https://doi.org/10.2352/EI.2025.37.3.MOBMU-321

Enhancing Health Promotion Communication through Domain-Tailoring Techniques in ChatGPT

Ebrahim Mellatdoust Pordel; Department of Electrical & Computer Engineering, The University of Texas at San Antonio, Texas, USA
Farin Ahmed; Department of Electrical & Computer Engineering, The University of Texas at San Antonio, Texas, USA
David Akopian; Department of Electrical & Computer Engineering, The University of Texas at San Antonio, Texas, USA

Abstract

With the exponential growth of large language models (LLMs), enhancing model adaptability for diverse real-world applications has become crucial. This study critically examines domain-specific fine-tuning of ChatGPT and explores the potential of In-Context Learning (ICL) as a complementary strategy, highlighting the delicate balance between generalizability and specificity in health promotion communication. Employing two distinct fine-tuning strategies—single-prompt interactions and multi-turn conversation models—the research advances current methodologies for tailoring LLMs to specialized domains. By incorporating approaches such as data augmentation, transfer learning, and adaptive fine-tuning, alongside structured Meta-Prompting, the study systematically evaluates ChatGPT's adaptability in handling health-specific dialogues, comparing model performance across varied interaction types. Case studies and targeted customization strategies underscore the practical utility and significant impact of these adaptations in applied health communication contexts, demonstrating the enhanced contextual understanding in multi-turn interactions. Results indicate the superior efficacy of the multi-turn approach in managing nuanced, contextually rich dialogues, underscoring the capacity of the model for sustained engagement in health-related discourse. ICL with Meta-Prompting, on the other hand, demonstrates notable flexibility and resource efficiency. These findings have significant implications for advancing AI in health communication, suggesting a developmental trajectory that integrates technological sophistication with a focus on empathetic user engagement.

Index Term *- Large Language Models, Domain-Specific Fine-Tuning, Health Promotion Communication, Adaptive Fine-Tuning, Meta-Prompting, In-Context Learning, ChatGPT.*

Introduction

Advanced AI language models like ChatGPT represent a major breakthrough in how humans interact with machines and with each other. This advancement comes from refining and perfecting the principles behind Generative Pre-trained Transformers (GPT) technology [1]. These systems demonstrate remarkable capability in natural language processing (NLP) tasks, leveraging their transformer-based neural network architecture to effectively process and generate contextually appropriate textual content [2]. Most notably, ChatGPT is trained upon vast textual datasets, allowing it to notice and continue complex patterns and interactions in the language, making it indispensable for many NLP tasks from translation to content generation. Nevertheless, in the NLP context, especially in health promotion communication, methods for customizing AI models to specific domains are critically important. The fine-tuning process is essential for developing an AI system capable of delivering personalized health messages. This approach can significantly enhance the effectiveness of health communication campaigns and behavioral interventions [3]. AI models are capable of producing highly relevant language while also adapting to domain-specific outputs, such as healthcare. This adaptation involves fine-tuning the model using observation-driven data, allowing it to deliver personalized advice and support tailored to individual users [4].

The emergence of LLMs has revolutionized NLP tasks, enabling sophisticated applications in conversational AI, healthcare support, and knowledge retrieval. Among these models, In-Context Learning (ICL) has gained considerable attention for its ability to adapt to domain-specific tasks without the need for explicit parameter updates or fine-tuning. ICL leverages examples and instructions embedded within prompts to guide responses, offering a lightweight and flexible approach to task-specific performance enhancement [5]. This method leverages analogy-based reasoning, where demonstration examples formatted in natural language templates provide the context necessary for the model to generate predictions. Unlike traditional supervised learning, which demands computationally intensive parameter optimization and retraining, ICL offers a lightweight, efficient, and training-free paradigm for adapting models to task-specific requirements. Additionally, Wei et al discuss the decision-making process in ICL mirrors human reasoning by learning from analogy, enhancing its intuitive appeal for practical applications [6]. Within the ICL paradigm, Meta-Prompting emerges as a refined technique that enhances the capabilities of LLMs by structuring prompts with explicit roles, goals, and contextual cues. Unlike traditional ICL, which relies solely on providing a few examples in the prompt, Meta-Prompting integrates task-specific instructions, enabling models to better align with desired outcomes [7].

