ROSHAN POUDEL - Problem 3 HW#3

How ChatGPT Actually Works
https://www.assemblyai.com/blog/how-chatgpt-actually-works/

ChatGPT is a Large Language Model (LLM) based on GPT-3 architecture, designed to communicate with humans using appropriate responses. It has been fine-tuned for various language generation tasks, such as translation, summarization, text completion, and question-answering. It is a transformer-based neural network that has over 175 billion parameters, making it one of the largest language models. It is pre-trained and can multitask, simultaneously translating, summarizing, and answering questions. ChatGPT processes user input by tokenizing the text, embedding it, using encoder-decoder attention, generating a probability distribution for all possible outputs, and then generating an output answer. The user receives a text response in real time.

A model which is only trained to predict the next word (or a masked word) in a text sequence, may not necessarily be learning some higher-level representations of its meaning. As a result, the model struggles to generalize to tasks or contexts that require a deeper understanding of language. The creators of ChatGPT used both Supervised Learning and Reinforcement Learning techniques to mitigate this misalignment problem of LLM. The developers employed a specific technique called Reinforcement Learning from Human Feedback (RLHF) which integrates human feedback into the training process to minimize the generation of harmful, inaccurate, and biased outputs.

Reinforcement Learning from Human Feedback consists of three steps:

1. Supervised fine-tuning
2. Mimic human preferences
3. Proximal policy optimization.

In the supervised fine-tuning step, a small, high-quality curated dataset is used to fine-tune a pre-trained language model. In the second step, labelers rank the outputs generated by the fine-tuned model, creating a new dataset that is used to train a reward model. Finally, in the proximal policy optimization step, reinforcement learning is used to further fine-tune and improve the supervised fine-tuned model using the reward model.

The performance evaluation of language models trained with the RLHF methodology involves rating the model outputs based on three criteria: helpfulness, truthfulness, and harmlessness. However, the evaluation is based on human input and can be influenced by subjective factors such as the preferences of labelers and researchers, as well as the choice of prompts.

For Extra Credit:

1. **LaMDA** (**Language Model for Dialogue Applications**) - Google
   https://blog.google/technology/ai/lamda/

2. **OPT 175-B - MetaAI**
   https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/

3. **OpenAI**
   https://openai.com/