Review

In-Context Learning in Large Language Models: A Comprehensive Survey

[Clyde Highmore](#) *

Article

In-Context Learning in Large Language Models: A Comprehensive Survey

Clyde Highmore

Scholar-Chat Research; research@scholar-chat.com

Abstract: This survey provides a comprehensive overview of in-context learning (ICL) in large language models (LLMs), a phenomenon where models can adapt to new tasks without parameter updates by leveraging task-relevant information within the input context. We explore the definition and mechanisms of ICL, investigate the factors contributing to its emergence, and discuss strategies for optimizing and effectively utilizing ICL in various applications. Through a systematic review of recent literature, we first clarify what ICL is, distinguishing it from traditional fine-tuning approaches and highlighting its unique characteristics. We then delve into the underlying causes of ICL, examining theories ranging from implicit meta-learning during pre-training to the emergence of task vectors in LLMs. The survey also covers various approaches to enhance ICL performance, including prompt engineering techniques, demonstration selection strategies, and methods for improving generalization across diverse tasks. Additionally, we discuss the limitations and challenges of ICL, such as its sensitivity to demonstration ordering and potential biases. By synthesizing findings from numerous studies, we aim to provide researchers and practitioners with a clear understanding of the current state of ICL research, its practical implications, and promising directions for future investigation. This survey serves as a valuable resource for those seeking to leverage ICL capabilities in LLMs and contributes to the ongoing discourse on the remarkable adaptability of these models.

Keywords: large language model; in-context learning

1. Introduction

The advent of large language models (LLMs) has introduced a paradigm shift in machine learning, where models exhibit remarkable capabilities in understanding and generating human language. A critical avenue of exploration within this domain is In-Context Learning (ICL), a process allowing LLMs to learn from and adapt to new information presented in their input context. This literature review aims to provide a comprehensive overview of ICL in LLMs, elucidating what ICL is, the factors contributing to its emergence and optimization, and strategies for effective utilization. We review the characteristics of ICL, examine the mechanisms underpinning it, and discuss approaches to enhance its performance and generalization. This survey not only captures the advancements in ICL research but also underscores the existing challenges that prompt ongoing inquiry, thereby serving as an informative resource for researchers and practitioners alike[1].

2. Defining In-Context Learning in Large Language Models

2.1. Characteristics of ICL

Large language models (LLMs) manifest emergent abilities and improved in-context learning (ICL) as they scale, but with limitations and increased social biases [1,2].

In the paper by Wei et al. [1], LLMs are observed to develop emergent abilities that are not present in smaller models. These abilities, which arise unpredictably as the model scales, indicate that further scaling could expand the capabilities of LLMs, potentially enhancing ICL characteristics.

---

[1]   This survey can be reproduced with Scholar-Chat AI

Srivastava et al. [2] introduce the BIG-bench to assess the capabilities of LLMs and identify that quantitative improvements in model performance also translate to qualitative enhancements in ICL. However, the study also notes that despite the scaling advantages, LLMs exhibit poor absolute performance in calibration when compared to human raters and that social biases tend to increase with model size. This suggests that while ICL characteristics improve with scale, they are accompanied by challenges that need addressing.

2.2. Research Gaps and Suggested Agenda

Large language models demonstrate emergent abilities and improved task performance with scale, but the mechanisms of in-context learning remain underexplored [1,2]. To advance our understanding, we suggest the following research agenda:

- Investigate the threshold at which in-context learning emerges in language models of varying sizes and architectures.
- Explore the differences in in-context learning capabilities between models trained on diverse datasets.
- Examine the impact of model scaling on the robustness and generalizability of in-context learning.

3. Theoretical Foundations of ICL

The chapter explores the theoretical foundations of In-context Learning (ICL) in large language models, revealing its connections to implicit meta-learning, Bayesian inference, and feature learning relevant to downstream tasks. Research by [3] and [4] provides mathematical and empirical evidence for these connections, while [5] further shows transformers' ability to select and implement various machine learning algorithms in-context. The chapter concludes with the insight from [6] that "induction heads" within transformers may be instrumental in facilitating ICL, thereby enhancing our understanding of the adaptability of these models.

3.1. Implicit Meta-Learning During Pre-training

The theoretical underpinnings of In-context Learning (ICL) suggest that large language models develop implicit meta-learning capabilities during pre-training, which enable them to adapt to new tasks without explicit re-training. In the work by [3], it is argued that in-context learning can be seen as a form of implicit Bayesian inference, emerging when pretraining involves documents with long-range coherence, and the model must infer latent concepts to generate coherent text. This understanding is supported by experiments on a synthetic dataset where both Transformers and LSTMs exhibit in-context learning behaviors. Meanwhile, [4] explores the success of autoregressive language models in downstream tasks, providing mathematical reasoning that connects learning the distribution of natural language with the efficacy of solving downstream tasks. They show that language models that minimize cross-entropy also learn features beneficial for natural linear classification tasks. Lastly, [5] demonstrates that transformers can implement a variety of machine learning algorithms in-context and, more crucially, perform in-context algorithm selection, which mimics the adaptability of a human statistician in choosing the right method for a given problem. Collectively, these works provide a mathematical and empirical foundation for understanding the phenomenon of ICL in large language models.