Domain adaptation in NLP is a crucial technique that enables models to effectively transfer knowledge from one domain to another. This process allows models to generalize their learned knowledge to new, previously unseen domains. The significance of domain adaptation lies in its ability to bridge the gap between training data and real-world applications, where the model may encounter diverse domains. By effectively adapting to new domains, NLP models can maintain their performance and relevance across various applications and contexts [8]. Wang et al, employed the domain adaptation techniques in ChatGPT also encompass the Supervised Fine-Tuning (SFT) Model via Proximal Policy Optimization (PPO) and Expert-Oriented Black-box Tuning. The SFT Model, implemented using the PPO algorithm, refines the SFT Policy by employing reinforcement learning techniques. This process involves optimizing the reward system to produce desirable outputs, ensuring improved performance and alignment with user expectations [9]. Expert-Oriented Black-box Tuning entails engaging domain specialists in the tuning procedure, wherein they offer feedback and recommendations to enhance the performance of AI systems like ChatGPT. This approach is guided by particular evaluation criteria established by experts to ensure domain relevance and precision. The tuning process typically includes assessing the current performance, updating the Co-Pilot's long-term memory, and performing derivative-free optimization based on expert recommendations [10]. Additionally, techniques such as AugGPT leverage ChatGPT for text data augmentation, enhancing its adaptability across various applications. These augmentation methods aim to expand and diversify training datasets, further strengthening ChatGPT's contextual understanding and adaptability [11].

In recent years, the rapid advancement of LLMs has introduced various strategies for task adaptation, notably ICL and fine-tuning. Within the ICL framework, Meta-Prompting addresses limitations of standard ICL, such as its sensitivity to prompt design and example selection, and significantly improves the adaptability of LLMs for dynamic tasks. In contrast, fine-tuning involves adjusting the model's parameters using task-specific data to enhance performance on designated tasks. While ICL offers flexibility and ease of implementation, emerging research indicates that fine-tuning often yields superior performance. For instance, Mosbach et al. conducted a comprehensive comparison and found that fine-tuned language models can generalize well out-of-domain, challenging the notion that ICL inherently provides better generalization [12]. Similarly, Liu et al. demonstrated that few-shot parameter-efficient fine-tuning not only surpasses ICL in accuracy but also incurs lower computational costs. These findings suggest that, despite the convenience of ICL, fine-tuning remains a more effective approach for achieving optimal performance in specific tasks [13].

This study aims to comprehensively evaluate the relative effectiveness of ICL and fine-tuning strategies in adapting LLMs for domain-specific tasks, focusing on single-prompt and multi-turn conversations separately. The emphasis on resource-constrained datasets highlights the practical challenges associated with such settings and explores how these constraints impact the performance and adaptability of each method. By conducting a comprehensive analysis, the study seeks to determine which method offers superior performance in terms of accuracy, generalization, and resource efficiency. This investigation addresses the ongoing debate in the NLP community regarding the optimal approach for task adaptation, providing insights that could inform future applications and model development [14].

The primary objective is to determine which method yields superior performance in generating contextually appropriate and semantically accurate responses across both single-prompt and multi-turn conversational settings. This investigation seeks to address the following research questions:

- Performance Comparison: How do ICL and fine-tuning differ in terms of their ability to generate accurate and contextually relevant responses in both single-prompt and multi-turn conversations?
- Generalization Capability: To what extent can models adapted through ICL or fine-tuning generalize to unseen conversational contexts, particularly in specialized domains such as smoking cessation support?
- Resource Efficiency: What are the computational and data requirements associated with each adaptation method, and how do these impact their practicality for real-world applications?

By systematically comparing ICL and fine-tuning across these dimensions, the study aims to provide insights that inform the selection of adaptation strategies for deploying language models in complex, domain-specific conversational tasks.

Recent research highlights the superior accuracy and computational efficiency of fine-tuning approaches, especially in low-resource scenarios and complex downstream tasks. These observations emphasize the need for further investigation into the comparative efficacy of these methods to optimize their deployment across diverse machine-learning applications [15].

