Book Review

Large Language Models and Prompt Engineering

A Comprehensive Guide with Practical Applications

Publication Details

- Authors: Mohammad Amin Kuhail, Mohammad Tubishat, Sarah Kohail, and Maha Hadid
- Publisher: CRC Press (First Edition, 2027)
- ISBNs: 9781041041184 (hbk) | 9781041041160 (pbk) | 9781003626909 (ebk)
- DOI: 10.1201/9781003626909

Overview

Large Language Models and Prompt Engineering provides a structured, accessible introduction to the theory, design, and real-world application of Large Language Models (LLMs). As generative AI becomes deeply embedded across various industries, this book offers a balanced perspective that seamlessly blends technical foundations with practical guidance and critical reflection.

LLMs have fundamentally revolutionized fields such as customer support, education, and mental health care. The daily use of these models has significantly boosted productivity by automating repetitive tasks and streamlining decision making processes. This book serves as a comprehensive roadmap for understanding and leveraging this transformative technology.

Technical Foundations: How LLMs Work

The authors provide a clear, step-by-step breakdown of how LLMs generate text through a structured sequence of computational steps:

1. Tokenization: Input text is broken into smaller units called tokens and converted into numerical representations.

2. Embedding: These tokens pass through an embedding layer, mapping them to high-dimensional vectors that capture semantic meanings and relationships.

3. Positional Encoding: Positional data is added to the embeddings, informing the model of each word's sequence order.

4. Self-Attention: The vectors pass through a self-attention mechanism, allowing the model to compute attention weights and determine which words most strongly influence one another.

5. Feed-Forward Processing: The attention-weighted data is refined by a feed forward neural network.

6. Probability Generation: The output generates a set of logits (unnormalized scores for potential next tokens), which are converted into probability distributions (Cramer, 2003).

7. Token Selection: The model selects the token with the highest probability as the next word, iteratively repeating this process to produce coherent, contextually appropriate text.

The Art of Prompt Engineering

Simply put, prompt engineering is the technique of guiding LLMs to perform specific tasks by embedding task descriptions directly into the input (Reddy et al., 2024). Researchers utilize these techniques to expand LLM capabilities across complex tasks like question answering and arithmetic reasoning (Shi et al., 2024), while developers use them to design robust interfaces between LLMs and external tools (promptingguide.ai, 2025).

Practical Application and Tools

To maximize the learning experience, the book relies heavily on hands-on examples designed for cloud-based development environments. The authors highly recommend Google Colab, a platform that allows users to write, execute, and share Python code via interactive Jupyter Notebooks in a web browser. Colab significantly lowers the barrier to entry for AI and machine learning experimentation, providing necessary computational resources for rapid prototyping and research. All source code is provided within the book and hosted on GitHub.

Chapter-by-Chapter description

- Chapter 2: Prompt Engineering Techniques: Explores how prompt engineering shapes LLM outputs. It begins with basic techniques (zero shot and few shot prompting) before advancing to complex problem solving methods like chain of thought, self consistency, tree of thought, self ask, generated knowledge, role play, and ReAct. It concludes with orchestration strategies (meta prompting, prompt chaining, and synthetic prompting) to manage multi-step tasks.

- Chapter 3: Frameworks for AI Solutions: Introduces structured approaches that scale individual prompts into complete systems. It divides frameworks into two categories: *prompting frameworks* (which formalize prompt construction) and *application frameworks* (which scale these principles for orchestration and integration).

- Chapter 4: LLM Architecture: Dives into the transformer based systems powering modern AI. By explaining attention mechanisms, context windows, and tokenization, this chapter bridges the gap between prompt design and underlying model functionality, explaining why certain input styles yield better results.

- Chapters 5–7: Domain-Specific Applications: Explores high-impact use cases across three major sectors:

o *Healthcare (Ch. 5):* Clinical documentation and patient communication.

o *Customer Support (Ch. 6):* Automating responses and enhancing customer experience.

o *Education (Ch. 7):* Tutoring, lesson planning, and simulation based learning.

- Chapter 8: Evaluation and Fine-Tuning Focuses on continuous performance monitoring. It introduces evaluation metrics, feedback mechanisms, human review, scoring systems, and task-specific tuning to refine prompts and optimize AI systems over time.

- Chapter 9: Ethics and Future Trends Previews crucial governance topics, including bias, data privacy, misinformation, and accountability. It also looks ahead to emerging frontiers like multimodal AI, autonomous agents, and prompt automation, encouraging responsible AI development.

Conclusion

By the end of *Large Language Models and Prompt Engineering*, readers will possess a clear mental model of LLMs, spanning from foundational machine learning concepts to modern transformer architectures.

The book ensures readers gain practical mastery over prompt engineering and frameworks, enabling them to reliably steer, evaluate, and optimize model behavior. Furthermore, it equips professionals to translate these skills into real world applications across healthcare, education, and customer support, all while maintaining a critical, ethical perspective necessary for the responsible and scalable use of AI.

References

[1] Cramer, J. S. (2003). The Origins and Development of the Logit Model. Cambridge: Cambridge UP.

[2] Reddy, C. K. K., Anoushka, P., Draksharapu, A., Doss, S. (2024). Beyond text: Analyzing artificial intelligence models through prompt engineering. In: I. U. Khan, H. Taherdoost, M. Madanchian, M. Ouaissa, S. El Hajjami H. Rahman, eds. Future Tech Startups and Innovation in the Age of AI. Boca Raton, FL: CRC Press.

[3] Shi, C., Zhai, R., Song, Y., Yu, J., Li, H., Wang, Y., Wang, L. (2023). Few-shot sentiment analysis based on adaptive prompt learning and contrastive learning. Information Technology and Control, 52(4), 1058–1072.

Hathairat Ketmaneechairat
Department of Information and Production Technology Management
College of Industrial Technology
King Mongkut's University of Technology North Bangkok
Thailand