Analytics Vidhya | DATAHACK SUMMIT

Workshop

NLP using Generative AI

Speaker

Raghav Bali
Staff Data Scientist, Delivery Hero

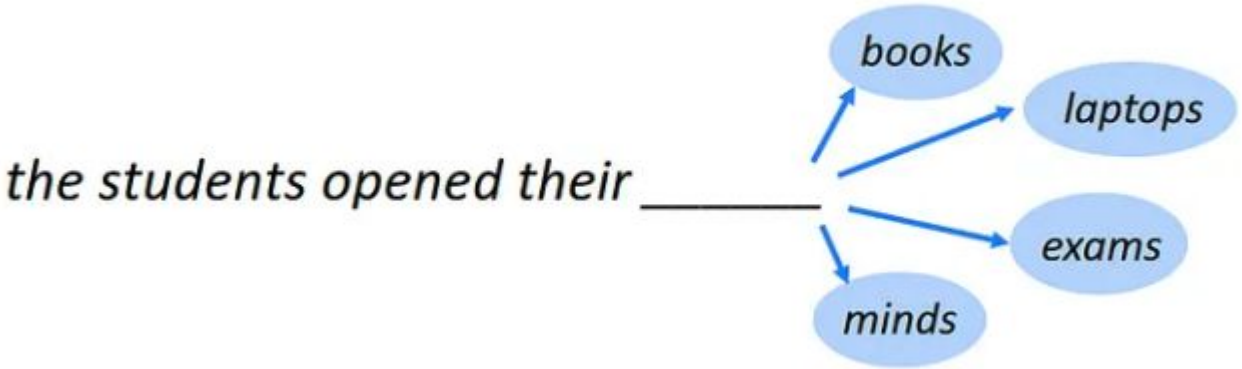
Amar Lalwani
Co-founder, Stealth Startup
Ex - Head of AI and R&D, funtoot