In this study, we explore the use of ICL combined with Meta-Prompting and compare it with a fine-tuned model using a dataset that includes question-and-answer (Q&A) pairs in the smoking cessation domain. Meta-Prompting involves structuring prompts with explicit roles, goals, and context, enabling the model to deliver domain-relevant outputs. This method addresses the limitations of zero-shot prompting by integrating task-specific instructions and clarifications directly within the prompt [14]. The motivation for employing Meta-Prompting stems from its ability to handle multi-turn conversations, where maintaining context and coherence is critical. Doimo et al. has demonstrated that prompt engineering techniques, including Meta-Prompting, can significantly improve language model performance by optimizing input-output alignment [16]. We evaluate the proposed method using a dataset of smoking cessation-related Q&A pairs and assess performance with metrics such as BLEU, ROUGE, and BERTScore. Results highlight the model's ability to capture semantic relevance and produce motivational responses, although challenges remain in lexical alignment and structural coherence.

Methodology

This study employs a mixed-method research design which incorporates both quantitative and qualitative approaches to evaluate the performance of ICL and fine-tuning methods for smoking cessation support. Lexical and semantic performance are evaluated quantitatively using statistical assessment measures such as BERTScore, ROUGE, and BLEU. Meanwhile, the qualitative analysis examines contextual relevance, coherence, and user engagement through manual evaluation. This structure enables a comprehensive assessment of both models, utilizing numerical metrics and qualitative observations to determine strengths and weaknesses in their performance. This combination enables a thorough assessment of the technical performance and practical applicability.

The dataset utilized in this study comprises Q&A pairs specifically focused on smoking cessation. The data were sourced from credible repositories, including publicly available health databases such as the Centers for Disease Control and Prevention (CDC). It encompasses a diverse set of queries addressing behavioral strategies, pharmacological interventions, and motivational techniques. The dataset was split into training and validation subsets, with the training set used to design prompts for Meta-Prompting and fine-tuning, and the validation set used for performance evaluation. Selection criteria prioritized relevance to smoking cessation topics and clarity of responses.

The study followed a systematic procedure to implement and evaluate the proposed methodology. Initially, prompts were developed and refined using the training dataset, incorporating contextual instructions and task-specific objectives. These prompts were then tested with ChatGPT-4o-mini in both Meta-Prompting and fine-tuned configurations. The study employed two distinct experimental configurations. Two distinct experimental setups were utilized. The first involved ICL with Meta-Prompting, where prompts were carefully designed to define the assistant's role, establish clear objectives, and address ambiguous queries. The second configuration employed a fine-tuned variant of the model, trained through supervised learning on the same dataset. Responses generated from these configurations were assessed using both quantitative and qualitative methods. Quantitative evaluations utilized BLEU, ROUGE, and BERTScore metrics to measure lexical overlap, sequence alignment, and semantic accuracy. Qualitative analysis focused on thematic analysis of coherence, user engagement, and usability, ensuring the model's practical effectiveness.

Model Setup

ChatGPT-4o-mini is implemented as the underlying LLM, leveraging its capabilities for natural language understanding and generation due to its optimal balance between computational efficiency and performance. Its streamlined architecture enables us to effectively implement both fine-tuning and ICL with Meta-Prompting, while significantly reducing resource demands compared to larger models. The model was evaluated in two configurations:

In-Context Learning with Meta-Prompting

The prompts were designed to:

- Define the role and expertise of the assistant.
- Specify goals for concise, motivational, and actionable responses.
- Context and clarifying questions for ambiguous queries.

Fine-Tuned Model

The fine-tuned configuration employed supervised learning to adapt ChatGPT-4o-mini for smoking cessation counseling. Separate training and validation datasets were used for single-prompt and multi-turn conversation configurations to evaluate different interaction patterns.

Meta-Prompt Design

The Meta-Prompting framework was meticulously structured to achieve the following objectives:

- Establish the assistant's role as a certified smoking cessation counselor.
- Include instructions for handling incomplete or ambiguous queries.
- Incorporate motivational strategies for smoking cessation based on behavioral and pharmacological interventions.
- Maintain context coherence across multi-turn conversations.

The prompt template adopts structured input formatting to ensure that the assistant recognizes its predefined role, adheres to task-specific instructions, and maintains a professional tone throughout multi-turn interactions. The model was evaluated separately in two configurations: single prompt and multiturn conversation. After designing structured prompts based on the training dataset. These prompts are crafted to align with the goals of the study, ensuring relevance to the smoking cessation domain. Next, responses are generated using two configurations: Meta-Prompting and fine-tuned model approaches. The generated predictions are then evaluated using BLEU, ROUGE, and BERTScore metrics to measure lexical similarity, semantic relevance, and alignment with the ground truth. Following this, the performance metrics are compared to analyze coherence, relevance, and contextual consistency. Finally, qualitative assessments are conducted to evaluate response usability and consistency in real-world applications. The evaluation further examines the model's ability to sustain coherence in multi-turn conversations, ensuring that follow-up questions and clarifications align with previous exchanges to simulate realistic counseling scenarios. Insights gained from the evaluations help refine the model's response quality, focusing on the flow and its capacity to provide actionable guidance.

