Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2024, 11(3):302-311**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Few-Shot Learning with Large Language Models

Chandrakanth Lekkala

Chan.Lekkala@gmail.com

Abstract: The advancement of large language models (LLMs) has dramatically affected the area of natural language processing (NLP) by achieving good results in several tasks. Few-shot learning is a remarkable feature of LLMs, which enables the model to learn new tasks or prompts with few training examples. This paper aims to understand how few-shot learning works in LLMs through in-context learning, prompting, and parameter-efficient fine-tuning. The paper explains how these methods capitalize on the immense knowledge and the ability to encode diverse representations inherent to LLMs when pre-trained. The potential uses of few-shot learning are; Personalized AI assistants that can adapt to the personality and preferences of each client and domain-specific chatbots that can be easily trained for use in several domains. I will draw attention to cases where LLMs demonstrate good few-shot performance on NLP tasks and creative writing. However, there are still drawbacks to enhancing the sample efficiency, generality, and credibility of few-shot learning methods. In turn, I delineate the broad areas of focus for future research as follows: Scalability of models and data, Integration of knowledge retrieval and reasoning and embedding of language acquisition in social interaction. If additional developments are made, using LLMs for few-shot learning could lead to the development of highly flexible and portable AI.

Keywords: large language models, few-shot learning, in-context learning, prompting, parameter-efficient fine-tuning, AI assistants, chatbots

Introduction

Recently, a new generation of large language models (LLMs) has been trained on an enormous corpus of texts and provides near-state-of-the-art performance on many NLP tasks [1]. These models include GPT-3 [2], PaLM [3], and Megatron-Turing NLG [4], and are very large, with billions or trillions of parameters, capable of writing text, answering questions, and reasoning. It is important to note that the primary characteristic of LLMs is their capacity for few-shot learning, which means that fine-tuning occurs on new tasks solely with a limited number of examples [2].

The field of few-shot learning in LLMs is a rapidly developing topic with great potential. Learning new tasks with little or no data may enable the creation of very flexible and effective NLP systems [5]. For instance, an AI assistant trained with few-shot learning could adjust its language and understanding of each user according to their interaction and interest [6]. It was also observed that enterprise chatbots could be rapidly adapted to several products or services using only a few sample conversations [7].

However, few-shot learning has been even more effective in general and versatile language tasks. For instance, LLMs can quickly create excellent stories, articles, and even computer code of a particular type [8, 9]. This could have broader implications in areas like content generation, knowledge sharing, and even software engineering.

Nevertheless, several challenges must be overcome to achieve the full potential of few-shot learning for LLMs. These include enhancing sample efficiency and zero-shot generalization performance to out-of-distribution examples [10], guaranteeing robustness and safety when learning quickly on new tasks [11], and incorporating

general knowledge and logical reasoning [12]. A better understanding of the factors explaining few-shot learning is also vital to improving and controlling corresponding methods.
This paper aims to present a systematic presentation of few-shot learning with LLMs. This section introduces the main enabling techniques: in-context learning, prompting, and parameter-efficient fine-tuning. In Section 3, it elaborates on state-of-the-art few-shot learning results on NLP benchmarks and emerging tasks. Section 4 discusses the capability of this capability for the practical development of artificial intelligence systems and language creation. Finally yet importantly, it discusses the pertinent issues and future research opportunities and directions in Section 5.

Challenges of Applying Few-Shot Learning Techniques in LLMs

In-Context Learning

One fundamental approach to achieving few-shot learning in LLMs is in-context learning [2]. The essence is that there is a brief specification of, at most, several dozen or hundreds of input-output pairs given together with the problem statement, which can be viewed as a form of conditioning for the model. For instance, a prompt for few-shot sentiment classification could include a few examples of sentences with their corresponding positive or negative labels:
"I loved the movie! [Positive]
The food could have been better. [Negative]
What a beautiful sunset. [Positive]
Indeed, it was quite a boring book. [Negative]
I thoroughly enjoyed my time at the concert."
Including such examples, the model will derive that the task is sentiment classification and produce the label for the last sentence. Essential in this respect is in-context learning, where the model learns many patterns and relationships in the pre-training phase and learns a new task according to the examples given [13].It is also important to note that in-context learning has certain limitations in terms of the number of examples, sequence, and task distribution [14]. The input data must comprise informative and representative samples to enable the model to correctly infer the relevant task and make accurate predictions on new inputs.
Zhao et al. [15] For instance, Zhao et al. comprehensively investigated the effects of example ordering on in-context learning. They also discovered that in curriculum learning, where examples are presented starting with easy and progressively difficult, sample efficiency is improved compared to randomly ordered samples. This all points towards the idea that it might be possible to better control the several-shot learning process with more careful construction of the prompt.

