3	transaction_charged_twice	16
country_support	12	transfer_fee_charged	17
declined_card_payment	14	transfer_into_account	8
declined_cash_withdrawal	17	transfer_not_received_by_recipient	14
declined_transfer	10	transfer_timing	9
direct_debit_payment_not_recognised	10	unable_to_verify_identity	9
disposable_card_limits	8	verify_my_identity	7
edit_personal_details	7	verify_source_of_funds	7
exchange_charge	10	verify_top_up	11
exchange_rate	7	virtual_card_not_working	3
exchange_via_app	9	visa_or_mastercard	11
extra_charge_on_statement	12	why_verify_identity	9
failed_transfer	13	wrong_amount_of_cash_received	17
fiat_currency_support	9	wrong_exchange_rate_for_cash_withdrawal	15
get_disposable_virtual_card	7		

Table 4: **Banking_CG** test dataset statistics

Intent	#Instance	Intent	#Instance
Refund_not_showing_up	28	get_physical_card	18
activate_my_card	31	getting_spare_card	29
age_limit	23	getting_virtual_card	21
apple_pay_or_google_pay	22	lost_or_stolen_card	16
atm_support	20	lost_or_stolen_phone	24
automatic_top_up	24	order_physical_card	15
balance_not_updated_after_bank_transfer	25	passcode_forgotten	16
balance_not_updated_after_cheque_or_cash_deposit	30	pending_card_payment	26
beneficiary_not_allowed	37	pending_cash_withdrawal	29
cancel_transfer	29	pending_top_up	25
card_about_to_expire	31	pending_transfer	34
card_acceptance	20	pin_blocked	21
card_arrival	24	receiving_money	26
card_delivery_estimate	23	request_refund	36
card_linking	28	reverted_card_payment?	35
card_not_working	20	supported_cards_and_currencies	24
card_payment_fee_charged	17	terminate_account	11
card_payment_not_recognised	20	top_up_by_bank_transfer_charge	20
card_payment_wrong_exchange_rate	20	top_up_by_card_charge	19
card_swallowed	17	top_up_by_cash_or_cheque	31
cash_withdrawal_charge	34	top_up_failed	30
cash_withdrawal_not_recognised	34	top_up_limits	19
change_pin	18	top_up_reverted	27
compromised_card	18	topping_up_by_card	17
contactless_not_working	20	transaction_charged_twice	35
country_support	18	transfer_fee_charged	33
declined_card_payment	32	transfer_into_account	27
declined_cash_withdrawal	35	transfer_not_received_by_recipient	27
declined_transfer	26	transfer_timing	26
direct_debit_payment_not_recognised	16	unable_to_verify_identity	30
disposable_card_limits	21	verify_my_identity	21
edit_personal_details	27	verify_source_of_funds	26
exchange_charge	24	verify_top_up	29
exchange_rate	22	virtual_card_not_working	9
exchange_via_app	22	visa_or_mastercard	20
extra_charge_on_statement	36	why_verify_identity	22
failed_transfer	27	wrong_amount_of_cash_received	29
fiat_currency_support	23	wrong_exchange_rate_for_cash_withdrawal	28
get_disposable_virtual_card	23		

Table 5: **OOS_CG** training dataset statistics

Intent	#Instance	Intent	#Instance	Intent	#Instance
accept_reservations	50	greeting	33	reset_settings	6
account_blocked	21	how_busy	20	restaurant_reservation	22
alarm	13	how_old_are_you	47	restaurant_reviews	47
application_status	22	improve_credit_score	11	restaurant_suggestion	23
apr	27	income	52	rewards_balance	15
are_you_a_bot	20	ingredient_substitution	42	roll_dice	11
balance	22	ingredients_list	26	rollover_401k	19
bill_balance	18	insurance	29	routing	29
bill_due	15	insurance_change	26	schedule_maintenance	19
book_flight	19	interest_rate	26	schedule_meeting	21
book_hotel	17	international_fees	16	share_location	24
calculator	53	international_visa	33	shopping_list	24
calendar	34	jump_start	10	shopping_list_update	16
calendar_update	17	last_maintenance	14	smart_home	19
calories	41	lost_luggage	28	spelling	33
cancel	30	make_call	20	spending_history	17
cancel_reservation	19	maybe	26	sync_device	7
car_rental	12	meal_suggestion	29	taxes	24
card_declined	16	meaning_of_life	17	tell_joke	26
carry_on	32	measurement_conversion	23	text	27
change_accent	23	meeting_schedule	45	thank_you	25
change_ai_name	15	min_payment	23	time	29
change_language	33	mpg	26	timer	14
change_speed	27	new_card	24	timezone	39
change_user_name	47	next_holiday	17	tire_change	26
change_volume	11	next_song	24	tire_pressure	21
confirm_reservation	25	no	25	todo_list	12
cook_time	24	nutrition_info	28	todo_list_update	21
credit_limit	17	oil_change_how	17	traffic	16
credit_limit_change	13	oil_change_when	15	transactions	26
credit_score	6	order	49	transfer	17
current_location	18	order_checks	20	translate	24
damaged_card	17	order_status	21	travel_alert	55
date	31	pay_bill	23	travel_notification	27
definition	62	payday	22	travel_suggestion	29
direct_deposit	14	pin_change	10	uber	14
directions	40	play_music	42	update_playlist	27
distance	57	plug_type	12	user_name	23
do_you_have_pets	19	pto_balance	7	vaccines	23
exchange_rate	33	pto_request	35	w2	18
expiration_date	13	pto_request_status	18	weather	21
find_phone	11	pto_used	14	what_are_your_hobbies	26
flight_status	29	recipe	37	what_can_i_ask_you	6
flip_coin	13	redeem_rewards	19	what_is_your_name	26
food_last	43	reminder	52	what_song	28
freeze_account	23	reminder_update	26	where_are_you_from	23
fun_fact	18	repeat	14	whisper_mode	23
gas	13	replacement_card_duration	16	who_do_you_work_for	32
gas_type	12	report_fraud	11	who_made_you	43
goodbye	43	report_lost_card	25	yes	47

