DAFTAR PUSTAKA

[1] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large Language Models are Zero-Shot Reasoners," May 2022, [Online]. Available: http://arxiv.org/abs/2205.11916

[2] Z. Jie and W. Lu, "Leveraging Training Data in Few-Shot Prompting for Numerical Reasoning," May 2023, [Online]. Available: http://arxiv.org/abs/2305.18170

[3] A. Kong *et al.*, "Better Zero-Shot Reasoning with Role-Play Prompting," Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.07702

[4] R. Wang *et al.*, "Role Prompting Guided Domain Adaptation with General Capability Preserve for Large Language Models," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.02756

[5] A. Kong *et al.*, "Self-Prompt Tuning: Enable Autonomous Role-Playing in LLMs," Jul. 2024, [Online]. Available: http://arxiv.org/abs/2407.08995

[6] Z. Han and Z. Wang, "Rethinking the Role-play Prompting in Mathematical Reasoning Tasks," in *Proceedings of the 1st Workshop on Efficiency, Security, and Generalization of Multimedia Foundation Models*, New York, NY, USA: ACM, Oct. 2024, pp. 13–17. doi: 10.1145/3688864.3689149.

[7] W. Ma *et al.*, "Combining Fine-Tuning and LLM-based Agents for Intuitive Smart Contract Auditing with Justifications," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.16073

[8] Z. Wang *et al.*, "Re-TASK: Revisiting LLM Tasks from Capability, Skill, and Knowledge Perspectives," Aug. 2024, [Online]. Available: http://arxiv.org/abs/2408.06904

[9] L. Zheng *et al.*, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.05685

[10] H. Ye *et al.*, “Ontology-enhanced Prompt-tuning for Few-shot Learning,” in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, Association for Computing Machinery, Inc, Apr. 2022, pp. 778–787. doi: 10.1145/3485447.3511921.

[11] Y. He, J. Chen, H. Dong, and I. Horrocks, “Exploring Large Language Models for Ontology Alignment,” Sep. 2023, [Online]. Available: http://arxiv.org/abs/2309.07172

[12] H. Huang *et al.*, “An Empirical Study of LLM-as-a-Judge for LLM Evaluation: Fine-tuned Judge Model is not a General Substitute for GPT-4,” Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.02839

[13] K. Forth and A. Borrmann, “Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM,” *Journal of Building Engineering*, vol. 95, Oct. 2024, doi: 10.1016/j.jobe.2024.110312.

[14] L. Li, Y. Zhang, and L. Chen, “Prompt Distillation for Efficient LLM-based Recommendation,” in *International Conference on Information and Knowledge Management, Proceedings*, Association for Computing Machinery, Oct. 2023, pp. 1348–1357. doi: 10.1145/3583780.3615017.

[15] C. H. Song, J. Wu, C. Washington, B. M. Sadler, W.-L. Chao, and Y. Su, “LLM-Planner: Few-Shot Grounded Planning for Embodied Agents with Large Language Models.” [Online]. Available: https://osu-nlp-group.github.io/LLM-Planner/

[16] Z. Lai, X. Zhang, and S. Chen, “Adaptive Ensembles of Fine-Tuned Transformers for LLM-Generated Text Detection,” Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.13335

[17] T. Komarlu, M. Jiang, X. Wang, and J. Han, “OntoType: Ontology-Guided and Pre-Trained Language Model Assisted Fine-Grained Entity Typing,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2024, pp. 1407–1417. doi: 10.1145/3637528.3671745.

[18] X. Ma, G. Fang, and X. Wang, “LLM-Pruner: On the Structural Pruning of Large Language Models.” [Online]. Available: https://github.com/horseee/LLM-Pruner

[19] C. Kong *et al.*, “SHARP: Unlocking Interactive Hallucination via Stance Transfer in Role-Playing LLMs,” Nov. 2024, [Online]. Available: http://arxiv.org/abs/2411.07965

[20] C. Snell, J. Lee, K. Xu, and A. Kumar, “Scaling LLM Test-Time Compute Optimally can be More Effective than Scaling Model Parameters,” Aug. 2024, [Online]. Available: http://arxiv.org/abs/2408.03314

[21] B. Zhang, Z. Liu, C. Cherry, and O. Firat, “When Scaling Meets LLM Finetuning: The Effect of Data, Model and Finetuning Method,” Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.17193

[22] C. Jeong, “Fine-tuning and Utilization Methods of Domain-specific LLMs.”

[23] Y. Ge *et al.*, “OpenAGI: When LLM Meets Domain Experts.” [Online]. Available: https://github.com/agiresearch/OpenAGI.