Fine-Tuning Model Design

The training dataset included Q&A pairs, pre-processed to ensure consistency in formatting and content. Fine-tuning was conducted using OpenAI's API, leveraging the GPT-4o-mini architecture. Separate datasets were prepared for single-prompt and multi-turn conversations, each optimized to capture specific interaction patterns. Key hyperparameters such as batch size and learning rate were tuned. Early stopping criteria were implemented to prevent overfitting, and dropout layers were utilized to enhance generalization. The validation dataset, comprising Q&A pairs, was used to test model generalization and assess performance on unseen data. The evaluation phase included performance monitoring through BLEU, ROUGE, and BERTScore metrics to quantify lexical precision, coherence, and contextual consistency. The fine-tuned model was configured to output concise, motivational, and actionable responses, aligning closely with expert-generated references. Multi-turn interactions were tested to verify the model's ability to maintain contextual relevance and coherence across exchanges. The fine-tuned model was optimized for scalability and deployment, ensuring compatibility with web-based applications and research prototypes.

Evaluation Metrics

The effectiveness of the proposed framework is evaluated using the following metrics:

BLEU: Measures n-gram overlap to evaluate lexical similarity.

ROUGE: Captures recall-oriented overlap at unigram and bigram levels and sequence matching.

BERTScore: Evaluates semantic similarity through contextual embeddings.

Additionally, performance was analyzed across specific categories, such as motivational content, factual accuracy, and response engagement, to provide a multidimensional assessment of model behavior. The combination of these metrics ensures a comprehensive evaluation of both the surface-level text and the underlying semantic quality of the generated responses.

Results

The key evaluation metrics for Meta-Prompting and the fine-tuned model, including BLEU scores, ROUGE scores, and BERTScore F1 values, are summarized in Table 1.

Table 1: Evaluation Metrics for Meta-Prompting and Fine-Tuned Model

Model Type	BLEU	ROUGE-L	BERTScore
Meta-Prompting (Single Prompt)	0.2395	0.2077	0.84
Meta-Prompting (Multiturn Conversation)	0.4246	0.4179	0.87
Fine-Tuning (Single Prompt)	0.2715	0.2311	0.88
Fine-Tuning (Multiturn Conversation)	0.4751	0.4599	0.89
Baseline ChatGPT-4o-mini (Single Prompt)	0.0082	0.1120	0.79
Baseline ChatGPT-4o-mini (Multiturn Conversation)	0.1165	0.1917	0.81

The baseline ChatGPT-4o-mini (Single Prompt) model shows limited capability, with a BLEU score of 0.0892, reflecting short, underdeveloped responses. Its multi-turn counterpart improves slightly to 0.165, suggesting that incorporating additional turns provides some benefit even without specialized prompting or fine-tuning.

The BLEU score evaluation highlights a clear performance distinction between single QnA and multi-turn models. Under ICL, the single-prompt model achieves a BLEU of 0.2395, indicating moderate lexical overlap with reference answers. Estimated n-gram precisions (35% for unigrams and 20% for bigrams) and a brevity penalty of 0.89 point to somewhat concise outputs that still miss extended context. By contrast, the multi-turn ICL model attains a higher BLEU of 0.4246, supported by stronger n-gram precisions (48% for unigrams, 29% for bigrams) and a length ratio closer to the references (brevity penalty 0.85). These improvements highlight the advantage of multi-turn setups in capturing context over successive exchanges.

Turning to fine-tuning, the single-prompt model registers a BLEU of 0.2715, modestly surpassing the single-prompt ICL model. However, the multi-turn fine-tuned model achieves the highest BLEU overall at 0.4751, reflecting robust lexical and structural coherence. Its higher n-gram precision (53% for unigrams, 34% for bigrams) and balanced length ratio underscore the benefits of domain-targeted training, particularly in multi-turn conversations.

