1. Hello, my name is Kolton Hauck and I will be discussing the paper: Language Models are Few-Shot Learners by Tom b. Brown and company
2. Here is a little bit of background/foundational knowledge for the paper
   1. Natural language Technology, or NLT, encompasses these other commonly used terms:
      1. natural language processing
      2. natural language understanding
      3. natural language generation
3. there are nuance to these terms, but generally, and most often they are referred to as Natural language Processing (NLP).
   1. NLP: primary goal is to enable computers to understand, interpret, and generate human language in a valuable way
   2. today, NN are the most common tools and most effective tools for performing NLP tasks
4. There has been some evolution to how NLP has been accomplished
   1. All of these are still used today depending on the specific application
5. First, there were statistical models, such as naïve bayes and hidden Markov models that have been used to accomplish NLP tasks.
   1. There are also logical rules and embeddings-vector representations of text
6. NN were then applied to NLP. These required task-specific training data. Model sizes were modest and also required task-specific architecture, such as recurrent NN or long short-term memory models
7. Transformers is another NN specific architecture for applying NLP.
   1. The paper corresponding to transformers; ‘attention is all you need’
   2. BERT/GPT are transformers
      1. GPT: generative pre-trained transformer
      2. Bert: Bidirectional Encoder Representations from Transformers
   3. This also introduced training on general data and then fine-tuning w/task-specific data
8. The primary difference between other NN and LLMs is the size. They ramped up the size/number of parameters the NNs have and the task-specific fine-tuning dataset has also become really important
9. Now, although this is nice because yo udon’t have to collect a full training dataset related to your needs- just a smaller fine-tuning dataset, this can still be really hard/infeasible to do depending on your use case
10. That’s where zero/one/few-shot learning comes in
    1. This is an example of transfer learning: using the NN for a different application/different dataset then it was trained for
    2. Also highlights the importance/application of prompt engineering
       1. Your prompts/queries do matter!
    3. ZOF-shot learning is when you give the model zero/one/few examples of queries/answers and the model is expected to pick up the task/patterns without being explicitly told to do so
11. Here is a few-shot example.
    1. As can be seen, the gray text is what the model is given, the black is the actual model output.
    2. The model picks up the task automatically
12. So…the paper
    1. Evaluated GPT-3’s performance sing ZOF shot learning against other models, including other models fine-tuned for specific tasks
    2. Also investigated how increasing parameter size affects the models zof shot capabilities
13. Why does this matter?
    1. Well we can apply LLMs with few shot learning to clinical tasks
14. Here are some results
    1. Here, for a question-answering task, the 175 B parameter gpt-3 with few-shot was able to surpass the fine-tuned state-of-the-art model, with one/zero shot just a few points behind
15. This task is a question-answering task in a more conversational setting. Gpt-3 wasn’t as good as human or the sota benchmark, but still with 0, 1, or 5 examples, the model was able to achieve decent accuracy
16. Here is a task given to gpt-3 to generate fake news articles and humans were required to determine which articles are real or model-generated. This accuracy is the human’s ability to identify real/model-generated . the largest gpt-3 tricked nearly 50% of humans. Really, I don’t know if that’s too hard to do nowadays though
17. Cloze/completion task is a task where the model has to predict a missing word in a text, but usually the missing word has long-range dependencies on the preceding text. Gpt-3 was able to perform very well on this getting fairly close to human capabilities.
18. This paper is a couple years old-talking about gpt-3. Today, gpt-4 is in full swing and is much better in it’s capabilities than gpt-3. This paper goes into gpt-4’s capabilities which only increases our capabilities to exploit these models for clinical-related tasks using zof shot learning. And we’re not limited to gpt-4. There are other respectable models out there that have close or better accuracies than gpt-4, depending on the application.