SYSTEMS ENGINEERING RESEARCH CENTER

CENTER FOR COMPLEX SYSTEMS & ENTERPRISES

STEVENS INSTITUTE OF TECHNOLOGY 1870

Research and Application Workshop

AI4SE & SE4AI

The Use of Large Language Models/"Generative" Systems in Secure and Restricted Settings

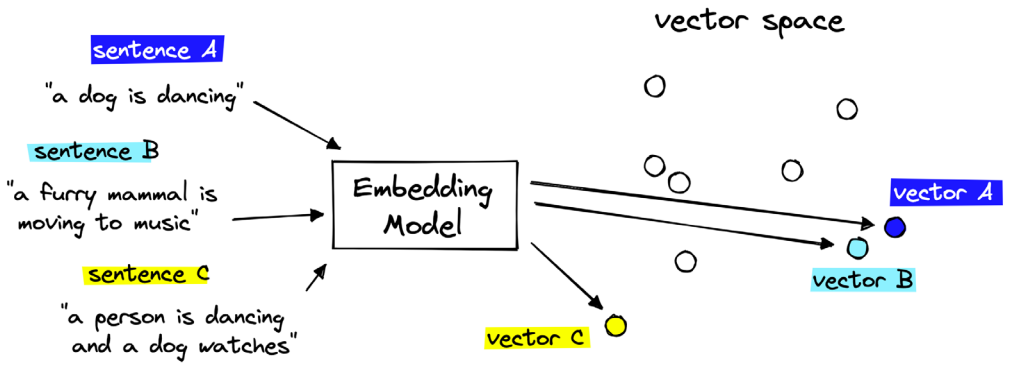
Carlo Lipizzi – clipizzi@stevens.edu

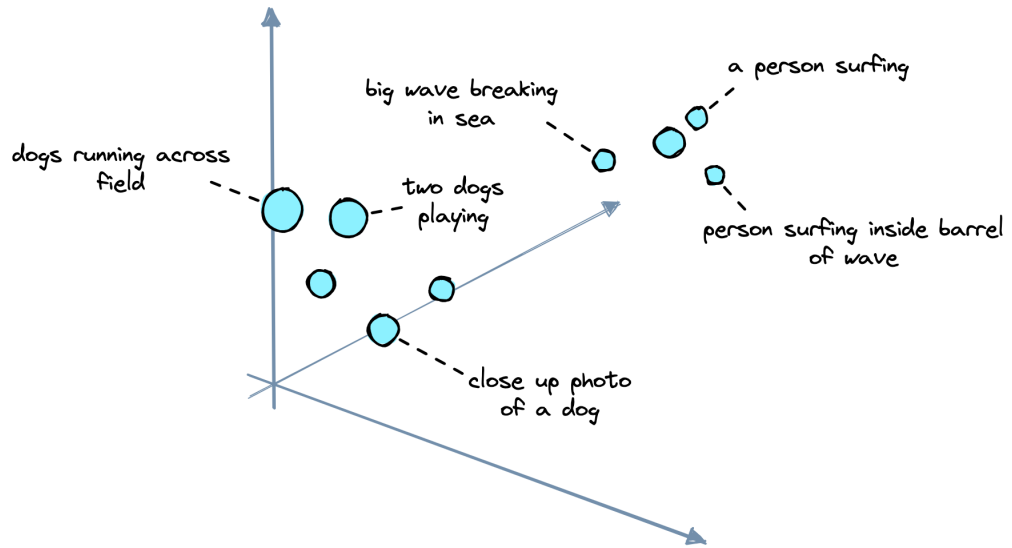
September 2023

What is in an LLM

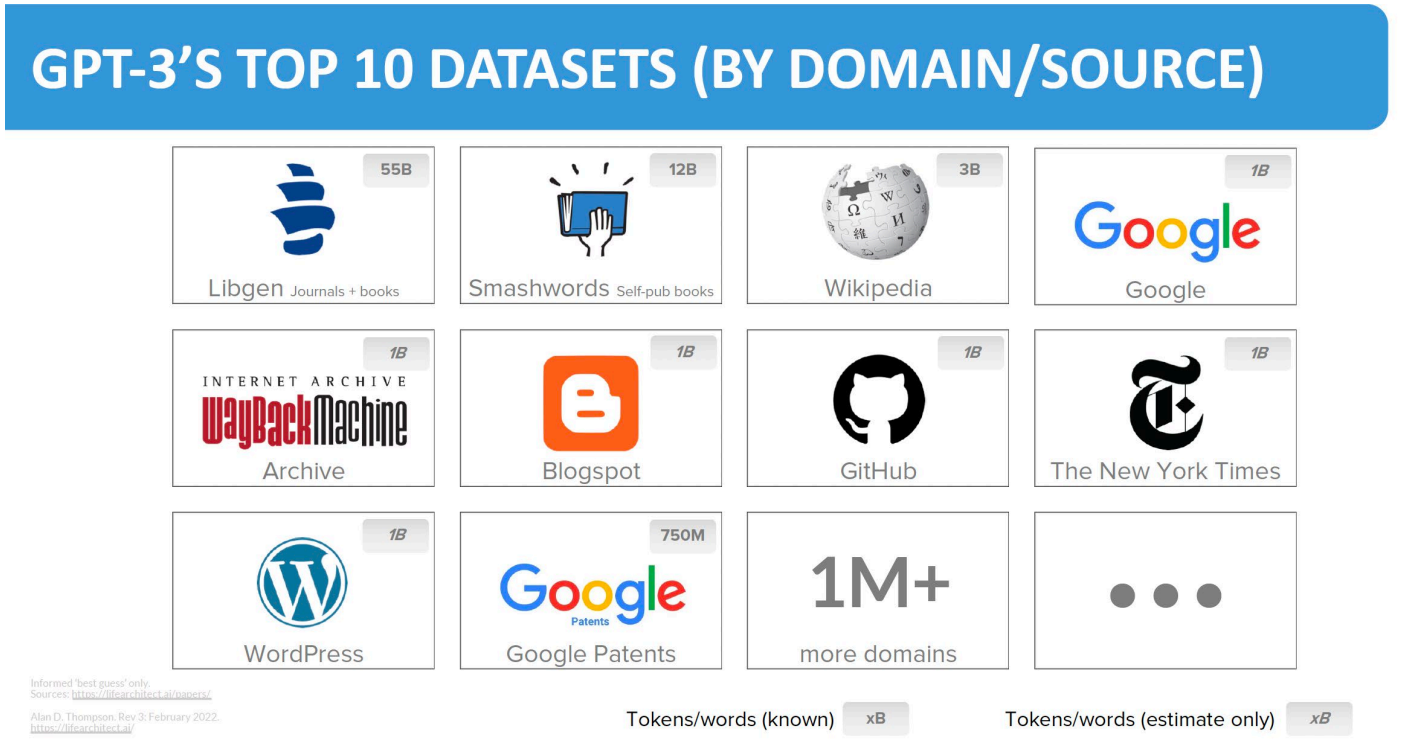
- LLMs are probability-driven pattern matchers with a conversational layer
- They do not “understand” the text
- Bias issue, in terms of data, human reviewers and possibly algorithms
- No traceability of the sources
- High inefficiency. Being based on a “brute force” approach, they perform much worst than human brain, that is using a fraction of the energy to perform much more complex tasks
- Leading LLMs – like ChatGPT – have limited/no domain-specific knowledge
- High cost of training

