O'REILLY®

Limitations of LLMs

Bert Gollnick

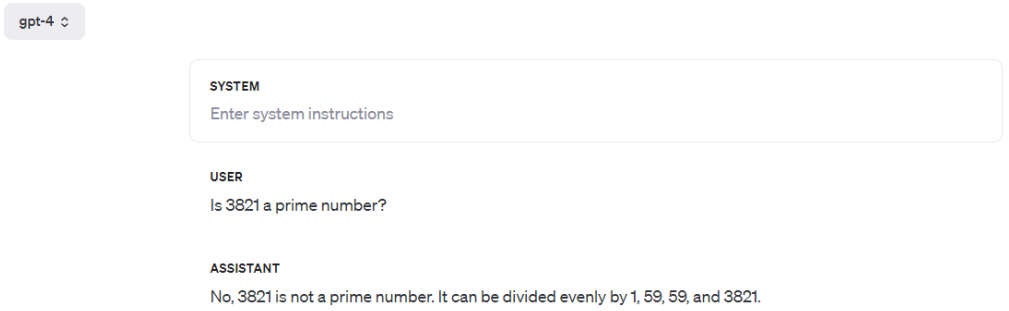
Learning Objectives

By the end of this module, you will:

- Understand the different limitations of LLMs

Hallucinations

- LLM always produces output, even if question hard to answer based on training data
- LLM generates responses that are:
  - factually incorrect
  - not connected to input prompt
  - non-sensical
  - conflicting to its own context



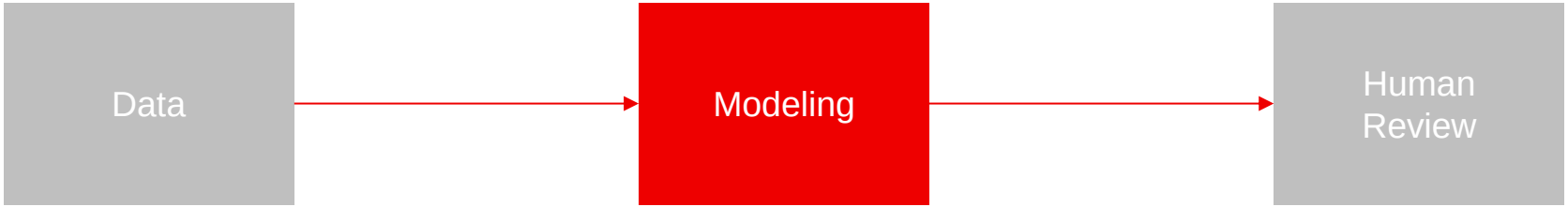
Source: own image created with ChatGPT

Biases: Data

- Historical bias
- Representation bias
- Measurement bias

Source: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

Biases: Modeling



- Evaluation bias
- Aggregation bias

Source: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

Biases: Human Review

- Bias of human reviewer introduced

Source: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

Ethical and social challenges

- Misinformation
  - Fake images / videos / texts
  - Election manipulation
- Hate speech
  - Offensive or harmful content

Source: left image…https://www.lmu.de/en/newsroom/news-overview/news/artificial-intelligence-stemming-the-tide-of-fake-facts.html; right image…https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion/index.html

Limited Reasoning

- Lack of true understanding
- Logical errors

BG

Q: Sofia has 7 apples in her basket. Her friend Emily gives her 3 more bags of apples. Each bag contains 4 apples. How many apples does Sofia have in total?
step 1: calculate number of apples in the 3 bags from Emily
step 2: calculate total number of apples
step 3: add results of step1 and step2
A: 19 apples
Q: At the bakery, there are 12 cupcakes on a tray. Sarah takes 3 cupcakes for herself. Her friend Alex then takes half of the remaining cupcakes. How many cupcakes are left on the tray?
A: ??

Typical math problem that LLMs struggled in the early days

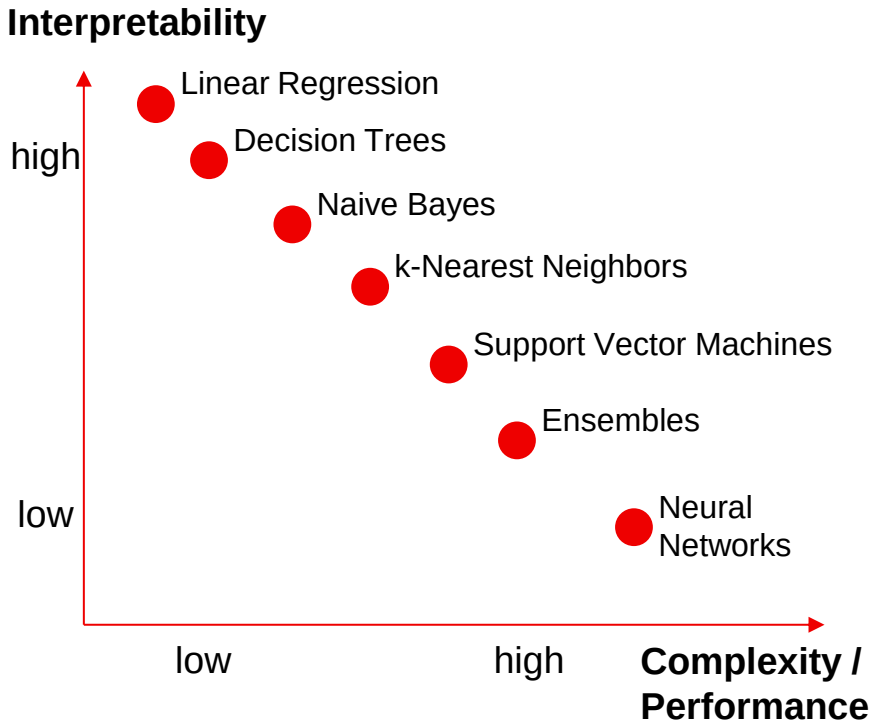
Legal, Regulatory, and IP Issues

- Use of personal data – privacy concerns
- Compliance with data protection regulations
- Content ownership…content might be generated based on copyrighted material