3.2. Emergence of Task Vectors

TLDR: Evidence suggests induction heads may drive in-context learning in transformers [6].

Olsson et al. [6] present a compelling argument that "induction heads" within transformer models are critical for the phenomenon of in-context learning, where the model improves its predictions as it processes more tokens in a sequence. The authors identify a correlation between the emergence of these induction heads and a marked increase in the model's in-context learning capabilities. This discovery is significant as it points to a possible mechanistic source of in-context learning, which could

be a foundational element for task-specific adaptations in large language models. The study offers both direct and indirect evidence across different model sizes, with stronger causal evidence for smaller models and correlational evidence for larger ones. This work contributes to the theoretical foundation of in-context learning (ICL) by shedding light on the internal structures, specifically induction heads, that facilitate this learning process, thereby advancing our understanding of the emergence of task vectors in transformer networks.

3.3. Research Gaps and Suggested Agenda

The theoretical underpinnings of in-context learning (ICL) in large language models are beginning to be explored, yet substantial gaps remain [3–6].

- Further investigation into the precise mechanisms by which transformers represent and utilize in-context information is needed, building on the preliminary identification of induction heads [6].
- There is a need for a deeper understanding of the distributional properties that facilitate ICL, especially in real-world data, beyond synthetic datasets [3].
- Expanding on the theoretical frameworks that link language modeling to downstream task performance, including the exploration of alternative objectives to cross-entropy, could yield insights into more efficient pretraining strategies [4].

4. Mechanisms Underlying ICL

This chapter delves into the research question of "Mechanisms Underlying ICL," revealing that In-Context Learning in large language models is a complex process influenced by model scale and the semantic role of label words, rather than relying solely on accurate ground truth demonstrations. It encapsulates the evolving understanding that ICL may involve gradient descent-like optimization, with empirical and theoretical insights suggesting that models can meta-learn from context and exhibit rapid performance leaps termed 'Eureka-moments.' The chapter also underscores the importance of the representation of label space, input distribution, and the semantic anchoring role of label words in driving ICL, pointing toward a future where these insights inform more effective model training and error diagnosis strategies.

4.1. Gradient Descent-like Behavior

In the exploration of the mechanisms underlying In-Context Learning (ICL), recent studies have proposed that language models may exhibit gradient descent-like behavior during ICL. Dai *et al.* [7] argue that the Transformer's attention mechanism implicitly performs meta-optimization akin to gradient descent, where demonstrations generate meta-gradients that fine-tune the model in-context. This hypothesis is supported by empirical evidence showing that ICL resembles explicit fine-tuning at various levels of model behavior. Additionally, they propose a novel momentum-based attention mechanism that outperforms the standard attention mechanism, further supporting the gradient descent analogy.

In contrast, Natan *et al.* [8] highlight gaps in the evaluation of the ICL-gradient descent correspondence, particularly in the use of problematic metrics and insufficient baselines. Their findings suggest that even untrained models can achieve similar ICL-GD similarity scores without demonstrating actual ICL. They introduce a new gradient descent-based optimization that respects layer causality, which aligns more closely with the information flow in ICL and improves similarity scores.

Lastly, Hoffmann *et al.* [9] observe 'Eureka-moments' in transformers during multi-step tasks, where performance rapidly improves after a prolonged plateau. They attribute this phenomenon to the softmax function within the self-attention mechanism and demonstrate that addressing this issue can enhance the training of both language models and ICL.

Overall, these findings collectively contribute to a nuanced understanding of the gradient descent-like processes that may underpin ICL in large language models.

4.2. Statistical Learning in Transformers

The key findings from recent literature suggest that in-context learning (ICL) in large language models (LLMs) does not rely heavily on ground truth demonstrations, and the mechanism of learning can be influenced by the model's scale and the nature of label words. In [10], it is demonstrated that performance on classification tasks is not significantly impacted even when labels in demonstrations are randomly replaced, suggesting that label correctness is not critical for ICL. Instead, aspects such as label space representation, input text distribution, and sequence format are vital. The study in [11] elaborates that larger LLMs can override their semantic priors based on the input-label mappings provided in-context, emphasizing that the capacity to learn from contextually presented exemplars scales with model size. Furthermore, [12] presents label words as informational anchors in ICL, where semantic information consolidates into these anchors in early computation layers, guiding the model's final predictions. This understanding has led to methods for enhancing ICL performance and diagnosing errors. Finally, [13] offers a data generation perspective on the mechanism of ICL, distinguishing between skill learning and skill recognition, and suggesting a unifying technical foundation for interpreting and advancing ICL research.

4.3. Research Gaps and Suggested Agenda

In-context learning (ICL) mechanisms in large language models are partially understood, with diverse perspectives but lacking a unified theory [7–13]. Research gaps include:

- The precise nature of the gradient descent-like processes that occur during ICL and their correlation with explicit fine-tuning needs further investigation [7,8].
- Understanding the role of softmax in attention mechanisms and its impact on abrupt learning improvements can be valuable for enhancing ICL [9].
- A more systematic approach is needed to integrate the various perspectives on ICL, such as the data generation view and the role of demonstration examples [10,13].

