Retrieval-Augmented Generation for  
Knowledge-Intensive NLP Tasks

Outline

1. LLM目前所面臨的問題
2. 帶入優化LLM的方式(i)Prompt Engineer (ii)Fine-tuning (iii)RAG 相互比較
3. RAG應用場景介紹
4. 論文細節描述
5. Architecture
   1. Query Encoder
   2. MIPS method
   3. Document Index
   4. Generator
6. Methods
   1. Retriever: pre-trained retriever
   2. Generator: pre-trained encoder-decoder
7. Models
   1. RAG-Sequence model
   2. RAG-Token model
   3. Retriever: Dense Passage Retriever (DPR)
   4. Generator: (BERT)
   5. Training
   6. Decoding
8. Experiment
   1. Open-Domain Question Answering
   2. Abstract Question Answering
   3. Jeopardy Question Answering
   4. Fact Verification
9. Result
10. Conclusion
11. Demo

<https://arxiv.org/pdf/2005.11401v1.pdf>