Source: „The New York Times' Copyright Lawsuit Against OpenAI Threatens the Future of AI and Fair Use" https://www.linkedin.com/pulse/new-york-times-copyright-lawsuit-against-ocq1e

Interpretability and Transparency

- Lack of transparency…unclear how LLMs arrive at specific outputs often challenging due to complex internal working
- Interpretability…explaining why a model produced a certain result can be difficult, limiting trust and usability
- Illustrate black box nature of LLMs

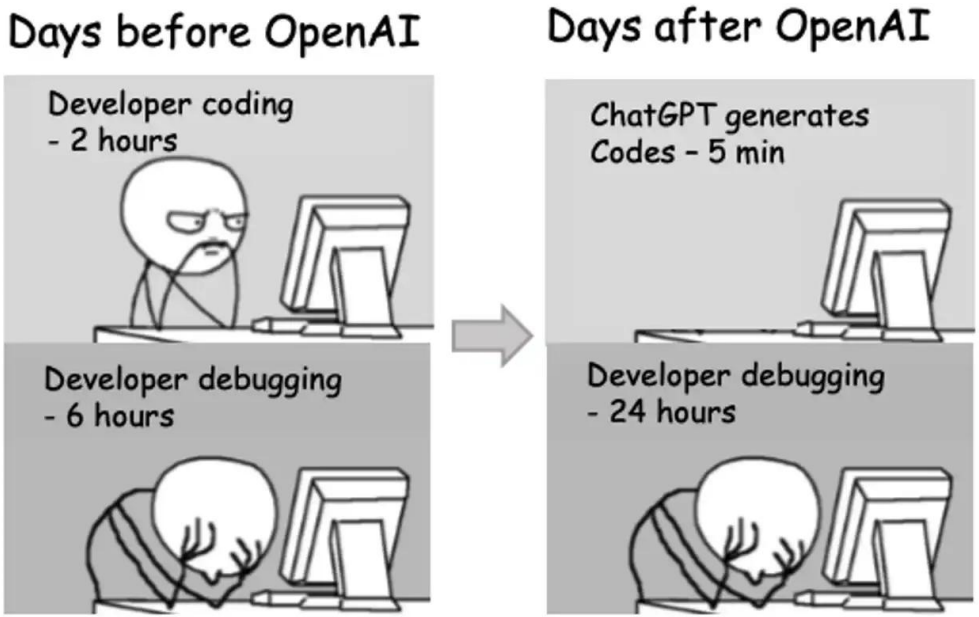


own graph; adapted from https://docs.aws.amazon.com/whitepapers/latest/model-explainability-aws-ai-ml/interpretability-versus-explainability.html

Debugging

- Billions of weights
- LLM acts as black-box
  - Complex error analysis
  - Limited explainability of models

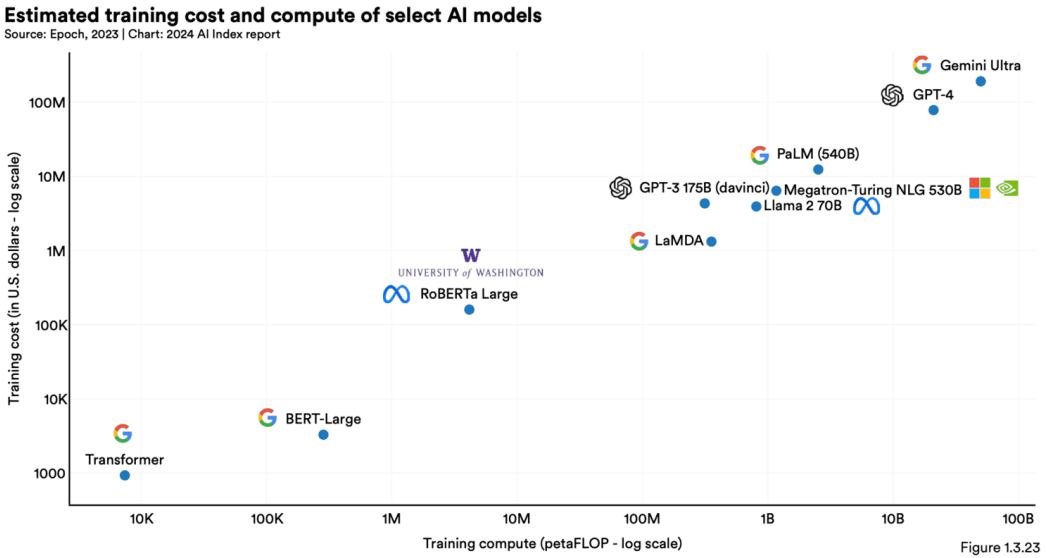
„With great power comes low explainability“



Source: https://grapevine.in/post/76efff35-e220-46b4-8a7f-fe55dd06245e

Computational Resources

- Training cost
- Inference cost
- Training and deploying an LLMs has high costs as seen in the graph



Source: https://aiindex.stanford.edu/report/


O'REILLY®