Table 6: **OOS_CG** development dataset statistics

Intent	#Instance	Intent	#Instance	Intent	#Instance
accept_reservations	15	greeting	13	reset_settings	13
account_blocked	11	how_busy	15	restaurant_reservation	10
alarm	14	how_old_are_you	11	restaurant_reviews	9
application_status	12	improve_credit_score	9	restaurant_suggestion	19
apr	6	income	6	rewards_balance	9
are_you_a_bot	10	ingredient_substitution	13	roll_dice	13
balance	14	ingredients_list	12	rollover_401k	9
bill_balance	11	insurance	11	routing	4
bill_due	9	insurance_change	9	schedule_maintenance	16
book_flight	18	interest_rate	9	schedule_meeting	8
book_hotel	15	international_fees	15	share_location	14
calculator	10	international_visa	5	shopping_list	5
calendar	9	jump_start	17	shopping_list_update	13
calendar_update	16	last_maintenance	11	smart_home	18
calories	5	lost_luggage	15	spelling	11
cancel	18	make_call	14	spending_history	16
cancel_reservation	19	maybe	18	sync_device	13
car_rental	11	meal_suggestion	14	taxes	11
card_declined	11	meaning_of_life	15	tell_joke	13
carry_on	18	measurement_conversion	15	text	9
change_accent	15	meeting_schedule	8	thank_you	15
change_ai_name	13	min_payment	8	time	3
change_language	8	mpg	12	timer	14
change_speed	14	new_card	7	timezone	8
change_user_name	14	next_holiday	11	tire_change	8
change_volume	15	next_song	9	tire_pressure	11
confirm_reservation	12	no	15	todo_list	12
cook_time	9	nutrition_info	11	todo_list_update	5
credit_limit	4	oil_change_how	4	traffic	11
credit_limit_change	12	oil_change_when	8	transactions	14
credit_score	5	order	13	transfer	9
current_location	11	order_checks	16	translate	15
damaged_card	11	order_status	16	travel_alert	11
date	10	pay_bill	8	travel_notification	12
definition	12	payday	9	travel_suggestion	14
direct_deposit	6	pin_change	13	uber	11
directions	10	play_music	16	update_playlist	6
distance	6	plug_type	11	user_name	11
do_you_have_pets	14	pto_balance	9	vaccines	8
exchange_rate	14	pto_request	9	w2	11
expiration_date	9	pto_request_status	10	weather	16
find_phone	5	pto_used	12	what_are_your_hobbies	6
flight_status	7	recipe	11	what_can_i_ask_you	12
flip_coin	16	redeem_rewards	6	what_is_your_name	8
food_last	11	reminder	9	what_song	10
freeze_account	4	reminder_update	12	where_are_you_from	19
fun_fact	18	repeat	13	whisper_mode	12
gas	10	replacement_card_duration	10	who_do_you_work_for	9
gas_type	11	report_fraud	14	who_made_you	9
goodbye	16	report_lost_card	8	yes	12

Table 7: **OOS_CG** test dataset statistics

Intent	#Instance	Intent	#Instance	Intent	#Instance
accept_reservations	17	how_busy	22	restaurant_reservation	17
account_blocked	15	how_old_are_you	16	restaurant_reviews	20
alarm	30	improve_credit_score	8	restaurant_suggestion	21
application_status	19	income	17	rewards_balance	12
apr	11	ingredient_substitution	19	roll_dice	17
are_you_a_bot	17	ingredients_list	20	rollover_401k	11
balance	15	insurance	11	routing	5
bill_balance	17	insurance_change	11	schedule_maintenance	12
bill_due	10	interest_rate	17	schedule_meeting	11
book_flight	15	international_fees	22	share_location	22
book_hotel	17	international_visa	6	shopping_list	5
calculator	14	jump_start	11	shopping_list_update	17
calendar	12	last_maintenance	18	smart_home	22
calendar_update	19	lost_luggage	24	spelling	15
calories	15	make_call	20	spending_history	14
cancel	23	maybe	20	sync_device	12
cancel_reservation	20	meal_suggestion	17	taxes	16
car_rental	15	meaning_of_life	23	tell_joke	15
card_declined	9	measurement_conversion	21	text	18
carry_on	22	meeting_schedule	16	thank_you	25
change_accent	14	min_payment	24	time	10
change_ai_name	18	mpg	25	timer	29
change_language	17	new_card	5	timezone	12
change_speed	15	next_holiday	21	tire_change	8
change_user_name	14	next_song	8	tire_pressure	12
change_volume	16	no	23	todo_list	18
confirm_reservation	14	nutrition_info	25	todo_list_update	10
cook_time	8	oil_change_how	9	traffic	13
credit_limit	10	oil_change_when	11	transactions	20
credit_limit_change	21	oos	1200	transfer	15
credit_score	28	order	17	translate	23
current_location	21	order_checks	17	travel_alert	13
damaged_card	16	order_status	20	travel_notification	14
date	15	pay_bill	6	travel_suggestion	15
definition	17	payday	13	uber	20
direct_deposit	19	pin_change	13	update_playlist	8
directions	19	play_music	21	user_name	7
distance	15	plug_type	18	vaccines	12
do_you_have_pets	25	pto_balance	9	w2	23
exchange_rate	21	pto_request	15	weather	26
expiration_date	17	pto_request_status	15	what_are_your_hobbies	13
find_phone	9	pto_used	20	what_can_i_ask_you	18
flight_status	19	recipe	19	what_is_your_name	15
flip_coin	24	redeem_rewards	19	what_song	13
food_last	16	reminder	14	where_are_you_from	21
freeze_account	9	reminder_update	17	whisper_mode	22
fun_fact	11	repeat	22	who_do_you_work_for	13
gas	16	replacement_card_duration	11	who_made_you	12
gas_type	15	report_fraud	15	yes	16
goodbye	22	report_lost_card	15		
greeting	21	reset_settings	14		