Module 3

Diving into GPT

Language Models

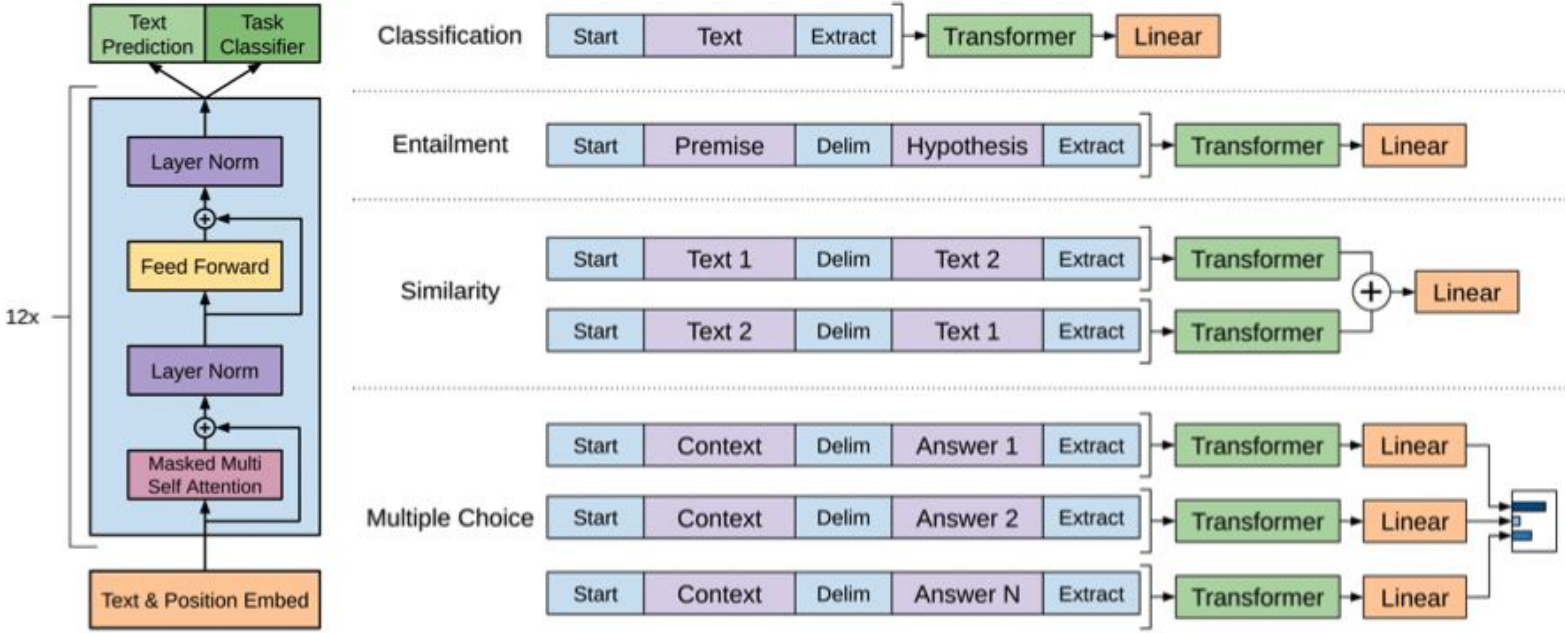
Sophisticated Next Word Predictors



https://medium.com/@antonio.lopardo/the-basics-of-language-modeling-1c8832f21079

Improving Language Understanding by Generative Pre-Training

- followed by (task-agnostic) Discriminative Fine-Tuning



https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

GPT-2: Unsupervised Multitask Learner

- Pre-training transformer language models
- Task specific fine-tuning
  - Eliminating need of task specific architectures
- Limitation: while architecture is task-agnostic, still task specific datasets and fine-tuning needed

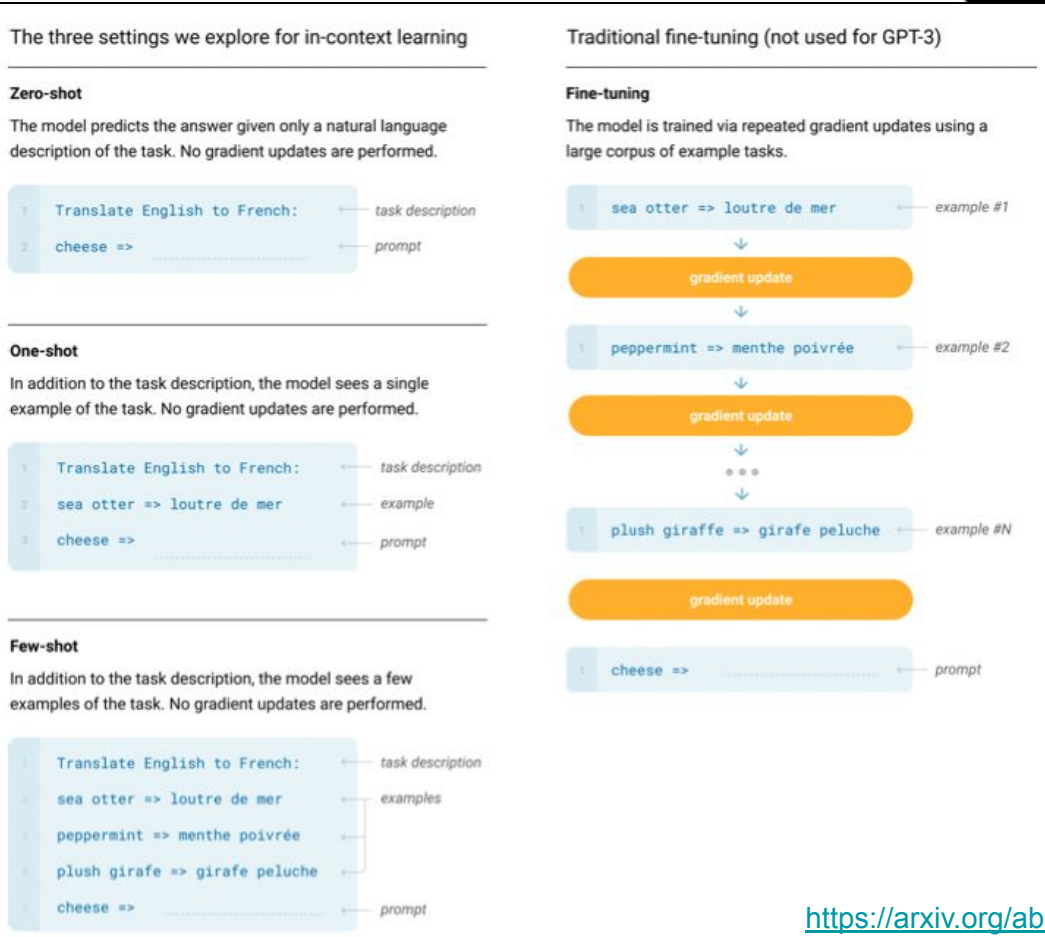
➔ For each task, large high quality dataset needed. May not be possible always