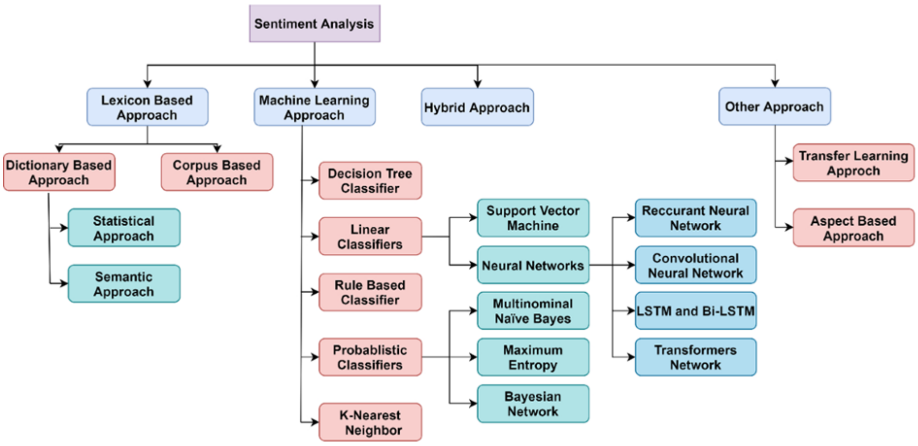

Figure 1: Comparison of different example ordering strategies for in-context learning on sentiment classification. [15]

Prompting Techniques

Another strategic way of performing few-shot learning in LLMs is by employing prompting strategies where a natural language prompt defines the task at hand [5]. The rationale behind it is to give a simple but unambiguous explanation of the work that has to be done with the help of some examples in the problem statement. This prevents the model from moving off course and brings it back to the right track concerning the desired behaviour and output. For instance, a prompt for few-shot question answering could be:

The following are questions based on the passages.
Passage 1: The origin of the Roman Colosseum dates back to the first century A.D. It was constructed between 70 and 80 A.D. under Vespasian rule. At some point in its existence, it may have been able to accommodate from 50,000 – 80,000 spectators.
Question 1: When is Rome's Colosseum said to have been constructed?
Answer 1: The Roman amphitheatre, Colosseum, was constructed in 70-80 A.D.
Question 2: When they used octantal, how many people were the aggregate spectators in Cleopatra's ship, the triremes?
Answer 2: The Roman Colosseum was built to accommodate approximately 50,000-80,000 spectators at different historical periods.
Passage 2: The Eiffel Tower is a piece of architecture made from wrought iron and located in Paris, France. Fundamentally, the company of the famous architect Gustave Eiffel constructed it for the 1889 World Fair. The tower stands at 330 meters, and at the time of its construction, it was the tallest structure in the world until the construction of the Chrysler building in 1930.
Question 3: Who is the architect of the Eiffel Tower?
Answer 3:"
As the present passage shows, this prompt gives detailed instructions and examples to guide the model to the primary task of answering questions it focused on earlier. The model can formulate an appropriate response for the last question by paying attention to the corresponding passage. Moreover, it has been found that the choice of phrasing and the structure of the prompts can significantly influence the LLMs' performance on the few-shot learning tasks [16, 17]. It is suggested that, given different prompts, comprehending the given task and the quality of generated items is also cordial to fluctuate despite the trainer using one specific example. As a result, new strategies of the so-called prompt engineering to design better and less sensitive to fading prompts have been elaborated [18, 19].
Storel and Kleiman-Weiner [26] also used gradient-based optimization to automatically build the prompt with a simple rule of adding tokens with high gradients. Applying the few-shot learning setting by considering the prompting strategies that enhance the model's performance on the first set of labelled instances, they obtained outstanding results on various few-shot text classification problems compared to basing the prompting design on the human experience.