ROUGE metrics confirm the BLEU trends. The baseline single-prompt model records ROUGE-1: 0.20, ROUGE-2: 0.06, and ROUGE-L: 0.1120, indicating limited recall of key words and phrases. Multi-turn baseline results (ROUGE-1: 0.28, ROUGE-2: 0.12, ROUGE-L: 0.1917) show slight improvement but remain relatively low.

Under ICL, the single-prompt configuration posts ROUGE-1: 0.31, ROUGE-2: 0.15, and ROUGE-L: 0.2077, reflecting moderate coverage of reference text structures. The multi-turn meta-prompting model, however, improves substantially to ROUGE-1: 0.50, ROUGE-2: 0.34, and ROUGE-L: 0.4179, capturing more of the content and maintaining better sentence flow across multiple turns.

Fine-tuning yields a similar pattern. The single-prompt fine-tuned model sees ROUGE-1: 0.35, ROUGE-2: 0.18, and ROUGE-L: 0.2311, a modest step up from single-prompt ICL. In contrast, the multi-turn fine-tuned model achieves the highest ROUGE scores overall (ROUGE-1: 0.57, ROUGE-2: 0.38, ROUGE-L: 0.4599), demonstrating effective handling of sequential dependencies and nuanced text structures.

BERTScore provides a deeper view of semantic alignment. For baseline ChatGPT-4o-mini, single-prompt responses yield a BERTScore of 0.79 (precision: 0.78, recall: 0.80, F1: 0.79), while multi-turn improves slightly to 0.81 (precision: 0.80, recall: 0.82, F1: 0.81). Under meta-prompting, the single-prompt model scores 0.84 overall (precision: 0.83, recall: 0.85, F1: 0.84), whereas the multi-turn model jumps to 0.87 (precision: 0.86, recall: 0.88, F1: 0.87), reflecting better contextual understanding and nuanced responses.

Fine-tuning boosts semantic similarity further. The single-prompt fine-tuned model achieves a BERTScore of 0.89 (precision: 0.88, recall: 0.90, F1: 0.89), indicating strong alignment with reference answers. Notably, the multi-turn fine-tuned model also reports 0.89 (precision: 0.89, recall: 0.89, F1: 0.89), matching the single-prompt precision but preserving context over multiple exchanges. These results highlight the efficacy of fine-tuning in refining embeddings for domain-specific content.

The multi-turn variants consistently outperform their single-prompt counterparts, emphasizing the importance of sustained context in generating coherent, accurate, and semantically rich answers. Meanwhile, fine-tuned models provide incremental gains over ICL, especially in multi-turn settings, underlining the value of targeted training for domain-specific tasks. Taken together, these findings demonstrate that combining multi-turn dialogue structures with fine-tuning strategies yields the most robust performance in both lexical fidelity and semantic alignment.

Qualitative Assessment

Engagement and relevance: The single-prompt model provides accurate but static responses, making it less engaging in interaction. In contrast, the multi-turn model actively engages users and transitions between topics in a meaningful way, enhancing the conversational experience. Fine-tuning further improves engagement by delivering more contextually aligned responses that feel natural and interactive compared to ICL.

Contextual awareness: The single-prompt model treats each query as an isolated interaction, showing limited ability to maintain continuity. The multi-turn model, however, builds upon previous inputs to create coherent and contextually aware dialogues. Fine-tuned models outperform ICL in maintaining context over multiple turns, enabling deeper and more responsive dialogues.

User intent understanding: While the single-prompt model effectively addresses explicit queries, it often misses nuanced or underlying user intents. The multi-turn model excels at recognizing and addressing these subtler intents. Fine-tuning enhances intent recognition, especially in complex queries, resulting in responses that better address implicit needs.

Guidance and support: The single-prompt model offers basic information but lacks detailed, step-by-step guidance. On the other hand, the multi-turn model provides comprehensive guidance and anticipates user needs for a more supportive interaction. Fine-tuning models provide even more structured and detailed responses, making them ideal for scenarios requiring multi-step instructions or elaborate problem-solving approaches.

Conversational depth: The single-prompt model is limited to factual answers and lacks motivational or deeper conversational layers. The multi-turn model engages in deeper, motivational, and meaningful discussions. Fine-tuning provides a higher level of depth by leveraging domain-specific training data, enabling responses that integrate behavioral strategies and motivational cues more effectively than ICL.