➔ Out of distribution generalization decays with
  - ◆ Expressiveness of the model
  - ◆ Narrowness of the training distribution

➔ Humans to do not require large training datasets to learn most natural language tasks
  - ◆ Brief directive in natural language or at-most a tiny number of demonstrations is enough
  - ◆ This allows humans to seemingly mix together or switch between various tasks and skills

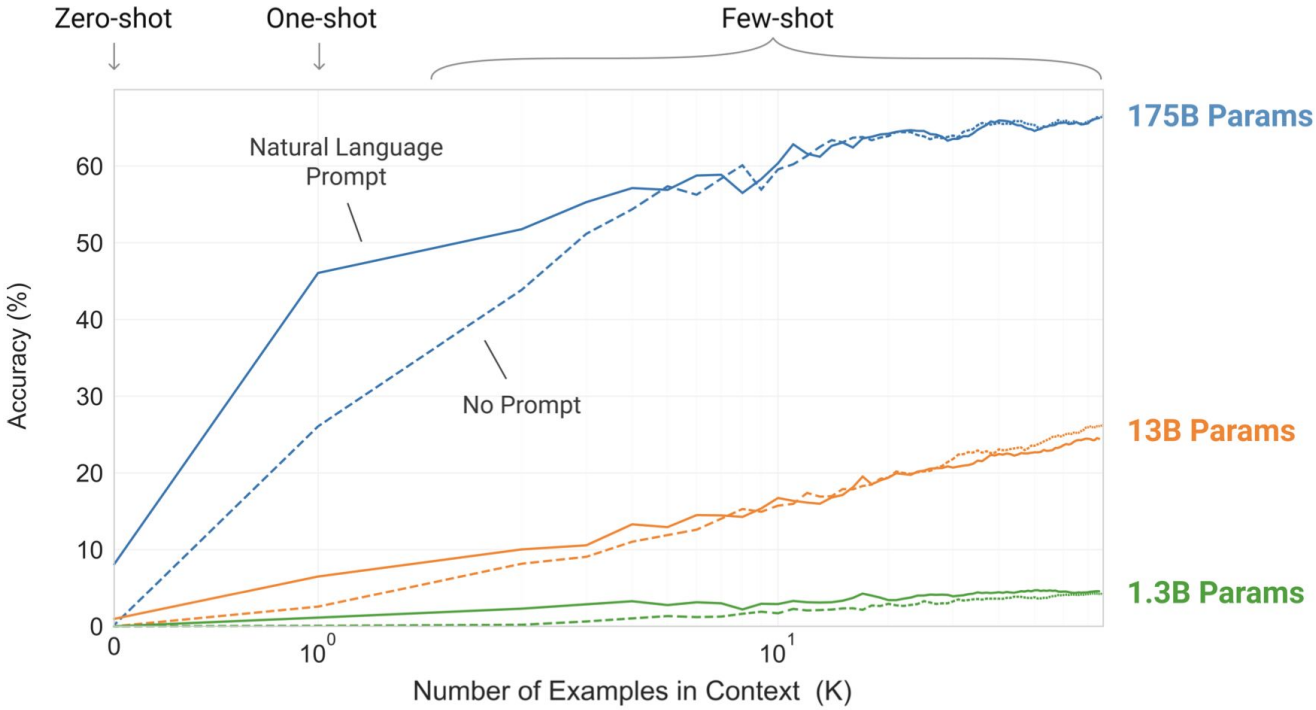
https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