Table 8: **StackOverflow_CG** training dataset statistics

Intent	#Instance	Intent	#Instance	Intent	#Instance	Intent	#Instance	Intent	#Instance
ajax	158	drupal	132	linq	105	osx	139	spring	154
apache	143	excel	106	magento	112	qt	147	svn	116
bash	104	haskell	126	matlab	132	scala	135	visual-studio	200
cocoa	233	hibernate	130	oracle	124	sharepoint	140	wordpress	155

Table 9: **StackOverflow_CG** development dataset statistics

Intent	#Instance	Intent	#Instance	Intent	#Instance	Intent	#Instance	Intent	#Instance
ajax	74	drupal	70	linq	78	osx	91	spring	67
apache	80	excel	70	magento	66	qt	83	svn	75
bash	80	haskell	71	matlab	73	scala	72	visual-studio	42
cocoa	66	hibernate	68	oracle	63	sharepoint	75	wordpress	58

Table 10: **StackOverflow_CG** test dataset statistics

Intent	#Instance	Intent	#Instance	Intent	#Instance	Intent	#Instance	Intent	#Instance
ajax	160	drupal	128	linq	131	osx	193	spring	151
apache	170	excel	140	magento	132	qt	172	svn	149
bash	139	haskell	122	matlab	164	scala	124	visual-studio	106
cocoa	161	hibernate	128	oracle	127	sharepoint	171	wordpress	137

Table 11: Examples of Rouge scores and corresponding utterance pairs

Training instance	Someone might be using my card that is not me.
Test instance	I don't recognize some of the transactions on my card, I think someone must have gotten my card info and used it.
Rouge-L score	0.13
Training instance	Someone might be using my card that is not me.
Test instance	What should I do if I think that someone else may be using my card.
Rouge-L score	0.33
Training instance	Someone might be using my card that is not me.
Test instance	I think someone got my card details and used it because there are transactions i don't recognize. What do I do now?
Rouge-L score	0.13

Table 12: Examples of ChatGPT's Paraphrases

(a) Banking_CG Dataset

Original utterance	i have a pending top-up
ChatGPT's paraphrase	A top-up transaction is still being processed.
Original utterance	What are the currency types that I can keep money in.
ChatGPT's paraphrase	Which currencies are allowed for me to keep my finances?
Original utterance	I think my card is broken or something. I tried using it and it wouldn't go through, please help.
ChatGPT's paraphrase	It appears that my card may be broken or faulty, as I tried to use it but it didn't work. Could you provide me with some assistance, please?

(b) OOS_CG Dataset

Original utterance	what is the equivalent of, 'life is good' in french
ChatGPT's paraphrase	What is the corresponding phrase for 'life is good' in French?
Original utterance	how many inches and centimeters would i need if i had 10 feet of something
ChatGPT's paraphrase	How many inches and centimeters would be needed for 10 feet of an object?
Original utterance	please put taco night on the house calendar date of january 6
ChatGPT's paraphrase	Kindly mark January 6th as the date for taco night on the house calendar.

(c) StackOverflow_CG Dataset

Original utterance	Why this keyboard intercepting kernel extension doesn't work?
ChatGPT's paraphrase	Can you provide an answer for the non-functionality of the keyboard interception kernel extension?
Original utterance	NSColor with calibrated values works differently than regular color??
ChatGPT's paraphrase	Calibrated NSColor showcases a diverse behavior from a typical color.
Original utterance	Tips for using CVS or Subversion as a backup framework for office documents
ChatGPT's paraphrase	Guidelines for utilizing CVS or Subversion as a backup solution for office documents.

Table 13: ADB experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	54.32	81.06	55.66	71.10	45.65	91.86	46.83	85.63	66.48	88.37	70.13	82.55
	1	54.36	79.78	55.63	70.94	52.36	91.18	53.35	84.77	60.63	80.47	63.94	72.94
	2	53.99	79.04	55.24	71.10	49.16	91.46	50.24	84.64	59.12	80.55	62.69	72.87
	3	53.55	80.15	54.88	71.78	49.81	89.61	50.83	82.57	59.95	85.19	64.16	78.04
	4	59.76	81.31	60.84	73.31	54.07	91.97	55.04	85.68	58.86	78.74	62.17	70.67
	5	49.72	81.85	51.33	72.52	50.03	90.73	51.07	84.03	55.42	69.00	57.68	61.17
	6	48.92	78.28	50.39	68.88	55.00	89.41	55.88	82.74	59.18	81.67	62.93	73.22
	7	50.79	83.28	52.41	74.21	44.05	89.56	45.22	82.46	61.03	84.09	64.88	76.90
	8	52.18	83.82	53.76	75.11	40.09	89.34	41.35	81.75	53.36	81.30	58.02	72.60
	9	57.33	82.40	58.58	74.10	51.07	91.41	52.11	85.32	49.71	61.01	51.59	52.56
50%	0	57.44	72.01	57.82	66.51	51.47	83.75	51.89	75.55	75.71	82.67	76.34	80.07
	1	62.04	68.92	62.22	65.72	52.52	83.81	52.93	75.30	70.56	73.05	70.78	70.60
	2	61.66	70.50	61.89	66.93	50.64	85.10	51.09	76.29	70.76	78.05	71.42	74.18
	3	58.88	69.09	59.14	64.98	54.26	84.22	54.65	76.35	70.45	74.29	70.80	71.70
	4	62.11	69.96	62.31	66.30	55.06	84.01	55.44	75.63	68.29	71.15	68.55	69.12
	5	58.56	72.02	58.91	66.30	52.78	84.17	53.20	75.69	74.22	77.70	74.54	76.01
	6	55.92	67.11	56.21	62.08	51.95	83.97	52.37	75.94	73.29	80.95	73.99	77.45
	7	57.74	66.71	57.97	61.87	49.92	83.73	50.37	74.83	70.24	75.19	70.69	72.70
	8	60.93	71.71	61.21	67.41	49.19	82.40	49.62	73.49	69.59	73.54	69.95	71.53
	9	63.98	68.29	64.09	65.66	55.37	84.78	55.75	77.53	71.41	74.80	71.72	72.46
75%	0	65.47	56.31	65.31	64.50	53.11	75.08	53.30	66.82	77.73	68.20	77.13	74.97
	1	66.17	51.73	65.93	63.71	53.75	76.63	53.95	68.83	74.97	57.43	73.87	69.60
	2	64.73	52.93	64.53	63.08	52.09	76.01	52.30	67.21	75.13	65.37	74.52	71.94
	3	63.41	49.66	63.18	61.50	55.98	75.57	56.15	68.86	75.72	57.86	74.60	69.98
	4	64.99	52.81	64.79	63.24	56.38	76.64	56.56	69.49	76.82	61.84	75.88	72.87
	5	64.98	59.67	64.89	64.87	51.57	78.20	51.81	68.78	78.43	59.41	77.24	72.25
	6	62.19	53.26	62.04	60.86	53.56	76.88	53.77	68.89	74.44	58.53	73.44	68.74
	7	59.66	51.17	59.51	59.02	52.58	76.27	52.79	68.01	78.70	66.63	77.95	75.22
	8	65.18	54.89	65.00	63.87	52.86	74.40	53.05	66.33	75.25	57.61	74.15	69.81
	9	66.26	51.22	66.01	63.50	56.87	76.72	57.05	70.10	74.12	62.73	73.40	70.43