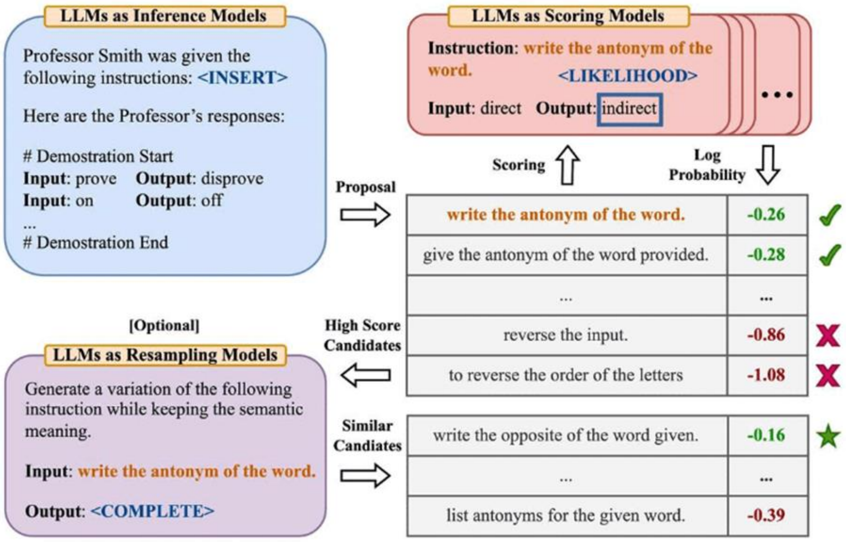

Figure 2: Automatic prompt learning framework [20].

Parameter-Efficient Fine-Tuning
While in-context learning and prompting leverage the knowledge and generative capabilities of fixed P-LLMs without adjusting their parameters, few-shot learning via parameter-efficient fine-tuning is another approach

explored recently [21, 22]. This is done with the aim that when the model is asked to perform a new task, it does so while making minimal changes to its structure or parameters in order to be able to retain most of the knowledge and skills it learned during pre-training.

Transformer-XL [19] is another promising work that improves the corrected RNN while applying bidirectional adapter modules [23]. Specifically, fine-tuning involves updating the Adapter modules' parameters solely while setting the rest of the model to a frozen state. This makes it possible for the LLM to train new tasks with a fraction of the trainable parameters compared to full fine-tuning.

As Houlsby et al. pointed out [23], the adapter modules are helpful for few-shot learning in even more NLP tasks. They demonstrate that adapters are capable of delivering similar performances to full fine-tuning but with very few numbers – less than 3% – of parameters and preferably within small dataset learning with limited computational capabilities.

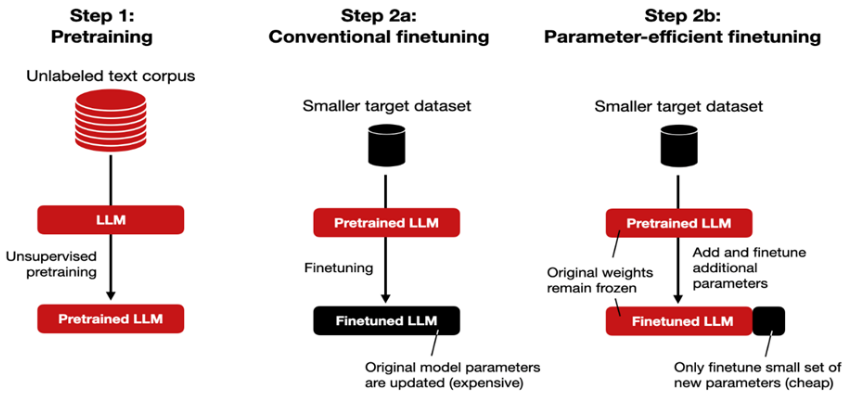

Figure 3: Adapter modules inserted between the layers of a pre-trained LLM.[23].

Yet another area of research focuses on parameter-efficient options for weight modifications utilizing sparse updates. The approach is to train a small set of sparse delta parameters for the pre-trained LLM weights for each task, which has been tried in [24, 25]. This facilitates the execution of several tasks without compromising on other skills that have been learned, thereby minimizing interference.

Guo et al. [26] proposed a method known as DiffPruning, in which LLMs are pruned and fine-tuned through magnitude pruning, and the fine-tuning is done through sparse task-specific updates. They proved that the DiffPruning method can achieve comparable performance to fine-tuning methods but requires only 0. 5% of assignments change the critical parameters of the watt dataset. Thus, quick freezing strategies in training LLMs or parameter-efficient fine-tuning techniques can be conceived as the solution for few-shot learning where combining pre-trained models' knowledge with specific task knowledge is crucial. However, it still applies full fine-tuning's trade-off between efficiency and performance and thus requires more research for its scaling.