5. Optimizing ICL Performance

In addressing the research question "Optimizing ICL Performance," this chapter reviews recent advancements in In-Context Learning (ICL) that focus on improving the efficiency and effectiveness of Large Language Models (LLMs) through novel prompt engineering methods and example selection strategies. These advancements span various techniques, including contrastive learning for response customization, adversarial optimization for prompt refinement, and state vector adjustments for incremental improvements, as well as innovative algorithms for demonstration selection that prioritize relevance and quality, offering a narrative of progress towards more sophisticated ICL paradigms.

5.1. Prompt Engineering Techniques

TLDR: Recent studies have introduced novel prompt engineering techniques to enhance the performance of In-Context Learning (ICL) in Large Language Models (LLMs), utilizing contrastive learning, adversarial optimization, and state vector refinement.

In the pursuit of optimizing ICL performance, Gao and Das [14] proposed a method that leverages contrastive examples to customize LLM responses, significantly improving content generation on platforms like StackExchange and Reddit. The technique uses positive examples to highlight desired output characteristics, while negative examples teach the model what to avoid, resulting in fine-tuned responses aligned with user intent.

Meanwhile, Do *et al.* [15] introduced Adversarial In-Context Learning (adv-ICL), an innovative prompt optimization framework. By adopting an adversarial setup, adv-ICL uses a generator and discriminator LLM to iteratively refine prompts. This approach has demonstrated substantial enhancements across a variety of tasks, outperforming existing methods in efficiency and adaptability, particularly in low-resource scenarios.

Lastly, Li *et al.* [16] explored the optimization of ICL through state vectors, analogous to trained parameters. The paper presented inner and momentum optimization techniques to incrementally refine these vectors, achieving state-of-the-art results on tasks such as Llama-2 and GPT-J, even when examples are too lengthy for conventional ICL approaches.

Collectively, these studies contribute to the evolving field of prompt engineering, offering robust, adaptable, and efficient techniques for optimizing the performance of LLMs.

5.2. Demonstration Selection Strategies

Demonstration selection strategies are crucial for optimizing In-Context Learning (ICL) performance, with recent studies proposing various approaches to enhance the relevance and quality of selected examples. Li et al. [17] introduced the Unified Demonstration Retriever (UDR), which is a multi-task list-wise ranking framework capable of retrieving demonstrations across a wide array of tasks, showing significant improvements over baselines in diverse scenarios. Ye et al. [18] focused on subset selection using Determinantal Point Processes (DPPs) through their Compositional Exemplars for In-context Learning (CEIL) approach, demonstrating state-of-the-art performance with benefits in transferability and compositionality. Mavromatis et al. [19]'s AdaICL algorithm leverages uncertainty and diversity sampling within an active learning framework for ICL to achieve both effectiveness and budget efficiency. Luo et al. [20] enhanced retrieval-based ICL methods, showing that task-specific retrievers and simple similarity measures like BM25 can outperform random selection. Lastly, Li et al. [21]'s LENS method employs a two-stage process to filter and search for supportive examples, introducing a novel informativeness metric and demonstrating notable improvements against various baselines.

Together, these studies present a narrative of advancing ICL by prioritizing example selection based on relevance, diversity, and language model feedback—each contributing to more effective and efficient learning paradigms.

5.3. Example Ordering Methods

TLDR: Recent studies [22], [23], and [24] provide novel methodologies for optimizing in-context learning performance by addressing the challenges posed by example ordering.

In [22], the authors propose Batch-ICL, which treats in-context learning as a meta-optimization process, demonstrating its effectiveness and efficiency. Batch-ICL is agnostic to the order of examples due to its strategy of using separate one-shot computations and aggregating meta-gradients, which consistently outperforms standard ICL permutations and reduces computational demand.

The work in [23] introduces EASE, an approach that uses a neural bandit algorithm to optimize example ordering, significantly impacting LLM performance. EASE operates without extra test-time computation, providing an efficient solution to select and order exemplars, and can be extended to optimize the instruction component of prompts as well.

Lastly, [24] formulates example ordering as an optimization problem, proposing principles guided by model probability predictions. These principles are shown to enhance classification accuracy, reduce model miscalibration, and select better in-context examples across various datasets and LLMs.

5.4. Research Gaps and Suggested Agenda

Despite advancements in in-context learning (ICL) optimization, challenges remain in example selection, ordering, and computational efficiency [14–25].

- Investigate the interplay between example selection and ordering to understand their combined influence on model performance and calibration [18,24].
- Develop methods to reduce the computational cost of ICL, especially in real-time applications, without sacrificing the quality of the generated content [22,25].
- Explore the potential of unsupervised or self-supervised methods for demonstration retrieval and optimization to enhance ICL across diverse tasks and domains [17,23].