Table 14: DA-ADB experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	51.66	88.27	53.49	80.01	33.93	92.54	35.43	86.62	70.98	90.81	74.29	85.96
	1	55.33	84.21	56.77	76.21	37.62	91.84	39.01	85.55	66.88	87.19	70.26	81.03
	2	50.86	84.62	52.55	77.16	35.91	91.44	37.33	84.91	60.42	82.93	64.17	74.84
	3	55.55	85.63	57.05	78.22	39.84	91.12	41.16	84.61	66.89	88.05	70.41	82.03
	4	58.68	86.10	60.05	79.06	45.06	92.47	46.27	86.81	60.39	84.17	64.35	76.76
	5	51.81	86.97	53.57	78.96	44.04	92.12	45.27	86.29	62.54	81.87	65.76	74.22
	6	46.76	83.06	48.57	74.47	44.73	91.67	45.93	85.71	63.63	87.01	67.52	80.07
	7	49.18	87.05	51.08	78.74	35.03	90.76	36.46	83.92	64.84	86.87	68.51	80.45
	8	54.86	87.38	56.49	80.33	30.67	91.08	32.22	84.33	52.25	83.31	57.43	74.63
	9	58.64	88.16	60.11	81.17	35.87	91.93	37.31	85.74	54.41	76.14	58.04	66.85
50%	0	49.77	75.22	50.42	66.72	31.77	84.11	32.46	74.15	80.15	86.31	80.71	84.17
	1	52.47	71.43	52.96	65.19	34.71	83.17	35.34	73.24	77.32	81.37	77.69	78.93
	2	57.29	74.85	57.74	69.30	31.49	83.23	32.17	72.91	72.93	81.07	73.67	77.18
	3	53.55	72.26	54.03	66.14	34.22	83.41	34.86	73.82	74.07	77.89	74.42	75.52
	4	55.62	73.03	56.07	66.77	37.00	82.77	37.60	72.96	74.59	81.62	75.23	78.66
	5	52.39	76.15	53.00	68.57	37.80	83.32	38.40	73.68	75.76	80.38	76.18	78.18
	6	50.45	77.04	51.13	67.93	35.98	83.47	36.61	73.84	79.23	85.97	79.84	83.17
	7	55.90	74.81	56.38	68.04	32.06	83.22	32.73	72.80	76.72	83.76	77.36	81.31
	8	58.98	77.10	59.44	71.20	31.04	82.92	31.72	72.60	71.85	77.45	72.36	74.87
	9	59.34	72.63	59.68	67.83	30.50	83.46	31.19	73.68	77.05	81.71	77.47	79.38
75%	0	54.14	56.52	54.18	58.97	28.11	72.94	28.51	60.71	81.00	71.38	80.40	78.00
	1	55.05	51.20	54.99	56.75	28.66	71.49	29.04	59.55	75.19	63.00	74.43	71.46
	2	54.38	49.80	54.30	55.85	29.51	72.10	29.89	60.27	77.36	67.09	76.72	74.04
	3	55.38	48.25	55.26	55.27	28.06	71.12	28.44	59.36	78.03	61.71	77.01	72.74
	4	55.04	48.45	54.93	55.49	30.00	71.20	30.37	59.00	81.17	69.09	80.41	77.80
	5	52.59	55.11	52.63	56.80	29.10	73.25	29.49	60.79	79.46	62.63	78.40	74.01
	6	54.04	56.19	54.07	58.02	30.41	71.54	30.77	59.80	76.96	62.10	76.04	71.57
	7	51.78	51.07	51.77	54.32	29.22	71.38	29.59	59.42	81.15	70.17	80.47	77.97
	8	55.68	55.04	55.67	58.23	30.14	70.19	30.50	58.73	77.66	62.87	76.74	72.94
	9	59.35	52.95	59.24	59.70	32.61	72.40	32.97	61.43	77.73	67.98	77.12	74.56

