CS-435 Reading Assignment (Ungraded)

Articles to Read and Key Points

1. **Retrieval-Augmented Generation for Large Language Models: A Survey**
   `https://arxiv.org/abs/2312.10997`

   - Understand how retrieval mechanisms address the limitations of LLMs (e.g., hallucinations, outdated knowledge).
   - Focus on how external databases can be leveraged for more accurate and traceable reasoning.
   - Note the similarities and differences in Naive, Advanced, and Modular RAG systems.
   - Examine the evaluation frameworks and benchmarks introduced for RAG.

2. **Large Language Models: A Survey**
   Shervin Minaee, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, Jianfeng Gao
   `https://arxiv.org/abs/2402.06196`

   - Review the evolution of popular LLM families (GPT, LLaMA, PaLM).
   - Pay attention to scaling laws and how training on vast text corpora leads to general-purpose language understanding.
   - Note the common datasets used for fine-tuning and evaluation of large models.
   - Observe open challenges such as model bias, hallucination, and potential future research directions.

3. **Instruction Tuning for Large Language Models: A Survey**
   `https://arxiv.org/pdf/2308.10792`

   - Focus on how supervised fine-tuning (SFT) can align model outputs with human intentions.
   - Understand the (INSTRUCTION, OUTPUT) training paradigm and its impact on controllability.
   - Examine the various approaches and datasets used for instruction tuning.
   - Investigate known pitfalls, such as data quality issues, and proposed mitigation strategies.

4. **Machines of Loving Grace: How AI Could Transform the World for the Better**
   Dario Amodei
   `https://darioamodei.com/machines-of-loving-grace`

   - Focus on the essay's optimistic vision for AI's impact over the next decade.

- Note the arguments for potential transformations in biology, health, governance, and economic development.
- Consider how societal factors (e.g., regulations, inequalities) may interact with rapidly progressing AI.
- Reflect on the balance between leveraging AI for human benefit and mitigating risks.