6. Enhancing ICL Generalization

The chapter investigates the research question of "Enhancing ICL Generalization," revealing strategies to improve In-context Learning (ICL) without extensive fine-tuning. Recent studies have introduced various methods, such as in-context knowledge editing, bias calibration, and direct preference optimization, which show promise in enabling large language models to update knowledge, mitigate biases, and align responses more effectively across diverse tasks. Additionally, innovations such as Self-ICL, Z-ICL, and *k*NN Prompting enhance zero-shot and few-shot learning, while approaches for document information extraction further the ability of models to generalize across different contexts.

6.1. Cross-task Transfer in ICL

The collective findings from recent literature suggest that In-context Learning (ICL) can be enhanced for better cross-task generalization without the need for costly fine-tuning processes. Zheng et al. [26] demonstrated that in-context knowledge editing (IKE) could effectively update factual knowledge in large language models (LLMs) like GPT-J and OPT-175B, with fewer side effects such as over-editing or knowledge forgetting. Meanwhile, Fei et al. [27] addressed label biases in ICL and proposed a bias calibration method that significantly improves performance across a variety of tasks by controlling for estimated biases. Lastly, Song et al. [28] introduced In-Context Direct Preference Optimization (ICDPO), a method that enhances LLMs' alignment capabilities by borrowing from superior models without fine-tuning, showing promise in generating well-aligned responses across different tasks. Together, these studies contribute to the understanding of how ICL can be refined to facilitate better transfer of knowledge and preferences across diverse tasks.

6.2. Few-shot and Zero-shot ICL Approaches

Research in enhancing the generalization capabilities of In-Context Learning (ICL) has led to the development of methods that improve few-shot and zero-shot performance. [29] introduces Self-ICL, a method that generates its own demonstrations for zero-shot ICL, demonstrating superior performance on various tasks without real demonstrations. [30] presents Z-ICL, which constructs pseudo-demonstrations from a raw text corpus, achieving significant improvements in zero-shot performance and matching few-shot learning with actual labeled data. [31] proposes *k*NN Prompting, which scales beyond-context learning and eliminates the need for calibration, showing effectiveness across a vast range of training data sizes and model scales. Finally, [32] explores ICL for document information extraction, employing a diverse demonstrations updating strategy that allows LLMs to better generalize to different documents, achieving top performance in both in-distribution and out-of-distribution settings.

6.3. Research Gaps and Suggested Agenda

While significant progress has been made in the field of In-Context Learning (ICL), there are still important research gaps that need to be addressed to enhance ICL's generalization capabilities. Here we identify pivotal areas for future research.

- **Mitigating Label Biases:** Despite attempts to control label biases in ICL, as highlighted by [27], further research is necessary to develop more robust bias mitigation strategies that can handle the nuanced and diverse label biases that can occur across different domains and model scales.
- **Knowledge Editing and Updating:** The ability to edit and update factual knowledge within LLMs without retraining is a crucial area of exploration. Studies like [26] show the potential of in-context knowledge editing, but more scalable and efficient methods need to be developed to handle the dynamic nature of knowledge.
- **Diverse Demonstration Effectiveness:** While the use of diverse demonstrations for ICL has been proposed ([32]), understanding the optimal strategies for selecting and updating these demonstrations to maximize generalization across various tasks remains an open question.

7. Applications and Use Cases of ICL

The chapter on "Applications and Use Cases of ICL" illustrates the effectiveness and adaptability of in-context learning (ICL) across a spectrum of areas, including medicine, cloud services, visual tasks, and multimodal challenges. Studies within the medical domain indicate that ICL, as exemplified by ChatGPT's performance on USMLE exams, has the potential to revolutionize medical education and clinical decision-making. In the realm of cloud services, ICL has been shown to enhance incident diagnosis without the need for extensive training. Furthermore, advancements in visual and multimodal applications demonstrate ICL's capability to interpret complex prompts and improve task performance through innovative methodologies such as MMICL and Prompt Diffusion. Overall, the chapter underscores ICL's transformative impact on diverse domains by enabling models to learn and generalize from context with minimal specialized training.

7.1. Natural Language Processing Tasks

The application of in-context learning (ICL) in natural language processing (NLP) tasks showcases its potential to transform educational and professional realms, particularly in the medical field and incident diagnosis within cloud services. Kung et al. [33] demonstrated that ChatGPT can perform at or near the passing threshold on USMLE exams, indicating its capability as an assistive tool in medical education and potentially in clinical decision-making. Gilson et al. [34] also evaluated ChatGPT's performance on USMLE Step 1 and Step 2, finding that it outperformed other large language models and provided logical justifications for its answers, reinforcing its suitability as an educational aid. Li et al. [35] introduced KB-BINDER, a framework that utilizes ICL for answering queries over diverse knowledge bases without specialized training, proving its effectiveness by outperforming or matching state-of-the-art trained models. Seegmiller et al. [36] explored ICL's scope in extracting medical temporal constraints from unstructured texts, achieving notable success and suggesting its use for enhancing patient-centric healthcare applications. Lastly, Liu et al. [37] and Zhang et al. [38] respectively advanced our theoretical understanding of ICL's capabilities in NLP tasks and demonstrated its practical application in automated root causing for cloud incidents, with Zhang et al. showing significant improvements over previous models without the need for fine-tuning. Collectively, these studies underscore ICL's versatility and efficacy in various NLP tasks, heralding a new era of AI-assisted applications across different domains.