Conversational coherence and flow: The single-prompt model provides responses that are relevant but lack logical sequencing. Conversely, the multi-turn model ensures consistent and logical conversational flow, making interactions smoother and more natural. Fine-tuning surpasses ICL by maintaining better coherence and continuity across multi-turn conversations, reducing redundancies, and improving the clarity of responses.

Discussion

The results demonstrate that fine-tuning outperforms ICL with Meta-Prompting across all evaluation metrics. Multi-turn configurations consistently performed better than single-prompt setups, highlighting their ability to capture conversational context and improve response quality. The results reinforce the importance of context retention and task-specific optimization in conversational AI, advancing theories on prompt engineering and transfer learning. These insights can inform the development of AI-driven counseling tools for healthcare and behavioral interventions, providing scalable and effective solutions.

Several recent studies exploring ICL have utilized GPT-4o. Although fine-tuning GPT-4 is accessible through an API, the specific fine-tuning method employed via the API remains unclear. Comparable large models are publicly available; however, the computational resources required to implement them exceeded the scope of this study. While we anticipated that the patterns observed in this research would extend to larger models, empirical verification was not feasible. Moreover, all experiments were conducted using English language models due to the lack of publicly available datasets of similar quality in other languages. Lastly, we focused exclusively on basic fine-tuning and ICL techniques, even though more sophisticated approaches, such as calibration, may enhance performance for both methods. It is worth noting that such techniques can generally be applied to either approach, implying that advancements in one are likely to benefit the other.

Future work could incorporate user feedback, test real-time deployment scenarios, and explore hybrid approaches combining fine-tuning with prompt engineering. Moreover, a detailed error analysis of the generated responses could pinpoint specific challenges in context retention and semantic accuracy, guiding targeted improvements in both fine-tuning and ICL approaches.

Conclusion

This study investigated and compared the effectiveness of ICL with Meta-Prompting and fine-tuning approaches for generating responses in the smoking cessation domain. Key findings highlight that multi-turn models consistently outperformed single-prompt models across all evaluation metrics, including BLEU, ROUGE, and BERTScore. Fine-tuning demonstrated superior performance, particularly in handling multi-turn conversations, capturing contextual nuances, and maintaining semantic coherence.

The primary contributions of this research include a systematic comparison of ICL and fine-tuning approaches, providing insights into their respective strengths and limitations. The study emphasizes the importance of leveraging fine-tuning for tasks requiring domain-specific knowledge and multi-turn conversational depth, while also acknowledging the utility of ICL for simpler, single-prompt interactions.

In conclusion, this research underscores the need for ongoing exploration of advanced prompting techniques and fine-tuning strategies to further enhance model performance. Future studies should investigate the integration of more sophisticated methodologies, such as calibration and hybrid approaches, to improve both ICL and fine-tuning capabilities. Researchers are encouraged to adapt these methods to broader domains and diverse datasets, addressing evolving challenges in conversational AI systems.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 2017, pp. 5998–6008.

[2] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems 33 (NeurIPS 2020), Virtual Conference, 2020, pp. 1877–1901.

[3] S. Lim and R. Schmälzle, "Artificial Intelligence for Health Message Generation: Theory, Method, and an Empirical Study Using Prompt Engineering," arXiv preprint arXiv:2212.07507, 2022.

[4] J. Cosentino, A. Belyaeva, X. Liu, N. A. Furlotte, Z. Yang, C. Lee, E. Schenck, Y. Patel, J. Cui, L. D. Schneider, R. Bryant, R. G. Gomes, A. Jiang, R. Lee, Y. Liu, J. Perez, J. K. Rogers, C. Speed, S. Tailor, M. Walker, J. Yu, T. Althoff, C. Heneghan, J. Hernandez, M. Malhotra, L. Stern, Y. Matias, G. S. Corrado, S. Patel, S. Shetty, J. Zhan, S. Prabhakara, D. McDuff, and C. Y. McLean, "Towards a Personal Health Large Language Model," arXiv preprint arXiv:2406.06474, 2024.

[5] Q. Dong, L. Li, D. Dai, C. Zheng, J. Ma, R. Li, H. Xia, J. Xu, Z. Wu, B. Chang, X. Sun, L. Li, and Z. Sui, "A Survey on In-context Learning," in Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pp. 1107–1128, Miami, Florida, USA, Association for Computational Linguistics, 2024.