Table 15: ADB+GPTAUG-F4 experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	58.04	85.27	59.40	77.48	52.90	92.93	53.92	87.78	66.07	87.09	69.58	80.90
	1	57.11	81.53	58.33	73.63	56.85	92.69	57.77	87.33	63.04	83.53	66.46	76.56
	2	59.83	83.40	61.01	76.53	54.62	92.73	55.60	87.22	62.06	83.26	65.59	76.11
	3	58.91	83.93	60.16	76.74	53.28	90.81	54.24	84.64	59.22	84.74	63.48	77.52
	4	61.81	82.72	62.86	74.79	56.91	92.21	57.82	86.48	58.39	78.63	61.76	70.50
	5	53.76	84.91	55.31	77.00	56.47	92.70	57.40	87.47	57.87	71.56	60.15	63.86
	6	49.84	78.68	51.28	69.57	59.21	91.51	60.04	86.10	63.44	86.24	67.24	79.14
	7	51.00	84.02	52.65	75.37	45.39	89.72	46.53	82.76	61.32	84.23	65.14	77.11
	8	57.43	84.96	58.81	77.37	44.15	91.12	45.35	84.64	59.58	86.30	64.03	79.14
	9	59.54	84.31	60.78	76.58	55.61	92.91	56.56	87.89	58.38	80.57	62.08	72.36
50%	0	59.88	72.93	60.21	67.99	55.53	85.53	55.92	78.30	74.80	82.30	75.48	79.48
	1	64.52	71.24	64.69	68.72	56.29	84.85	56.67	77.56	73.32	79.95	73.93	76.63
	2	64.21	72.72	64.43	69.67	53.62	85.84	54.04	77.92	69.22	78.59	70.07	74.04
	3	60.36	71.55	60.65	67.77	55.85	85.28	56.24	78.19	70.77	76.59	71.30	73.43
	4	65.22	74.51	65.46	70.62	58.56	85.23	58.91	78.14	66.06	64.09	65.88	64.27
	5	60.60	75.58	60.98	70.15	55.21	85.54	55.60	77.97	73.83	77.12	74.13	75.32
	6	60.24	76.12	60.65	71.10	57.29	86.50	57.68	79.90	73.91	82.90	74.72	79.17
	7	61.76	72.72	62.04	68.04	54.12	85.37	54.53	77.75	71.47	76.16	71.90	73.94
	8	63.54	73.68	63.80	70.31	51.68	84.15	52.11	76.21	68.46	72.63	68.84	70.43
	9	65.21	70.97	65.35	67.99	55.47	85.59	55.87	78.77	73.98	79.95	74.53	77.01
75%	0	68.56	57.90	68.38	67.09	57.04	78.70	57.23	71.92	77.32	67.80	76.72	74.32
	1	68.72	53.37	68.46	65.98	58.07	77.79	58.25	71.45	76.73	62.79	75.86	72.43
	2	65.97	55.15	65.79	64.66	54.59	77.05	54.79	69.60	73.05	63.29	72.44	69.57
	3	65.24	50.29	64.98	62.97	56.72	76.96	56.90	71.01	75.24	59.45	74.25	69.64
	4	64.58	51.19	64.36	62.39	58.43	77.49	58.60	71.56	76.35	61.50	75.42	72.01
	5	66.94	58.90	66.81	65.88	53.77	78.05	53.98	70.43	77.68	59.26	76.53	71.60
	6	64.75	56.67	64.61	63.98	53.65	76.91	53.86	69.71	73.96	57.14	72.91	67.78
	7	67.40	55.04	67.19	65.72	53.77	77.22	53.98	70.04	78.29	65.85	77.51	74.60
	8	66.09	54.49	65.89	64.08	55.79	76.08	55.97	69.77	75.00	56.65	73.86	68.95
	9	68.24	55.17	68.02	66.14	58.04	77.31	58.21	71.48	73.54	63.11	72.88	69.91

Table 16: ADB+GPTAUG-F10 experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	58.71	86.00	60.07	78.59	53.03	92.85	54.05	87.67	66.19	86.48	69.58	80.17
	1	57.87	81.50	59.05	73.68	57.33	92.70	58.24	87.42	62.82	83.08	66.19	75.97
	2	60.23	82.65	61.35	75.74	55.26	92.90	56.22	87.50	58.46	78.12	61.74	70.19
	3	58.84	83.84	60.09	76.64	52.85	91.08	53.83	84.99	53.44	76.63	57.30	68.26
	4	63.74	85.02	64.80	78.01	58.23	92.79	59.11	87.56	58.66	80.21	62.25	72.22
	5	54.81	85.70	56.36	78.11	58.45	92.92	59.33	87.91	53.65	67.25	55.92	59.66
	6	50.23	79.56	51.70	70.73	59.61	91.42	60.43	85.90	58.26	81.54	62.14	73.18
	7	52.00	84.93	53.65	76.58	46.39	89.63	47.49	82.63	60.53	83.16	64.30	75.73
	8	57.18	85.03	58.57	77.43	45.55	91.60	46.73	85.46	58.85	85.62	63.31	78.18
	9	62.15	86.17	63.36	79.06	55.95	92.81	56.90	87.75	59.03	81.32	62.75	73.08
50%	0	59.45	73.31	59.81	68.51	56.05	86.22	56.45	79.32	74.30	82.36	75.03	79.35
	1	64.61	72.94	64.82	70.09	57.44	84.34	57.79	76.98	71.94	78.66	72.55	75.25
	2	63.61	73.12	63.85	69.99	53.01	86.08	53.45	78.61	68.62	78.56	69.52	74.11
	3	60.49	71.53	60.77	67.77	55.27	85.53	55.67	78.58	69.86	76.45	70.45	73.05
	4	65.83	74.73	66.06	71.26	58.02	85.55	58.38	78.69	67.31	68.90	67.45	67.37
	5	59.64	74.90	60.03	69.67	56.20	86.01	56.59	78.83	73.22	78.02	73.66	75.80
	6	60.02	76.72	60.45	71.62	56.64	86.21	57.03	79.71	74.21	83.34	75.04	79.62
	7	61.61	72.19	61.88	67.62	54.29	85.35	54.70	77.75	71.85	80.04	72.59	76.83
	8	62.66	74.10	62.95	70.57	51.47	84.50	51.90	76.82	69.22	75.75	69.81	72.87
	9	64.91	68.75	65.01	66.51	55.60	85.85	56.00	79.13	69.23	73.57	69.62	70.95
75%	0	68.27	57.05	68.08	66.51	56.60	78.20	56.80	71.50	77.85	68.49	77.27	74.91
	1	68.10	53.69	67.86	65.56	56.99	77.18	57.17	70.90	76.53	60.94	75.56	71.29
	2	65.71	56.59	65.56	64.98	53.54	76.59	53.74	69.71	72.94	62.67	72.30	69.36
	3	65.29	52.34	65.07	63.55	57.20	76.27	57.37	70.46	74.57	58.22	73.55	68.74
	4	64.32	50.80	64.09	62.24	58.62	77.76	58.79	72.00	76.26	62.65	75.41	72.08
	5	65.75	57.27	65.60	64.61	54.58	77.99	54.79	70.65	77.08	59.47	75.97	71.33
	6	64.39	54.56	64.22	63.13	52.86	76.82	53.07	69.63	73.89	59.24	72.98	68.40
	7	66.59	54.30	66.38	65.08	53.26	76.52	53.47	69.36	77.40	64.11	76.56	73.56
	8	64.78	54.95	64.62	63.66	54.11	75.50	54.30	69.27	75.57	56.72	74.40	69.12
	9	68.97	54.48	68.72	66.03	58.63	77.61	58.80	71.94	71.42	59.03	70.64	67.33