7.2. Visual and Multimodal Applications

In the realm of visual and multimodal applications, in-context learning (ICL) has been explored to enhance the capabilities of models in understanding complex prompts and tasks. Zhao et al. [39] introduced MMICL, a vision-language model that incorporates multi-modal in-context learning, and demonstrated its superior zero-shot performance on general vision-language tasks. Zhang et al. [40] focused on the selection of in-context examples for visual ICL, proposing both unsupervised and supervised prompt retrieval methods to improve performance. Wang et al. [41] presented Prompt Diffusion, a diffusion-based generative model framework that enables ICL by using vision-language prompts, showing high-quality in-context generation and generalization to new tasks. Lastly, Zhang et al. [42] investigated the use of ICL for compositional generalization in LLMs and introduced a human-guided tool manipulation framework that significantly enhances performance on complex compositional questions. Collectively, these papers underscore the potential of ICL in a variety of visual and multimodal applications, from improving task understanding to automating example selection and facilitating compositional generalization.

7.3. Research Gaps and Suggested Agenda

In-context learning (ICL) has shown promise in various domains, yet challenges remain in its broader application and understanding [33–42]. Research gaps and a future agenda could include:

- Exploring the mechanisms behind ICL's performance variability across different tasks and datasets, as evidenced by the variance in medical question-answering [34] and root causing of cloud incidents [38].
- Enhancing the in-context learning capabilities of VLMs for complex multi-modal prompts, building on the work in [39], and addressing language biases in these models.
- Developing more effective methods for selecting in-context examples to improve ICL performance, as the quality of examples is critical for tasks like visual in-context learning [40].

8. Conclusion

In-Context Learning (ICL) is a defining feature of large language models, enabling them to adapt and respond to new tasks with minimal intervention. Our literature survey has outlined the characteristics, mechanisms, optimization techniques, and applications of ICL, while also pinpointing the research gaps and proposing a future agenda. It is evident that while ICL has made impressive strides, particularly in domains such as healthcare and cloud services, significant challenges remain. These include understanding the emergence and optimization of ICL, mitigating biases, improving generalization, and enhancing computational efficiency. Addressing these challenges will require concerted research efforts to unlock the full potential of ICL in LLMs, which promises to revolutionize how we interact with and leverage artificial intelligence for complex problem-solving.

References

1. Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; Chi, E.; Hashimoto, T.; Vinyals, O.; Liang, P.; Dean, J.; Fedus, W. Emergent Abilities of Large Language Models 2022.
2. Srivastava, A.; Rastogi, A.; Rao, A.; Shoeb, A.A.M.; Abid, A.; Fisch, A.; Brown, A.R.; Santoro, A.; Gupta, A.; Garriga-Alonso, A.; Kluska, A.; Lewkowycz, A.; Agarwal, A.; Power, A.; Ray, A.; Warstadt, A.; Kocurek, A.W.; Safaya, A.; Tazarv, A.; Xiang, A.; Parrish, A.; Nie, A.; Hussain, A.; Askell, A.; Dsouza, A.; Slone, A.; Rahane, A.; Iyer, A.S.; Andreassen, A.; Madotto, A.; Santilli, A.; Stuhlmuller, A.; Dai, A.M.; La, A.; Lampinen, A.K.; Zou, A.; Jiang, A.; Chen, A.; Vuong, A.; Gupta, A.; Gottardi, A.; Norelli, A.; Venkatesh, A.; Gholamidavoodi, A.; Tabassum, A.; Menezes, A.; Kirubarajan, A.; Mullokandov, A.; Sabharwal, A.; Herrick, A.; Efrat, A.; Erdem, A.; Karakacs, A.; Roberts, B.R.; Loe, B.S.; Zoph, B.; Bojanowski, B.; Ozyurt, B.; Hedayatnia, B.; Neyshabur, B.; Inden, B.; Stein, B.; Ekmekci, B.; Lin, B.Y.; Howald, B.; Orinion, B.; Diao, C.; Dour, C.; Stinson, C.; Argueta, C.; Ram'irez, C.F.; Singh, C.; Rathkopf, C.; Meng, C.; Baral, C.; Wu, C.; Callison-Burch, C.; Waites, C.; Voigt, C.; Manning, C.D.; Potts, C.; Ramirez, C.; Rivera, C.; Siro, C.; Raffel, C.; Ashcraft, C.; Garbacea, C.; Sileo, D.; Garrette, D.H.; Hendrycks, D.; Kilman, D.; Roth, D.; Freeman, D.; Khashabi, D.; Levy, D.; Gonz'alez, D.; Perszyk, D.R.; Hernandez, D.; Chen, D.; Ippolito, D.; Gilboa, D.; Dohan, D.; Drakard, D.; Jurgens, D.; Datta, D.; Ganguli, D.; Emelin, D.; Kleyko, D.; Yuret, D.; Chen, D.; Tam, D.; Hupkes, D.; Misra, D.; Buzan, D.; Mollo, D.C.; Yang, D.; Lee, D.H.; Schrader, D.; Shutova, E.; Cubuk, E.D.; Segal, E.; Hagerman, E.; Barnes, E.; Donoway, E.; Pavlick, E.; Rodolà, E.; Lam, E.; Chu, E.; Tang, E.; Erdem, E.; Chang, E.; Chi, E.A.; Dyer, E.; Jerzak, E.; Kim, E.; Manyasi, E.E.; Zheltonozhskii, E.; Xia, F.; Siar, F.; Mart'inez-Plumed, F.; Happ'e, F.; Chollet, F.; Rong, F.; Mishra, G.; Winata, G.I.; Melo, G.d.; Kruszewski, G.; Parascandolo, G.; Mariani, G.; Wang, G.X.; Jaimovitch-L'opez, G.; Betz, G.; Gur-Ari, G.; Galijasevic, H.; Kim, H.; Rashkin, H.; Hajishirzi, H.; Mehta, H.; Bogar, H.; Shevlin, H.; Schutze, H.; Yakura, H.; Zhang, H.; Wong, H.M.; Ng, I.; Noble, I.; Jumelet, J.; Geissinger, J.; Kernion, J.; Hilton, J.; Lee, J.; Fisac, J.; Simon, J.B.; Koppel, J.; Zheng, J.; Zou, J.; Koco'n, J.; Thompson, J.; Wingfield, J.; Kaplan, J.; Radom, J.; Sohl-Dickstein, J.N.; Phang, J.; Wei, J.; Yosinski, J.; Novikova, J.; Bosscher, J.; Marsh, J.; Kim, J.; Taal, J.; Engel, J.; Alabi, J.O.; Xu, J.; Song, J.; Tang, J.; Waweru, J.W.; Burden, J.; Miller, J.; Balis, J.U.; Batchelder, J.; Berant, J.; Frohberg, J.; Rozen, J.; Hernández-Orallo, J.; Boudeman, J.; Guerr, J.; Jones, J.; Tenenbaum, J.; Rule, J.S.; Chua, J.; Kanclerz, K.; Livescu, K.; Krauth, K.; Gopalakrishnan, K.; Ignatyeva, K.; Markert, K.; Dhole, K.D.; Gimpel, K.; Omondi, K.; Mathewson, K.; Chiafullo, K.; Shkaruta, K.; Shridhar, K.; McDonell, K.; Richardson, K.; Reynolds, L.; Gao, L.; Zhang, L.; Dugan, L.; Qin, L.; Contreras-Ochando, L.; Morency, L.P.; Moschella, L.; Lam, L.; Noble, L.; Schmidt, L.; He, L.; Col'on, L.O.; Metz, L.; cSenel, L.K.; Bosma, M.; Sap, M.; Hoeve, M.t.; Farooqi, M.; Faruqui,

