

Reading List

1. LLM Basics:

- *Attention Is All You Need*. Vaswani, et al. <https://arxiv.org/abs/1706.03762>
- *GPT-4 Technical Report*. OpenAI. <https://arxiv.org/abs/2303.08774>

2. LLM for Code

- *Evaluating Large Language Models Trained on Code*. Chen, et al. <https://arxiv.org/abs/2107.03374>.
- *Code Llama: Open Foundation Models for Code*. Rozière, et al. <https://ai.meta.com/research/publications/code-llama-open-foundation-models-for-code/>.

3. LLM as Programming Assistant

- *A Large-Scale Survey on the Usability of AI Programming Assistants: Successes and Challenges*. Liang, et al. <https://dl.acm.org/doi/abs/10.1145/3597503.3608128>.
- *GitHub Copilot AI pair programmer: Asset or Liability?* Dakhel, et al. <https://arxiv.org/abs/2206.15331>.

4. LLM for Collaborative Coding

- *MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework*. Hong, et al. <https://arxiv.org/abs/2308.00352>.
- *Communicative Agents for Software Development*. Qian, et al. <https://arxiv.org/abs/2307.07924>.

5. Augmented LLM with Tools

- *Augmented Language Models: a Survey*. Mialon, et al. <https://arxiv.org/abs/2302.07842>.
- *Toolformer: Language Models Can Teach Themselves to Use Tools*. Schick, et al. <https://arxiv.org/abs/2302.04761>.

6. LLM for Unit Testing

- *Automated Unit Test Improvement using Large Language Models at Meta*. Alshahwan and et al. <https://arxiv.org/abs/2402.09171>.
- *An Empirical Evaluation of Using Large Language Models for Automated Unit Test Generation*. Schäfer and et al. <https://arxiv.org/abs/2302.06527>.

7. LLM for Bug Hunting

- *Large Language Models are Few-shot Testers: Exploring LLM-based General Bug Reproduction*. Kang and et al. <https://arxiv.org/abs/2209.11515>.
- *PentestGPT: An LLM-empowered Automatic Penetration Testing Tool*. Deng, et al. <https://arxiv.org/abs/2308.06782>.

8. LLM for Debugging

- *Teaching Large Language Models to Self-Debug*. Chen, et al. <https://arxiv.org/abs/2304.05128>.
- *Reflexion: Language Agents with Verbal Reinforcement Learning*. Shinn, et al. <https://arxiv.org/abs/2303.11366>.

9. Reasoning with LLM

- *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. Wei, et al. <https://arxiv.org/abs/2201.11903>.
- *Program of Thoughts Prompting: Disentangling Computation from Reasoning for Numerical Reasoning Tasks*. Chen, et al. <https://arxiv.org/abs/2211.12588>.

10. LLM for Theorem Proving

- *Generative Language Modeling for Automated Theorem Proving*. Polu and Sutskever.
<https://arxiv.org/abs/2009.03393>.
- *LeanDojo: Theorem Proving with Retrieval-Augmented Language Models*. Yand, et al.
https://proceedings.neurips.cc/paper_files/paper/2023/hash/4441469427094f8873d0fecb0c4e1cee-Abstract-Datasets_and_Benchmarks.html