Congratulations! You passed!

Grade received 80% **Latest Submission Grade** 80% **To pass** 80% or higher

Go to next item

1. **True or False.** Because of the knowledge cut-off, an LLM cannot answer questions about today's news. But with RAG to supply it articles from the news, it would be able to. **1 / 1 point**

- (•) True
- ( ) False

Correct
RAG provides an LLM with additional information and context from external documents that it can reason through to answer a question. So RAG would enable an LLM to answer questions about current news articles.

2. You want to build an application to answer questions based on information found in your emails. Which of the following is the most appropriate technique? **1 / 1 point**

- ○ Prompting (without RAG), where we iteratively refine the prompt until the LLM gets the answers right.
- ○ Pretraining an LLM on your emails.
- ◉ RAG, where the LLM is provided additional context based on retrieving emails relevant to your question.
- ○ Fine-tuning an LLM on your emails, whereby we take a pre-trained LLM and further train it on your emails.

Correct
RAG can be used to give an LLM access to new, external sources of information that it can reason through to formulate an answer to your question. So a RAG system that provides access to your emails is the best approach to get an LLM to answer your questions.

3. What does the idea of using an LLM as a reasoning engine refer to? **1 / 1 point**

- ◉ This refers to the idea of using an LLM not as a source of information, but to process information (wherein we provide it the context it needs, through techniques like RAG).
- ○ The idea of using an LLM to play games (like chess) that require complex reasoning, but having its output moves in the game.
- ○ Reasoning engine is another term for RAG.
- ○ This refers to pretraining an LLM on a lot of text so that it acquires general reasoning capabilities.

Correct
The ability of LLMs to process information is one of the features that makes them such powerful and useful tools.

4. **True or False.** By making trusted sources of information available to an LLM via RAG, we can reduce the risk of hallucination. **1 / 1 point**

- ○ False, because giving the LLM more information only confuses the LLM more and causes it to be more likely to hallucinate.
- ○ True, because the LLM is now restricted to outputting paragraphs of text exactly as written in the provided document, which we trust.
- ○ False, because the LLM has learned from a lot of text from the internet (perhaps >100 billion words) to hallucinate, so adding one more short piece of text to the prompt as in RAG won't make any meaningful difference.
- ◉ True, because RAG allows the LLM to reason through accurate information retrieved from a trusted source to arrive at the correct answer.

✓ **Correct**
RAG can be used to give an LLM access to new, trusted sources of information that it can reason through to formulate an answer to your question. This helps prevent the model from hallucinating because it doesn't know the answer.

5. An ecommerce company is building a software application to route emails to the right department (Apparel, Electronics, Home Appliances, etc.) It wants to do so with a small, 1 billion parameter model, and needs high accuracy. Which of these is an appropriate technique? **0 / 1 point**

- ○ Fine-tune a 1 billion parameter model on around 1 billion examples of emails and the appropriate department.
- ○ Fine-tune a 1 billion parameter model on around 1,000 examples of emails and the appropriate department.
- ○ Pretrain a 1 billion parameter model on around 1 billion examples of emails and the appropriate department.
- ◉ Pretrain a 1 billion parameter model on around 1,000 examples of emails and the appropriate department.

✗ **Incorrect**
Pretraining refers to the initial training of an LLM, where it learns to repeatedly predict the next word from billions to trillions of words from the internet. Pretraining a model on just 1000 email examples would not results in a useful LLM, so this is not the right technique to use.