M.; Mazeika, M.; Baturan, M.; Marelli, M.; Maru, M.; Quintana, M.J.R.; Tolkiehn, M.; Giulianelli, M.; Lewis, M.; Potthast, M.; Leavitt, M.L.; Hagen, M.; Schubert, M.; Baitemirova, M.; Arnaud, M.; McElrath, M.; Yee, M.; Cohen, M.; Gu, M.; Ivanitskiy, M.; Starritt, M.; Strube, M.; Swkedrowski, M.; Bevilacqua, M.; Yasunaga, M.; Kale, M.; Cain, M.; Xu, M.; Suzgun, M.; Walker, M.; Tiwari, M.; Bansal, M.; Aminnaseri, M.; Geva, M.; Gheini, M.; MukundVarma, T.; Peng, N.; Chi, N.A.; Lee, N.; Krakover, N.G.A.; Cameron, N.; Roberts, N.; Doiron, N.; Martinez, N.; Nangia, N.; Deckers, N.; Muennighoff, N.; Keskar, N.; Iyer, N.; Constant, N.; Fiedel, N.; Wen, N.; Zhang, O.; Agha, O.; Elbaghdadi, O.; Levy, O.; Evans, O.; Casares, P.A.M.; Doshi, P.; Fung, P.; Liang, P.; Vicol, P.; Alipoormolabashi, P.; Liao, P.; Liang, P.; Chang, P.; Eckersley, P.; Htut, P.M.; Hwang, P.B.; Milkowski, P.; Patil, P.; Pezeshkpour, P.; Oli, P.; Mei, Q.; Lyu, Q.; Chen, Q.; Banjade, R.; Rudolph, R.E.; Gabriel, R.; Habacker, R.; Risco, R.; Milliere, R.; Garg, R.; Barnes, R.; Saurous, R.; Arakawa, R.; Raymaekers, R.; Frank, R.; Sikand, R.; Novak, R.; Sitelew, R.; Bras, R.L.; Liu, R.; Jacobs, R.; Zhang, R.; Salakhutdinov, R.; Chi, R.; Lee, R.; Stovall, R.; Teehan, R.; Yang, R.; Singh, S.; Mohammad, S.M.; Anand, S.; Dillavou, S.; Shleifer, S.; Wiseman, S.; Gruetter, S.; Bowman, S.R.; Schoenholz, S.; Han, S.; Kwatra, S.; Rous, S.A.; Ghazarian, S.; Ghosh, S.; Casey, S.; Bischoff, S.; Gehrmann, S.; Schuster, S.; Sadeghi, S.; Hamdan, S.S.; Zhou, S.; Srivastava, S.; Shi, S.; Singh, S.; Asaadi, S.; Gu, S.; Pachchigar, S.; Toshniwal, S.; Upadhyay, S.; Debnath, S.; Shakeri, S.; Thormeyer, S.; Melzi, S.; Reddy, S.; Makini, S.; Lee, S.H.; Torene, S.; Hatwar, S.; Dehaene, S.; Divic, S.; Ermon, S.; Biderman, S.; Lin, S.; Prasad, S.; Piantadosi, S.T.; Shieber, S.M.; Misherghi, S.; Kiritchenko, S.; Mishra, S.; Linzen, T.; Schuster, T.; Li, T.; Yu, T.; Ali, T.; Hashimoto, T.; Wu, T.L.; Desbordes, T.; Rothschild, T.; Phan, T.; Wang, T.; Nkinyili, T.; Schick, T.; Kornev, T.; Tunduny, T.; Gerstenberg, T.; Chang, T.; Neeraj, T.; Khot, T.; Shultz, T.; Shaham, U.; Misra, V.; Demberg, V.; Nyamai, V.; Raunak, V.; Ramasesh, V.; Prabhu, V.U.; Padmakumar, V.; Srikumar, V.; Fedus, W.; Saunders, W.; Zhang, W.; Vossen, W.; Ren, X.; Tong, X.; Zhao, X.; Wu, X.; Shen, X.; Yaghoobzadeh, Y.; Lakretz, Y.; Song, Y.; Bahri, Y.; Choi, Y.; Yang, Y.; Hao, Y.; Chen, Y.; Belinkov, Y.; Hou, Y.; Hou, Y.; Bai, Y.; Seid, Z.; Zhao, Z.; Wang, Z.; Wang, Z.J.; Wang, Z.; Wu, Z. Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models **2022**.