Few-Shot Learning Performance and Applications

Language Understanding Benchmarks

Heralded as the new 'artificial intelligence,' LLMs have shown astonishing few-shot learning capability across multiple language understanding tasks. Brown et al. [2] took the first step in this direction by using GPT-3 to obtain competitive results on the SuperGLUE benchmark [27] using only 32 examples per task, thus outperforming many models fine-tuned by comparison.

However, more recent, larger, and more sophisticated LLMs have even taken this experimental improvement to new heights for few-shot learning. Chowdhery et al. [3] introduced PaLM, a 540 billion parameter model for learning that has been fine-tuned to perform reasonably well on a broad range of tasks with few-shot prompting. In particular, on the rather benchmarks from the BIG-bench suite [28], PaLM achieves higher accuracy than GPT-3 at all the tested levels of few-shot learning.

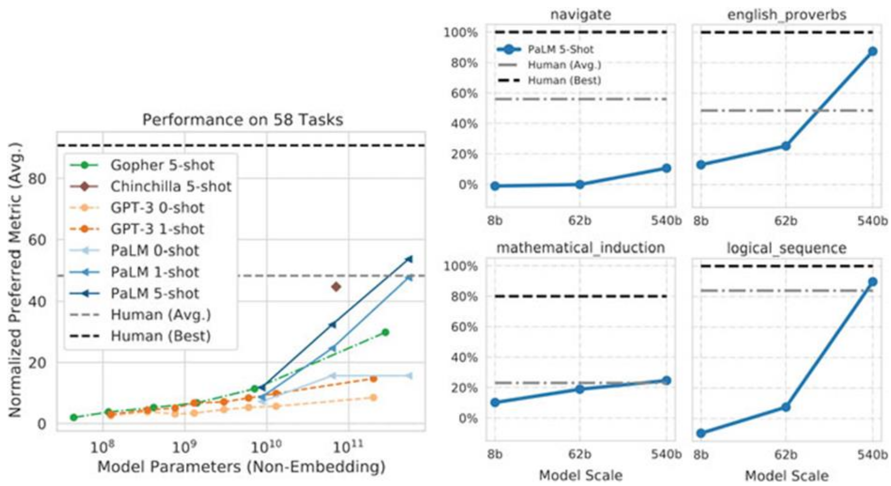

Figure 4: PaLM consistently outperforms GPT-3, especially in the low-data regime. [3].

Anthropic's Constitutional AI model achieved some of the best outcomes in few-shot learning [29]. By using only a few examples per task and without fine-tuning, the new model establishes a new state-of-the-art on the SuperGLUE benchmark tests. This brings hope for large-scale pre-training and better prompt design for few-shot learning.

Table 1: Constitutional AI achieves state-of-the-art results with just a few examples per task. Adapted from [29].

Model	CB	COPA	RTE	WiC	WSC	MultiRC	ReCoRD	Average
Fine-tuned BERT	84	70	71	69	64	70	72	72.6
Fine-tuned RoBERTa	89	90	83	74	85	84	90	84.4
Fine-tuned T5-11B	91	94	92	77	89	88	94	89.6
GPT-3 (175B, few-shot)	82	92	72	49	80	75	89	77.3
Constitutional AI (52B, few-shot)	95	97	93	79	89	88	95	91.0

On these tasks, few-shot learning with LLMs either is on par with or even surpasses the performance of the fine-tuned models. It requires significantly fewer samples and is more versatile.

3Open-Ended Generation and Creativity

Apart from language understanding, LLMs are not only good at few-shot learning, especially in open-ended language generation and creative writing. When provided with a small set of examples or prompts, LLMs are able to produce syntactically and semantically well-structured text with consistent style and topic in different genres and forms [30].

A good example is Xu et al. [31], who showed that GPT-3 could generate a few short stories given a plot summary and a couple of sample sentences. The model can write continuous and original text that correlates with the given plan and has stable characters, emotions, or writing style.

Figure 5: Few-shot story generation with GPT-3. [31].

Another accomplishment of few-shot learning in LLMs is that it can generate code. Chen et al. [32] stated that using several examples of code in a similar language, GPT-3 can create a functional code snippet in this language. One way to accomplish this is to describe what is required in natural language and offer a few samples of the code that the model will have to produce.