GPT-3: Few-shot Learner

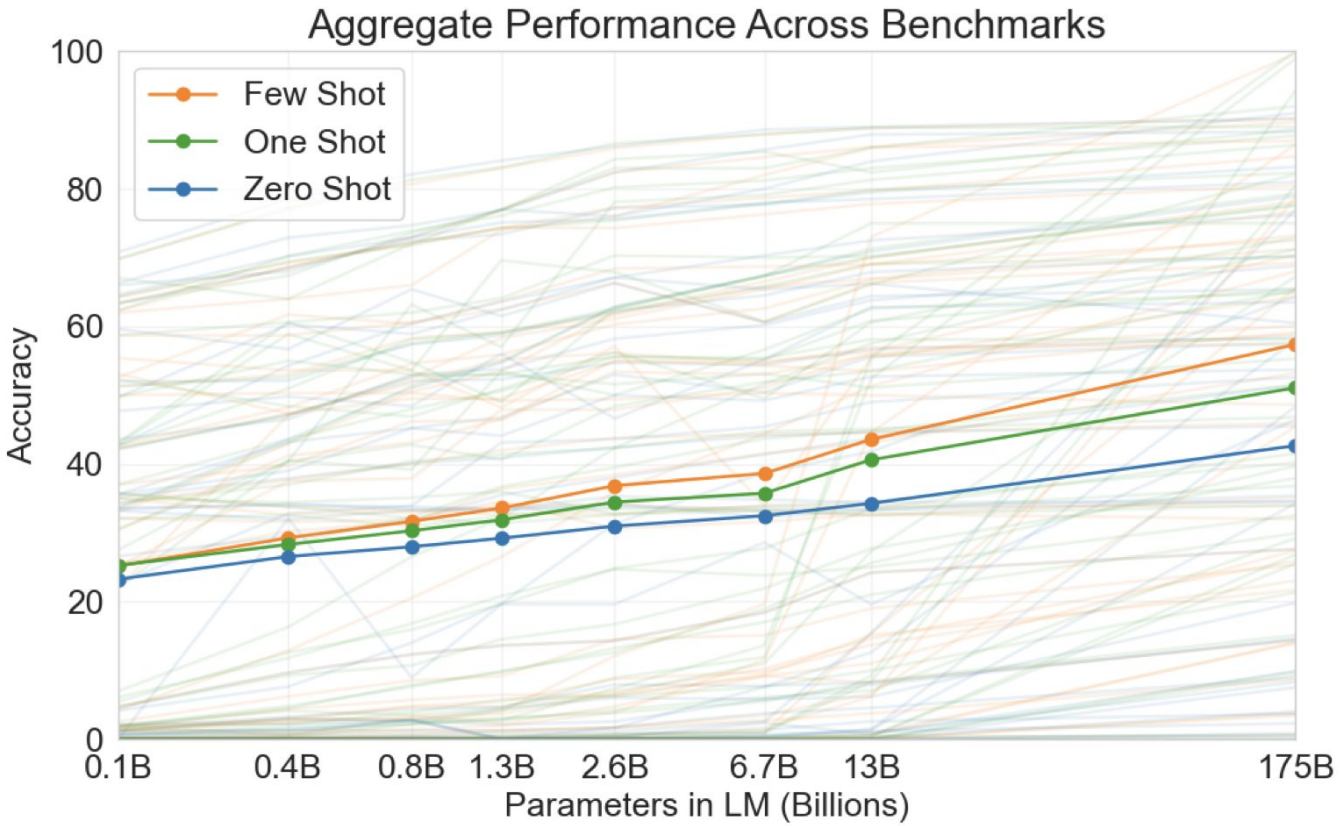


https://arxiv.org/abs/2005.14165

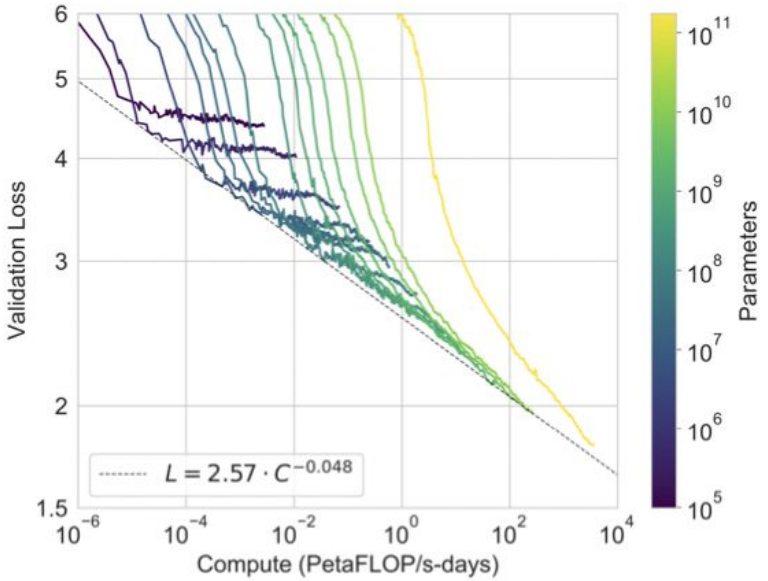