Table 17: ADB+GPTAUG-WP10 experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	53.68	74.20	54.71	63.82	42.78	88.54	43.95	80.23	55.82	55.25	55.72	50.26
	1	51.63	71.00	52.59	61.60	46.34	86.13	47.36	77.18	52.86	61.73	54.34	55.11
	2	50.59	67.94	51.46	60.44	47.38	89.55	48.46	81.55	50.06	60.37	51.78	53.56
	3	54.12	74.17	55.12	65.98	49.64	87.59	50.62	79.71	49.37	70.08	52.83	61.55
	4	58.30	77.65	59.27	69.57	54.61	91.64	55.56	85.27	56.11	74.04	59.10	65.92
	5	45.41	72.00	46.74	61.81	50.45	90.10	51.47	83.20	53.80	68.79	56.30	61.27
	6	47.58	67.77	48.59	58.54	57.17	89.23	58.00	82.65	50.65	68.32	53.60	59.35
	7	42.20	54.62	42.82	47.20	42.85	85.61	43.95	76.57	50.25	62.95	52.37	55.59
	8	45.03	72.01	46.38	61.50	41.98	88.94	43.19	81.22	45.41	54.21	46.88	47.13
	9	51.87	73.33	52.94	64.24	47.13	88.34	48.19	80.70	51.82	59.69	53.13	51.91
50%	0	58.94	62.37	59.03	60.13	52.23	82.91	52.64	74.45	71.75	74.91	72.04	73.29
	1	62.10	58.44	62.00	60.07	52.31	82.77	52.71	74.28	70.46	71.96	70.59	70.02
	2	61.63	65.70	61.73	63.61	51.46	84.34	51.89	75.39	64.61	65.84	64.73	64.20
	3	59.66	64.29	59.78	62.18	55.36	83.71	55.73	75.74	69.14	70.79	69.29	69.16
	4	62.01	59.84	61.95	60.50	56.44	83.40	56.79	74.72	56.88	29.67	54.41	46.92
	5	56.61	63.18	56.78	60.28	54.38	84.28	54.77	75.94	69.88	65.79	69.51	67.13
	6	56.27	58.66	56.33	57.49	53.58	83.42	53.97	75.33	62.22	55.86	61.64	58.42
	7	58.79	57.53	58.76	57.70	49.73	82.41	50.16	73.32	68.51	69.89	68.64	68.85
	8	60.05	61.67	60.09	60.44	49.75	80.34	50.15	71.23	65.71	64.05	65.56	64.54
	9	62.64	59.46	62.56	60.34	57.26	83.06	57.60	75.66	71.25	72.51	71.36	71.15
75%	0	67.13	55.05	66.93	65.14	53.69	74.77	53.88	66.63	76.61	63.64	75.80	72.98
	1	66.55	45.06	66.19	62.45	55.60	75.61	55.78	67.84	74.09	50.07	72.59	67.37
	2	63.93	47.66	63.66	61.55	53.02	75.63	53.22	67.10	73.16	55.16	72.04	68.19
	3	65.11	46.54	64.80	62.13	57.03	75.30	57.19	68.69	72.96	39.88	70.89	64.89
	4	66.03	46.44	65.70	62.18	56.62	76.20	56.79	68.92	76.12	54.41	74.77	70.71
	5	66.35	55.35	66.16	64.35	52.77	77.94	52.99	68.45	74.51	43.38	72.56	67.02
	6	64.69	50.50	64.45	61.87	55.38	76.83	55.57	68.81	73.15	49.34	71.67	66.16
	7	61.94	46.28	61.68	60.07	52.85	75.50	53.06	67.37	75.28	54.81	74.01	70.22
	8	65.27	47.83	64.98	62.13	53.96	73.84	54.14	66.22	73.53	41.78	71.54	65.34
	9	65.17	39.13	64.73	60.34	58.19	76.18	58.35	69.66	69.21	46.78	67.81	63.79