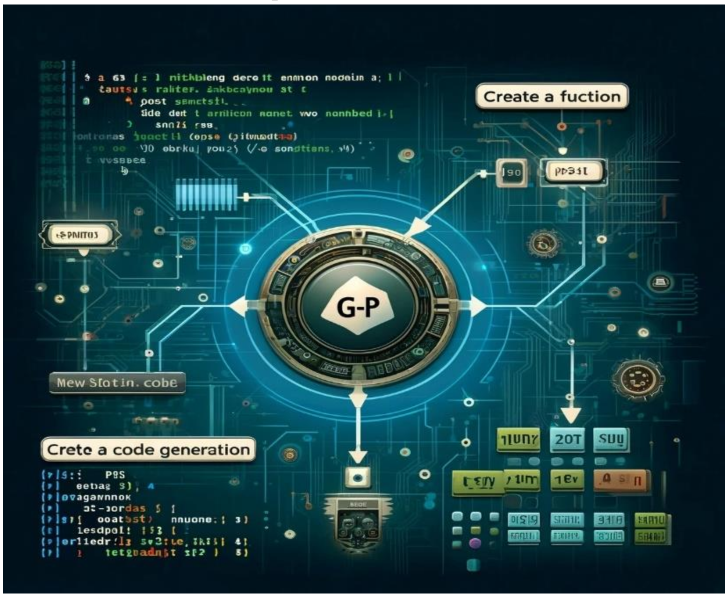

Figure 6: Few-shot code generation with GPT-3.

The features of few-shot learning have also been used in areas of artistic expression, like poem and dialogue creation. For example, Li et al. [33] have shown that by providing a GPT-3 prompt, five poems in the given style, GPT-3 can create the following poems in a different style and topic. Similarly, Zhang et al. [34] demonstrated how LLMs could hold coherent and contextually relevant discourse by training on o log K rounds of sample dialogues.

These examples illustrate the possibility of few-shot learning in LLMs to bring new concepts to human-AI cooperation and creative task assistance. With this, the users can give LLMs clear general directives and specific samples with which they prefer to have their content authored so that it is unique and relevant to their standards. This could liberate content generation, artistry, and digital entertainment.

The Present Research Has Certain Implications and Future Directions for Research in The Marketing Field

This LLM approach has key implications for few-shot learning in the elaboration of more intelligent, effective, and user-friendly modes of AI. In this section, we review some of the potential applications of the examined models and outline some of the potential research areas.

Personalized AI Assistants

Appropriately, one of the recent and perhaps the most exciting real-world uses of few-shot learning in LLMs is the creation of the PIA [6]. Using the in-context learning and prompting methods, an AI assistant should be able to assess language style, domain knowledge, and problem-solving skills based on the previous user's interaction history and feedback.

For instance, an intelligent writing assistant based on few-shot learning can take a small set of samples from a particular user and understand the best way to imitate their writing style or emotions. It can then be used to offer real-time suggestions and auto-completion practised by the user and assist them in articulating their ideas. Likewise, a virtual tutor could tailor the content of incoming messages/ chats by referencing the student's previous lessons, examples, and instruction. In other words, it could respond according to the student's ability levels and learning style.

Classification, the main problem of few-shot learning, brings several issues that must be solved to develop personalized AI assistants. They are in particular as follows: general label acquisition: acquisition of valuable examples from user tables; efficient example selection: determination of practical but little-known examples from all possible interactions with users; user data privacy and security: protection of users' privacy and data security; consistency and continuity of the assistant: stability of the assistant's actions over time and sessions. Integrating user feedback/corrections into the few-shot learning process also poses challenges for future work [35].

Domain-Specific Chatbots and Customer Support

Domain-adaptive few-shot learning in LLMs can help construct niche yet efficient Chatbots and customer support services [7]. Thus, using knowledge and language skills obtained during pre-training makes it possible to retrain the LLM on a given problem with the help of several specific examples and prompts. For example, the use case of a chatbot in customer support to answer queries related to a particular product/ service can be initially trained with a few typical questions that customers may ask regarding that specific product/ service, along with their respective answers. The users would be able to provide actual feedback on the interactions they had with the chatbot, and it could learn and adapt to improve its knowledge and language further, enabling it to take requests to the next level and provide accurate and helpful responses.