[6] J. W. Wei, L. Hou, A. K. Lampinen, X. Chen, D. Huang, Y. Tay, X. Chen, Y. Lu, D. Zhou, T. Ma, and Q. V. Le, "Symbol tuning improves in-context learning in language models," in Proceedings of *the 2023 Conference on Empirical Methods in Natural Language Processing*, Singapore, Dec. 2023, pp. 968–979. Association for Computational Linguistics.

[7] M. Suzgun and A. T. Kalai, "Meta-Prompting: Enhancing Language Models with Task-Agnostic Scaffolding," arXiv preprint arXiv:2401.12954, 2024.

[8] W. Lin, Q. Lin, L. Feng, and K. Tan, "Ensemble of Domain Adaptation-Based Knowledge Transfer for Evolutionary Multitasking," IEEE Trans. Evol. Comput., vol. 28, pp. 388-402, 2024.

[9] B. Wang, R. Zheng, L. Chen, Y. Liu, S. Dou, C. Huang, W. Shen, S. Jin, E. Zhou, C. Shi, S. Gao, N. Xu, Y. Zhou, X. Fan, Z. Xi, J. Zhao, X. Wang, T. Ji, H. Yan, L. Shen, Z. Chen, T. Gui, Q. Zhang, X. Qiu, X. Huang, Z. Wu, and Y. G. Jiang, "Secrets of RLHF in large language models part II: Reward modeling," arXiv preprint arXiv:2401.06080, 2024.

[10] S. Wang, Y. Zhu, Z. Li, Y. Wang, L. Li, and Z. He, "ChatGPT as your vehicle co-pilot: An initial attempt," IEEE Transactions on Intelligent Vehicles, 2023.

[11] H. Dai, Z. Liu, W. Liao, X. Huang, Y. Cao, Z. Wu, L. Zhao, S. Xu, W. Liu, N. Liu, S. Li, D. Zhu, H. Cai, L. Sun, Q. Li, D. Shen, T. Liu, and X. Li, "AugGPT: Leveraging ChatGPT for text data augmentation," arXiv preprint arXiv:2302.13007, 2023.

[12] M. Mosbach, T. Pimentel, S. Ravfogel, D. Klakow, and Y. Elazar, "Few-shot Fine-tuning vs. In-context Learning: A Fair Comparison and Evaluation," Findings of the Association for Computational Linguistics: ACL 2023, pp. 12284–12314, 2023.

[13] H. Liu, D. Tam, M. Muqeeth, J. Mohta, T. Huang, M. Bansal, and C. Raffel, "Few-Shot Parameter-Efficient Fine-Tuning is Better and Cheaper than In-Context Learning," arXiv preprint arXiv:2205.05638, 2022.

[14] Q. Yin, X. He, C. T. Leong, F. Wang, Y. Yan, X. Shen, and Q. Zhang, "Deeper Insights Without Updates: The Power of In-Context Learning Over Fine-Tuning," in Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, 2024, pp. 4138–4151.

[15] D. Doimo, A. P. Serra, A. Ansuini, and A. Cazzaniga, "The Representation Landscape of Few-Shot Learning and Fine-Tuning in Large Language Models," in Proceedings of the Thirty-Eighth Annual Conference on Neural Information Processing Systems, 2024.

[16] L. Wang, W. Xu, Y. Lan, Z. Hu, Y. Lan, R. K.-W. Lee, and E.-P. Lim, "Plan-and-Solve Prompting: Improving Zero-Shot Chain-of-Thought Reasoning by Large Language Models," in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, Canada, 2023, pp. 2609–2634.

Author Biography

Ebrahim Mellatdoust Pordel is currently a PhD student in the Electrical and Computer Engineering department at the University of Texas at San Antonio where his research focuses on Machine Learning and AI, and large language models. His work centers on developing innovative algorithms that bridge theoretical research with real-world applications.

Farin Ahmed received her BS in Electrical Engineering from BRAC University, Bangladesh and MS in Electrical Engineering from University of Texas at San Antonio. Her research interest includes wearable health-care device using Internet of Things, AI Domain Adaptation, and also hardware-software integration.

David Akopian received his Ph.D. degree from Tampere University, Finland. He is currently a Professor with The University of Texas at San Antonio (UTSA). Before joining UTSA, he was a senior scientist at Nokia. His current research interests include mobile computing and positioning, He is a Fellow of US National Academy of Inventors