Table 18: DA-ADB+GPTAUG-F4 experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	55.09	86.53	56.66	78.69	38.42	92.36	39.80	86.43	71.49	90.20	74.61	85.16
	1	57.86	82.82	59.11	75.32	42.45	91.97	43.72	85.82	45.42	44.39	45.25	37.59
	2	50.66	83.60	52.31	76.11	41.85	91.92	43.14	85.85	63.07	84.41	66.63	77.04
	3	56.91	83.70	58.25	76.16	46.46	91.05	47.60	84.75	63.46	87.15	67.41	80.59
	4	62.07	84.82	63.21	77.74	50.34	92.73	51.42	87.47	55.97	80.22	60.01	71.64
	5	51.81	85.38	53.49	77.37	48.41	92.23	49.53	86.65	56.85	54.46	56.45	46.47
	6	47.89	83.49	49.67	75.16	53.37	92.06	54.36	86.65	68.30	88.74	71.71	82.65
	7	47.67	85.26	49.55	76.58	38.10	90.42	39.44	83.54	62.50	86.83	66.55	80.07
	8	56.29	85.71	57.76	78.22	36.07	91.53	37.50	85.19	54.72	84.43	59.68	76.28
	9	59.50	86.93	60.88	79.75	44.37	92.27	45.60	86.51	57.94	79.49	61.53	70.43
50%	0	50.54	73.12	51.12	66.51	35.47	84.41	36.11	75.22	78.31	84.84	78.90	82.48
	1	59.47	75.31	59.88	70.83	40.56	83.88	41.13	74.92	54.14	13.38	50.43	37.56
	2	57.98	75.05	58.42	70.04	38.37	84.00	38.97	74.61	70.96	78.61	71.66	74.53
	3	55.67	72.29	56.10	67.25	38.04	83.84	38.65	74.81	70.70	75.70	71.15	72.87
	4	58.89	73.87	59.27	69.15	40.96	83.27	41.51	74.37	71.34	80.37	72.16	76.63
	5	54.31	74.02	54.81	68.09	39.35	83.48	39.93	74.26	76.41	80.97	76.82	79.00
	6	56.25	77.70	56.80	71.99	43.06	84.56	43.60	76.24	75.68	82.58	76.30	79.38
	7	55.99	74.21	56.46	68.04	37.55	84.02	38.16	74.56	75.92	83.00	76.56	80.34
	8	59.88	76.63	60.31	72.10	38.05	83.97	38.65	74.86	73.36	79.61	73.93	76.97
	9	61.65	74.48	61.98	70.09	37.13	84.52	37.75	75.74	75.95	81.64	76.46	79.04
75%	0	55.25	54.49	55.24	58.97	33.04	73.38	33.40	62.42	80.62	71.23	80.03	77.73
	1	57.69	50.92	57.57	58.18	33.93	72.34	34.27	61.70	78.27	65.96	77.50	74.32
	2	55.24	50.60	55.16	56.96	35.07	72.62	35.41	62.33	76.72	65.31	76.00	72.60
	3	53.68	46.91	53.57	54.17	32.50	72.04	32.85	61.59	77.15	60.04	76.08	71.33
	4	53.99	47.43	53.88	54.96	34.23	72.39	34.57	62.25	78.95	65.26	78.09	74.94
	5	54.30	54.09	54.30	57.81	31.94	73.10	32.30	61.59	78.56	61.99	77.52	73.05
	6	53.14	52.15	53.12	56.70	34.54	72.64	34.87	62.25	75.95	60.29	74.97	70.15
	7	57.41	52.71	57.33	58.91	32.39	71.39	32.73	60.57	73.60	50.75	72.17	66.57
	8	56.93	54.07	56.88	59.39	31.40	70.80	31.75	60.05	75.43	49.51	73.81	67.23
	9	61.08	52.53	60.94	60.60	35.98	73.35	36.31	63.33	73.46	57.54	72.47	67.13

Table 19: DA-ADB+GPTAUG-F10 experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	52.79	85.62	54.43	77.58	38.97	92.28	40.34	86.29	65.05	79.99	67.54	71.88
	1	56.34	81.59	57.61	73.73	45.12	92.08	46.32	86.07	58.02	81.47	61.93	72.91
	2	50.23	82.99	51.87	75.32	41.15	91.85	42.45	85.66	54.62	76.51	58.26	66.33
	3	57.28	83.52	58.59	75.95	44.85	90.85	46.03	84.39	70.38	90.84	73.79	85.78
	4	56.53	82.43	57.83	73.95	50.29	92.19	51.36	86.59	59.89	84.56	64.00	76.97
	5	52.06	85.16	53.72	77.06	46.97	92.03	48.13	86.32	58.56	75.63	61.41	66.54
	6	47.94	83.57	49.72	75.26	51.96	91.89	52.99	86.32	71.34	90.61	74.55	85.40
	7	49.23	84.65	51.00	76.00	37.52	90.29	38.87	83.31	63.71	88.52	67.84	82.34
	8	55.30	84.64	56.76	76.90	38.27	91.38	39.63	85.02	56.27	84.82	61.03	76.63
	9	57.45	85.85	58.87	78.22	46.91	92.54	48.08	87.00	40.31	0.17	33.62	16.18
50%	0	50.07	71.87	50.63	65.45	36.62	84.47	37.25	75.36	77.45	85.47	78.18	82.86
	1	58.32	75.07	58.75	70.46	39.97	83.92	40.55	74.86	55.61	27.05	53.01	42.10
	2	59.39	74.99	59.79	70.46	38.79	84.03	39.39	74.64	70.36	80.53	71.29	76.21
	3	53.08	71.80	53.56	66.30	38.38	83.62	38.98	74.56	67.74	76.56	68.54	72.74
	4	56.98	73.86	57.41	68.83	40.59	83.27	41.15	74.34	70.46	77.36	71.09	73.80
	5	53.81	74.10	54.33	67.99	40.70	83.56	41.26	74.42	75.73	81.43	76.25	79.00
	6	56.50	77.59	57.04	71.99	42.66	84.53	43.21	76.16	73.76	83.17	74.61	79.52
	7	55.97	74.49	56.44	68.30	39.05	83.96	39.64	74.70	71.88	81.52	72.76	78.18
	8	59.42	76.05	59.85	71.15	36.42	83.68	37.04	74.34	71.55	78.52	72.18	75.59
	9	61.70	74.40	62.03	69.99	37.01	84.37	37.63	75.58	68.63	76.20	69.32	72.67
75%	0	53.43	53.49	53.43	57.54	34.01	73.64	34.36	63.33	77.10	67.88	76.52	74.25
	1	56.61	49.72	56.49	57.01	34.39	72.33	34.73	61.98	76.49	64.84	75.76	72.91
	2	54.23	50.28	54.16	56.17	36.57	73.04	36.89	63.24	74.02	63.03	73.34	69.95
	3	53.09	45.94	52.97	53.43	31.37	71.57	31.73	60.66	72.03	52.61	70.81	65.47
	4	53.99	46.46	53.86	54.22	33.32	72.17	33.66	61.95	73.55	54.27	72.35	68.50
	5	54.03	53.96	54.02	57.54	32.15	73.15	32.52	61.81	76.89	60.00	75.83	71.19
	6	52.23	52.07	52.22	56.17	34.82	72.76	35.15	62.39	73.87	57.40	72.84	67.54
	7	54.64	50.03	54.56	56.01	33.02	71.70	33.36	60.96	77.18	67.03	76.54	74.35
	8	56.11	52.99	56.06	58.44	32.94	71.29	33.28	60.93	71.23	51.46	70.00	64.75
	9	60.22	51.66	60.07	59.92	35.47	73.15	35.80	63.19	64.18	9.64	60.77	52.05