Few-shot learning could also enhance chatbot practices and usages by allowing the creation and adaptation of new and alternative models for various industries, uses, and languages. Instead of developing numerous chatbots, which can be challenging for the system, an initial LLM could be trained for several domains, applying only specific illustrations and prompts associated with the particular domains. This could decrease the time and effort needed to build and improve superior chatting interfaces in various modes.

Nonetheless, two fundamental problems have emerged in current research on few-shot learned chatbots: how to make few-shot learned chatbots reliable, safe, and controllable? While transforming LLMs to chatbots, there could be a risk of producing indifferent, discriminatory, or sometimes-obscene responses in case of out-of-domain or adversarial queries [36]. This paper discusses various techniques for restraining the behaviour of few-shot learned chatbots in future works, like dialogue safety classifiers and methods of controlled generation.

Knowledge Retrieval and reasoning

Improving LLMs' abilities in knowledge search and logical reasoning is another avenue for progression in the few-shot learning framework [37]. Moreover, LLMs can contain a large amount of knowledge obtained during pre-training. Still, they can only succeed when faced with the desire to reason along multi-step questions or integrate new knowledge when relatively infrequently represented in the dataset seen during training [38].

Combining few-shot learning with knowledge-aware proposals like dense passage retrieval [39] and knowledge-aged-based generation [40] may help LLMs retrieve and apply relevant information for tasks better. For example, an LLM-based question-answering system could acquire relevant passages from an external knowledge base and train its answers based on the facts and the few-shot examples presented to it.

Machine learning can also be used in conjunction with few-shot learning approaches to bring out general technology, such as chain of thought prompt [41] and scratchpad reasoning [42], to enhance the generality of LLM-based reasoning. With the help of step-by-step reasoning traces demonstrated in the prompt, an LLM could apply more acute and consistent thought processes to complex issues and subsequently convey its reasoning as logic paths along with the conclusion, which benefits users in understanding and checking the model's answer.

There is a need for more extensive research to enhance complex query processing and natural language reasoning in LLMs employing few-shot learning. This may help build more trustworthy and interpretable systems to help humans with knowledge-processing tasks such as research, decision-making, and idea finding.

Multilingual & Multimodal Learning

Another avenue for improving few-shot learning in LLMs is expanding the presented method for multilingual and multimodal cases. Although the majority of existing LLMs are developed from English-language text, there is a current trend for the seed models to include more than one language and even different media, such as speech, images, or videos [43].

Few-shot learning can also be quite effective for fine-tuning LLMs for new languages or other forms when only limited data are available. For instance, a multilingual LLM could be further trained with limited parallel sentences in a new language for zero/few-shot translation learning [44]. Likewise, as a form of multimodal

LLM, it could be trained on this new modality, for example, speech or images, by conditioning on several examples and paired text-modality data [45].

However, just as few-shot learning is valuable in many scenarios; it also presents some particular challenges in the multilingual and multimodal context [46]. These align with the task of handling linguistic features that may include syntax, semantic and cultural variations in semantic mapping across different languages with different cultures and learning to align and fuse information from multiple modalities. Cross lingual and crosschecking are some areas that should be developed in the future, such as multilingual prompting and vision-language pre-training.

The advancement of more effective and scalable approaches to few-shot learning for multilingual and multimodal recommender systems could lead to further utilization in translation, cross-modal search, and multimodal generation [47]. This means it can eliminate language and modality barriers and encourage the development of AI systems that could become more multifaceted and accessible.

Conclusion

Learning from small to moderately sized datasets through training large language models has become one of the promising ways to build more flexible, less resource-hungry and user-friendly AI systems. Regarding Model 1, LLMs can quickly learn new operations from just a handful of samples following pre-training without undergoing elaborate fine-tuning exercises. This paper has provided an overview of what few-shot learning is, how it has been applied and the main techniques within the context of LLMs, such as in-context learning, prompting, and parameter-efficient fine-tuning. As it said, these methods can help LLMs improve test performance across many language understanding and generation tasks, with sometimes better results than fine-tuned models and significantly shorter training time.

The seemingly endless possibilities that few-shot learning brings to LLMs include applications such as learning companions and personal assistants that independently adapt to each user's idiosyncrasies or domain-driven chatbots that could easily be fine-tuned for particular industries and niches. Few-shot learning could also help enhance knowledge acquisition and knowledge processing in LLMs and extend LLMs to few-shot multilingual or multimodal setups. Thus, several key challenges and limitations remain in realizing the full potential of few-shot learning in LLMs. These include enhancing sample efficiency, generalization, and robustness in few-shot learning, verifying the stability and security of the adapted models for practical applications, and incorporating superior KR&R abilities.