GPT-3 In-context learning: remove random symbols from a word

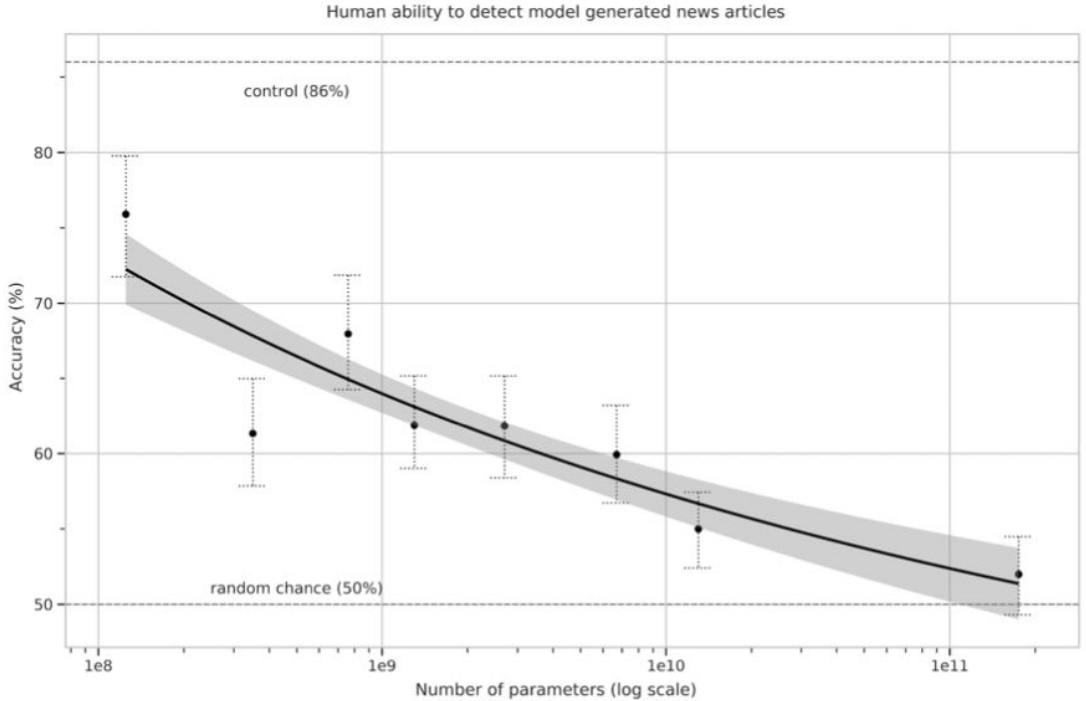


GPT-3: Few-shot Learner



Performance: Power Law





Let us do it!