Table 20: DA-ADB+GPTAUG-WP10 experimental results in detail

	Seed	Banking_CG				OOS_CG				StackOverflow_CG			
		F1-IND	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**	**F1-IND**	**F1-OOD**	**F1-All**	**Acc-All**
25%	0	53.70	86.98	55.37	78.06	38.21	92.80	39.61	87.00	63.39	81.83	66.46	74.73
	1	57.08	81.07	58.28	72.47	44.31	91.58	45.53	85.24	47.28	24.40	43.47	30.12
	2	55.92	81.19	57.18	73.58	39.26	91.75	40.61	85.41	47.71	29.83	44.73	31.57
	3	55.73	82.63	57.08	74.47	45.49	90.52	46.64	83.92	59.70	82.07	63.43	74.60
	4	62.46	83.43	63.51	75.79	50.07	92.43	51.15	86.81	58.29	77.75	61.53	69.74
	5	50.26	83.71	51.93	74.42	44.47	91.75	45.69	85.74	52.03	70.83	55.16	62.79
	6	49.43	77.93	50.85	68.78	49.18	91.79	50.28	86.10	60.47	80.97	63.88	73.22
	7	47.37	84.04	49.20	74.37	39.63	91.01	40.95	84.44	52.04	68.82	54.83	60.34
	8	55.20	85.40	56.71	77.74	37.16	90.51	38.53	83.65	47.95	60.58	50.06	52.32
	9	60.03	82.55	61.15	74.16	44.05	91.30	45.26	85.02	57.25	71.67	59.65	62.75
50%	0	53.41	73.46	53.92	66.09	38.59	84.64	39.20	75.39	77.29	82.41	77.76	80.24
	1	62.07	73.60	62.37	69.30	41.71	83.77	42.27	74.45	64.02	27.95	60.75	43.58
	2	56.87	64.58	57.06	60.60	39.02	84.46	39.62	75.06	57.58	37.95	55.80	47.02
	3	59.00	70.99	59.31	66.09	41.87	83.77	42.42	74.89	68.68	67.71	68.59	67.13
	4	60.97	70.86	61.22	66.56	42.25	83.81	42.80	74.64	70.05	73.94	70.40	71.39
	5	57.57	71.68	57.93	65.24	41.63	83.77	42.19	74.72	64.25	40.44	62.09	50.57
	6	52.47	70.47	52.93	63.13	42.35	84.68	42.91	76.21	79.65	85.86	80.21	83.20
	7	58.65	70.89	58.96	65.61	38.17	83.97	38.78	74.26	75.57	82.13	76.17	79.59
	8	60.21	68.56	60.43	64.19	39.95	83.01	40.51	73.54	71.58	74.68	71.86	72.98
	9	65.06	68.23	65.14	66.19	37.06	83.97	37.68	74.97	70.44	72.75	70.65	70.81
75%	0	61.38	56.01	61.29	62.34	38.23	74.93	38.56	64.34	81.59	71.79	80.98	78.62
	1	63.18	51.87	62.99	62.13	39.36	72.85	39.66	63.02	78.88	62.05	77.83	73.56
	2	61.95	52.63	61.80	61.18	40.57	74.43	40.87	64.26	77.74	65.95	77.01	73.91
	3	63.84	48.10	63.57	60.92	40.33	73.76	40.63	64.40	77.20	54.47	75.78	70.36
	4	62.31	49.46	62.10	60.86	41.52	74.05	41.80	64.45	80.47	65.99	79.56	76.49
	5	59.10	55.17	59.03	60.81	38.71	75.38	39.04	64.37	78.20	59.34	77.02	72.46
	6	60.86	56.98	60.80	61.81	40.03	74.08	40.33	64.04	77.30	60.53	76.25	71.81
	7	63.63	55.36	63.49	63.50	40.06	73.70	40.36	63.66	81.28	68.34	80.47	77.52
	8	62.06	55.62	61.95	62.24	39.52	71.95	39.81	62.31	70.73	37.47	68.65	60.83
	9	64.74	52.16	64.52	62.87	43.73	75.61	44.01	66.24	74.30	60.24	73.42	69.60