[24] A. Njifenjou, V. Sucal, B. Jabaian, and F. Lefèvre, “Role-Play Zero-Shot Prompting with Large Language Models for Open-Domain Human-Machine Conversation,” Jun. 2024, [Online]. Available: http://arxiv.org/abs/2406.18460

[25] R. Diky Dermawan, “Meningkatkan Kinerja Output ChatGPT Melalui Teknik Prompt Engineering Yang Dapat Dikustomisasi,” *INNOVATIVE: Journal Of Social Science Research*, vol. Volume 4, 2024.

[26] J. Sun *et al.*, “Enhancing Chain-of-Thoughts Prompting with Iterative Bootstrapping in Large Language Models,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.11657

[27] H. Jin *et al.*, “LLM Maybe LongLM: SelfExtend LLM Context Window Without Tuning,” 2024. [Online]. Available: https://github.com/datamllab/LongLM.

[28] N. Sager *et al.*, “Rethinking Retrieval Automated Fine-Tuning in an evolving LLM landscape,” 2024. [Online]. Available:

https://scholar.smu.edu/datasciencereviewAvailableat:https://scholar.smu.edu/datasciencereview/vol8/iss2/2http://digitalrepository.smu.edu.

[29] Z. Gekhman *et al.*, “Does Fine-Tuning LLMs on New Knowledge Encourage Hallucinations?,” May 2024, [Online]. Available: http://arxiv.org/abs/2405.05904

[30] M. R. J, K. VM, H. Warrier, and Y. Gupta, “Fine Tuning LLM for Enterprise: Practical Guidelines and Recommendations,” Mar. 2024, [Online]. Available: http://arxiv.org/abs/2404.10779

[31] B. Zhang, Z. Liu, C. Cherry, and O. Firat, “When Scaling Meets LLM Finetuning: The Effect of Data, Model and Finetuning Method,” Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.17193

[32] S. Li, L. Yao, J. Gao, L. Zhang, and Y. Li, “Double-I Watermark: Protecting Model Copyright for LLM Fine-tuning,” Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.14883

[33] H. T. Mai, C. X. Chu, and H. Paulheim, “Do LLMs Really Adapt to Domains? An Ontology Learning Perspective,” Jul. 2024, [Online]. Available: http://arxiv.org/abs/2407.19998

[34] F. Wu, N. Zhang, S. Jha, P. McDaniel, and C. Xiao, “A New Era in LLM Security: Exploring Security Concerns in Real-World LLM-based Systems,” Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.18649

[35] S. Das, A. Tariq, T. Santos, S. S. Kantareddy, and I. Banerjee, “Recurrent Neural Networks (RNNs): Architectures, Training Tricks, and Introduction to Influential Research,” in *Neuromethods*, vol. 197, Humana Press Inc., 2023, pp. 117–138. doi: 10.1007/978-1-0716-3195-9_4.

[36] F. Shan, X. He, D. J. Armaghani, and D. Sheng, “Effects of data smoothing and recurrent neural network (RNN) algorithms for real-time forecasting of tunnel boring machine (TBM) performance,” *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 16, no. 5, pp. 1538–1551, May 2024, doi: 10.1016/j.jrmge.2023.06.015.

[37] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, “Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews,” in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 728–735. doi: 10.1016/j.procs.2021.01.061.

[38] W. Xu, J. Chen, Z. Ding, and J. Wang, “Text Sentiment Analysis and Classification Based on Bidirectional Gated Recurrent Units (GRUs) Model.”

[39] Z. Niu *et al.*, “Recurrent attention unit: A new gated recurrent unit for long-term memory of important parts in sequential data.”

[40] M. K. H. Briman and B. Yildiz, “Beyond ROUGE: A Comprehensive Evaluation Metric for Abstractive Summarization Leveraging Similarity, Entailment, and Acceptability,” *International Journal on Artificial Intelligence Tools*, vol. 33, no. 5, Aug. 2024, doi: 10.1142/S0218213024500179.

[41] S. S. Amin and L. Ragha, “Text Generation and Enhanced Evaluation of Metric for Machine Translation,” 2021, pp. 1–17. doi: 10.1007/978-981-15-8530-2_1.

[42] C. Han and X. Lu, “Beyond BLEU: Repurposing neural-based metrics to assess interlingual interpreting in tertiary-level language learning settings,” *Research Methods in Applied Linguistics*, vol. 4, no. 1, Apr. 2025, doi: 10.1016/j.rmal.2025.100184.

[43] S. Musa Shafi’, H. U. Suru, and D. Gabi, “Evaluating English to Nupe Machine Translation Model Using BLEU.” [Online]. Available: https://www.iuokada.edu.ng/journals/nijesr/

[44] M. Barbella and G. Tortora, “ROUGE metric evaluation for Text Summarization techniques.” [Online]. Available: https://ssrn.com/abstract=4120317