Future research directions for few-shot learning in LLMs include:

Disambiguation of how to refine prior methods for selecting examples and defining prompts will help create more accurate and repeatable few-shot learning.

- Investigating how achieving the best representations and training goals is possible, considering that limited data can still reveal who and how a person learns a specific task.
- We propose fine-tuning few-shot learning methodologies with knowledge-related approaches and incorporating reasoning methods to develop better LLM-based systems that are more explainable and easier to control.
- Applying these principles to generalize few-shot learning into the multilingual and multimodal context to provide an increased understanding of human-centred AI solutions.
- Exploring methods for ensuring that few-shot learned models are safe, reliable, and relevant by developing and testing approaches to generating dialogue safety classifiers and techniques that integrate humans in constructing such learned models.

To sum up, the topic of few-shot learning for large language models is emerging as a ground-breaking field with great possibilities for evolving AI application development. As a potentially effective technique for achieving AI systems smarter than humans in fewer shots, few-shot learning might help to advance the long-held dream of developing artificial general intelligence (AGI). Nevertheless, achieving these advancements will require exceptionally focused research applications and cooperation involving machine learning, NLP, human-computer interaction, and AI ethics.

References

[1]. A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," Jun. 2018.
[2]. T. B. Brown et al., "Language models are few-shot learners," arXiv preprint arXiv:2005.14165, May 2020.
[3]. A. Chowdhery et al., "PaLM: Scaling language modeling with Pathways," arXiv preprint arXiv:2204.02311, Apr. 2022.
[4]. S. Smith et al., "Using DeepSpeed and Megatron to train Megatron-Turing NLG 530B, a large-scale generative language model," arXiv preprint arXiv:2201.11990, Jan. 2022.
[5]. P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," arXiv preprint arXiv:2107.13586, Jul. 2021.
[6]. H. Adel, Z. Ghahramani, and A. Weller, "Discovering interpretable representations for both deep generative and discriminative models," in International Conference on Machine Learning, Jul. 2018, pp. 50–59.
[7]. L. Deng and Y. Liu, Eds., Deep learning in natural language processing. Springer, Feb. 2018.
[8]. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," OpenAI blog, vol. 1, no. 8, p. 9, Feb. 2019.
[9]. M. Chen et al., "Evaluating large language models trained on code," arXiv preprint arXiv:2107.03374, Jul. 2021.
[10]. A. Lazaridou, E. Gribovskaya, W. Stokowiec, and N. Grigorev, "Internet-augmented dialogue generation," arXiv preprint arXiv:2107.07566, Jul. 2021.
[11]. E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big," in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Mar. 2021, pp. 610–623.
[12]. J. Wei et al., "Chain of thought prompting elicits reasoning in large language models," arXiv preprint arXiv:2201.11903, Jan. 2022.
[13]. S. Xie, A. Rastogi, S. Upadhyay, B. Van Durme, and G. Neubig, "In-context learning for few-shot dialogue state tracking," arXiv preprint arXiv:2203.08568, Mar. 2022.
[14]. T. Zhao, E. Wallace, S. Feng, D. Klein, and S. Singh, "Calibrate before use: Improving few-shot performance of language models," arXiv preprint arXiv:2102.09690, Feb. 2021.
[15]. Z. Zhao, E. Wallace, S. Feng, D. Klein, and S. Singh, "Strategically Ordered Prompts Improve Few-Shot Performance of Language Models," arXiv preprint arXiv:2112.05236, Dec. 2021.
[16]. L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," arXiv preprint arXiv:2102.07350, Feb. 2021.
[17]. S. Mishra, D. Khashabi, C. Baral, and H. Hajishirzi, "Cross-task generalization via natural language crowdsourcing instructions," arXiv preprint arXiv:2104.08773, Apr. 2021.
[18]. X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," arXiv preprint arXiv:2101.00190, Jan. 2021.
[19]. T. Shin, Y. Razeghi, R. L. Logan IV, E. Wallace, and S. Singh, "AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts," arXiv preprint arXiv:2010.15980, Oct. 2020.
[20]. X. Liu et al., "GPT understands, too," arXiv preprint arXiv:2103.10385, Mar. 2021.
[21]. B. Lester, R. Al-Rfou, and N. Constant, "The power of scale for parameter-efficient prompt tuning," arXiv preprint arXiv:2104.08691, Apr. 2021.
[22]. J. He, G. Neubig, and T. Berg-Kirkpatrick, "Efficient nearest neighbor language models," arXiv preprint arXiv:2109.04212, Sep. 2021.
[23]. N. Houlsby et al., "Parameter-efficient transfer learning for NLP," arXiv preprint arXiv:1902.00751, Feb. 2019.
[24]. D. Guo, A. M. Rush, and Y. Kim, "Parameter-efficient transfer learning with diff pruning," arXiv preprint arXiv:2012.07463, Dec. 2020.
[25]. B. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," arXiv preprint arXiv:2101.00190, Jan. 2021.