3.   Xie, S.M.; Raghunathan, A.; Liang, P.; Ma, T. An Explanation of In-context Learning as Implicit Bayesian Inference **2021**.

4.   Saunshi, N.; Malladi, S.; Arora, S. A Mathematical Exploration of Why Language Models Help Solve Downstream Tasks **2020**.

5.   Bai, Y.; Chen, F.; Wang, H.; Xiong, C.; Mei, S. Transformers as Statisticians: Provable In-Context Learning with In-Context Algorithm Selection **2023**.

6.   Olsson, C.; Elhage, N.; Nanda, N.; Joseph, N.; Dassarma, N.; Henighan, T.; Mann, B.; Askell, A.; Bai, Y.; Chen, A.; Conerly, T.; Drain, D.; Ganguli, D.; Hatfield-Dodds, Z.; Hernandez, D.; Johnston, S.; Jones, A.; Kernion, J.; Lovitt, L.; Ndousse, K.; Amodei, D.; Brown, T.B.; Clark, J.; Kaplan, J.; McCandlish, S.; Olah, C. In-context Learning and Induction Heads **2022**.

7.   Dai, D.; Sun, Y.; Dong, L.; Hao, Y.; Sui, Z.; Wei, F. Why Can GPT Learn In-Context? Language Models Secretly Perform Gradient Descent as Meta-Optimizers **2023**.

8.   Natan, T.B.; Deutch, G.; Magar, N.; Dar, G. In-context Learning and Gradient Descent Revisited **2023**.

9.   Hoffmann, D.T.; Schrodi, S.; Behrmann, N.; Fischer, V.; Brox, T. Eureka-Moments in Transformers: Multi-Step Tasks Reveal Softmax Induced Optimization Problems **2023**.

10. Min, S.; Lyu, X.; Holtzman, A.; Artetxe, M.; Lewis, M.; Hajishirzi, H.; Zettlemoyer, L. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work? **2022**.

11. Wei, J.W.; Wei, J.; Tay, Y.; Tran, D.; Webson, A.; Lu, Y.; Chen, X.; Liu, H.; Huang, D.; Zhou, D.; Ma, T. Larger language models do in-context learning differently **2023**.

12. Wang, L.; Li, L.; Dai, D.; Chen, D.; Zhou, H.; Meng, F.; Zhou, J.; Sun, X. Label Words are Anchors: An Information Flow Perspective for Understanding In-Context Learning **2023**.

13. Mao, H.; Liu, G.D.; Ma, Y.; Wang, R.; Tang, J. A Data Generation Perspective to the Mechanism of In-Context Learning **2024**.

14. Gao, X.; Das, K. Customizing Language Model Responses with Contrastive In-Context Learning **2024**.

15. Do, X.L.; Zhao, Y.; Brown, H.; Xie, Y.; Zhao, J.X.; Chen, N.F.; Kawaguchi, K.; Xie, M.Q.; He, J. Prompt Optimization via Adversarial In-Context Learning **2023**.

16. Li, D.; Liu, Z.; Hu, X.; Sun, Z.; Hu, B.; Zhang, M. In-Context Learning State Vector with Inner and Momentum Optimization **2024**.