[26]. D. Guo et al., "Parameter-efficient fine-tuning for cross-lingual summarization," arXiv preprint arXiv:2203.06869, Mar. 2022.
[27]. A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "GLUE: A multi-task benchmark and analysis platform for natural language understanding," arXiv preprint arXiv:1804.07461, Apr. 2019.
[28]. A. Srivastava et al., "Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models," arXiv preprint arXiv:2206.04615, Jun. 2022.
[29]. Y. Bai et al., "Constitutional AI: Harmlessness from AI Feedback," arXiv preprint arXiv:2212.08073, Dec. 2022.
[30]. G. Branwen, "GPT-3 creative fiction," Retrieved from https://www.gwern.net/GPT-3, Aug. 2020.
[31]. P. Xu et al., "Controllable story generation with external knowledge using large language models," arXiv preprint arXiv:2010.01029, Oct. 2020.
[32]. M. Chen et al., "Evaluating large language models trained on code," arXiv preprint arXiv:2107.03374, Jul. 2021.
[33]. X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting BERT for end-to-end aspect-based sentiment analysis," arXiv preprint arXiv:1910.00883, Oct. 2019.
[34]. Y. Zhang et al., "DIALOGPT: Large-scale generative pre-training for conversational response generation," arXiv preprint arXiv:1911.00536, Nov. 2019.
[35]. B. Hancock, A. Bordes, P. E. Mazare, and J. Weston, "Learning from dialogue after deployment: Feed yourself, chatbot!," arXiv preprint arXiv:1901.05415, Jan. 2019.
[36]. J. Xu et al., "Recipes for safety in open-domain chatbots," arXiv preprint arXiv:2010.07079, Oct. 2020.
[37]. E. Dinan et al., "The second conversational intelligence challenge (convai2)," arXiv preprint arXiv:1902.00098, Feb. 2019.
[38]. T. Yu et al., "GrailQA: Generalizable Question Answering over Knowledge Bases and Text," arXiv preprint arXiv:2011.07743, Nov. 2020.
[39]. V. Karpukhin et al., "Dense Passage Retrieval for Open-Domain Question Answering," arXiv preprint arXiv:2004.04906, Jul. 2020.
[40]. P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," arXiv preprint arXiv:2005.11401, May 2020.
[41]. J. Wei et al., "Chain of thought prompting elicits reasoning in large language models," arXiv preprint arXiv:2201.11903, Jan. 2022.
[42]. M. Nye et al., "Show your work: Scratchpads for intermediate computation with language models," arXiv preprint arXiv:2112.00114, Dec. 2021.
[43]. M. Tsimpoukelli, J. Menick, S. Cabi, S. M. Eslami, O. Vinyals, and F. Hill, "Multimodal few-shot learning with frozen language models," Advances in Neural Information Processing Systems, vol. 34, pp. 200–212, Dec. 2021.
[44]. G. I. Winata, S. Cahyawijaya, Z. Lin, Z. Liu, P. Xu, and P. Fung, "Multilingual generative language models for zero-shot cross-lingual transfer," arXiv preprint arXiv:2101.09465, Jan. 2021.
[45]. J. B. Alayrac et al., "Flamingo: a visual language model for few-shot learning," arXiv preprint arXiv:2204.14198, Apr. 2022.
[46]. X. V. Lin et al., "Few-shot learning with multilingual language models," arXiv preprint arXiv:2112.10668, Dec. 2021.
[47]. Z. Wang, J. Yu, A. W. Yu, Z. Dai, Y. Tsvetkov, and Y. Cao, "Simvlm: Simple visual language model pretraining with weak supervision," arXiv preprint arXiv:2108.10904, Aug. 2021.