17. Li, X.; Lv, K.; Yan, H.; Lin, T.; Zhu, W.; Ni, Y.; Xie, G.; Wang, X.; Qiu, X. Unified Demonstration Retriever for In-Context Learning **2023**.

18. Ye, J.; Wu, Z.; Feng, J.; Yu, T.; Kong, L. Compositional Exemplars for In-context Learning **2023**.

19. Mavromatis, C.; Srinivasan, B.; Shen, Z.; Zhang, J.; Rangwala, H.; Faloutsos, C.; Karypis, G. Which Examples to Annotate for In-Context Learning? Towards Effective and Efficient Selection **2023**.

20. Luo, M.; Xu, X.; Dai, Z.; Pasupat, P.; Kazemi, M.; Baral, C.; Imbrasaite, V.; Zhao, V. Dr.ICL: Demonstration-Retrieved In-context Learning **2023**.

21. Li, X.; Qiu, X. Finding Support Examples for In-Context Learning **2023**.

22. Zhang, K.; Lv, A.; Chen, Y.; Ha, H.; Xu, T.; Yan, R. Batch-ICL: Effective, Efficient, and Order-Agnostic In-Context Learning **2024**.

23. Wu, Z.; Lin, X.; Dai, Z.; Hu, W.; Shu, Y.; Ng, S.K.; Jaillet, P.; Low, B. Prompt Optimization with EASE? Efficient Ordering-aware Automated Selection of Exemplars **2024**.

24. Xu, Z.; Cohen, D.; Wang, B.; Srikumar, V. In-Context Example Ordering Guided by Label Distributions **2024**.

25. Ye, Q.; Beltagy, I.; Peters, M.E.; Ren, X.; Hajishirzi, H. FiD-ICL: A Fusion-in-Decoder Approach for Efficient In-Context Learning **2023**.

26. Zheng, C.; Li, L.; Dong, Q.; Fan, Y.; Wu, Z.; Xu, J.; Chang, B. Can We Edit Factual Knowledge by In-Context Learning? **2023**.

27. Fei, Y.; Hou, Y.; Chen, Z.; Bosselut, A. Mitigating Label Biases for In-context Learning **2023**.

28. Song, F.; Fan, Y.; Zhang, X.; Wang, P.; Wang, H. ICDPO: Effectively Borrowing Alignment Capability of Others via In-context Direct Preference Optimization **2024**.

29. Chen, W.L.; Wu, C.K.; Chen, H.H. Self-ICL: Zero-Shot In-Context Learning with Self-Generated Demonstrations **2023**.

30. Lyu, X.; Min, S.; Beltagy, I.; Zettlemoyer, L.; Hajishirzi, H. Z-ICL: Zero-Shot In-Context Learning with Pseudo-Demonstrations **2022**.

31. Xu, B.; Wang, Q.; Mao, Z.; Lyu, Y.; She, Q.; Zhang, Y. kNN Prompting: Beyond-Context Learning with Calibration-Free Nearest Neighbor Inference **2023**.

32. He, J.; Wang, L.; Hu, Y.; Liu, N.; Liu, H.j.; Xu, X.; Shen, H. ICL-D3IE: In-Context Learning with Diverse Demonstrations Updating for Document Information Extraction **2023**.

33. Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; Leon, L.D.; Elepaño, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G.; Maningo, J.; Tseng, V. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models **2022**.

34. Gilson, A.; Safranek, C.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.; Chartash, D. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment **2023**.

35. Li, T.; Ma, X.; Zhuang, A.; Gu, Y.; Su, Y.; Chen, W. Few-shot In-context Learning on Knowledge Base Question Answering **2023**.

36. Seegmiller, P.; Gatto, J.; Basak, M.; Cook, D.J.; Ghasemzadeh, H.; Stankovic, J.; Preum, S. The Scope of In-Context Learning for the Extraction of Medical Temporal Constraints **2023**.

37. Liu, S.; Cai, Z.; Chen, G.; Li, X. Towards Better Understanding of In-Context Learning Ability from In-Context Uncertainty Quantification **2024**.

38. Zhang, X.; Ghosh, S.; Bansal, C.; Wang, R.; Ma, M.J.; Kang, Y.; Rajmohan, S. Automated Root Causing of Cloud Incidents using In-Context Learning with GPT-4 **2024**.

39. Zhao, H.; Cai, Z.; Si, S.; Ma, X.; An, K.; Chen, L.; Liu, Z.; Wang, S.; Han, W.; Chang, B. MMICL: Empowering Vision-language Model with Multi-Modal In-Context Learning **2023**.

40. Zhang, Y.; Zhou, K.; Liu, Z. What Makes Good Examples for Visual In-Context Learning? **2023**.

41. Wang, Z.; Jiang, Y.; Lu, Y.; Shen, Y.; He, P.; Chen, W.; Wang, Z.; Zhou, M. In-Context Learning Unlocked for Diffusion Models **2023**.

42. Zhang, M.; He, J.; Lei, S.; Yue, M.; Wang, L.; Lu, C.T. Can LLM find the green circle? Investigation and Human-guided tool manipulation for compositional generalization **2023**.