The concept of “proximity” in LLMs

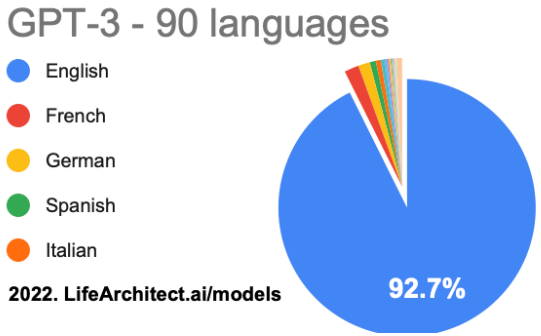




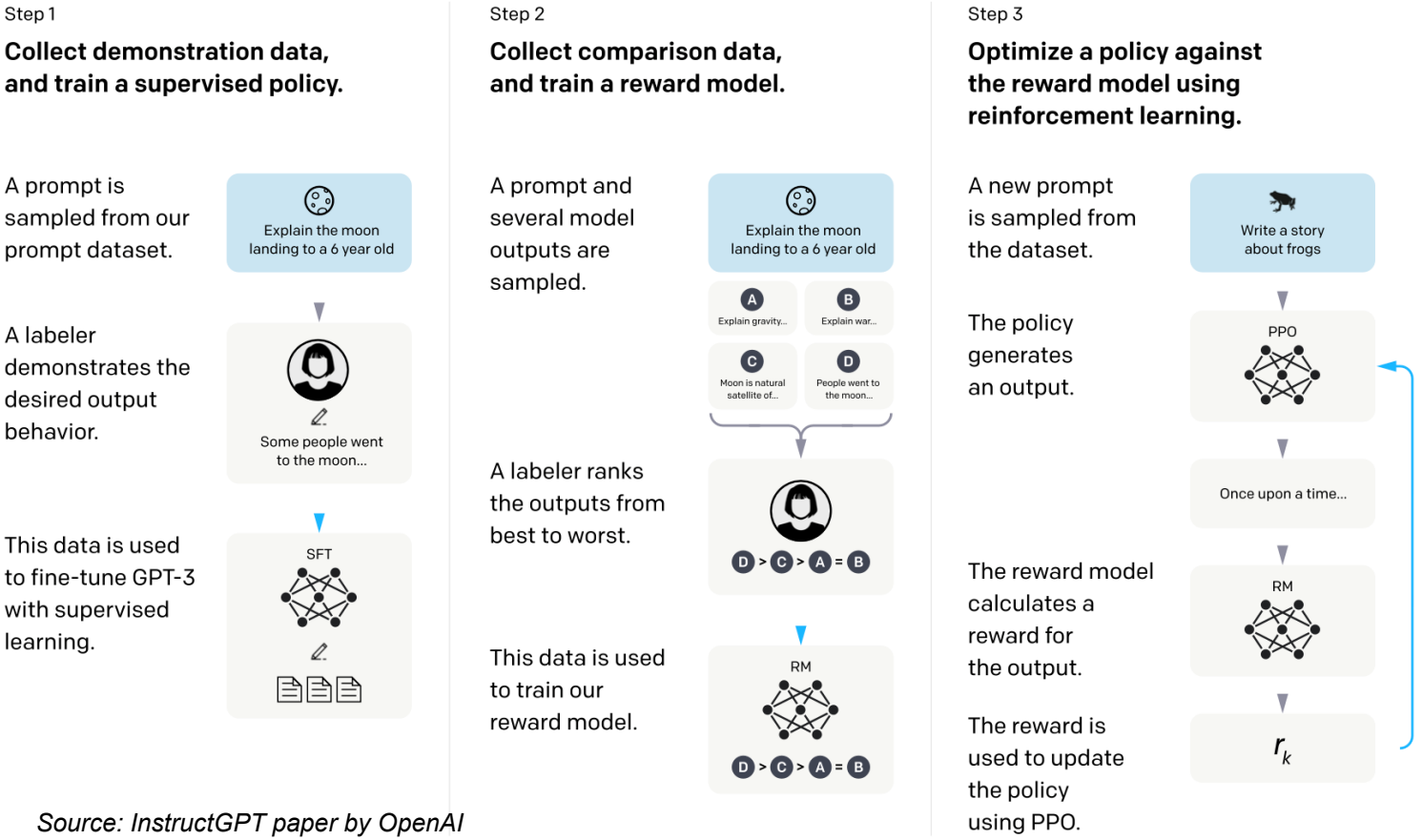
What is in the training data for LLMs



- The total size of the training dataset is estimated in **45TB of text**
- Being a data-driven model, there is an intrinsic bias, induced by the data used for the training



Training Large Language Models



Source: InstructGPT paper by OpenAI

- The estimated cost to train ChatGPT is **$4.6 million** with an estimated energy usage of **936 MWh**. This amount of energy is enough to power 30,632 American households for a day, or 97,396 average European households for a day

Bias in LLMs

- There are 3 different types of biases in a Large Language Model:
  - Data: The system is based on data. A biased dataset will cause biased results
  - Data Labeling: Humans are labeling the data, and their choices will influence the behavior of the system
  - Model: The algorithm used to encode the dataset is using a logic that may or may not be applicable to all the cases
- While this may not be a vital issue for many industries, it is critical for all the industries operating in restricted settings, such as Defense, Healthcare, Finance
- Besides a generic bias, there is the issue related to the low domain specialization of the generic Large Language Models, such as ChatGPT

Bias in LLMs - Remediations

- The first step when introducing an LLM into a production, the bias of the system should be evaluated. We are developing a method to measure at different stages generic and specific biases
- Once a measurement of the specific bias is determined, a specific remediation could be implemented. For example, by providing a proper dataset or performing a more controlled labeling
- One of the obvious options to reduce bias is to build your own LLM, assuming there is availability of the proper resources

Building your own LLM: Existing examples

Use Cases	Open-Source Models	Closed Models
Finance	**FinGPT** (Yang et al., 2023)	**BloombergGPT** (Wu et al., 2023)
Medicine	**BioMedLM** (Bolton et al., 2023), **BioGPT** (Luo et al., 2022), **GatorTron** (Yang, et al., 2023)	**BioGPT-JSL** (John Snow Labs, 2023)
Code	**CodeBERT** (Feng et al., 2020)	**Codex** (Chen et al., 2021)
Annotation		**AnnoLLM** (He et al., 2023)
Robotics	**PROGPROMPT** (Singh et al., 2023)	
Science	**Galatica** (Taylor et al., 2022) (shut down due to biased output)	

Building your own LLM: SSE_GPT project

- In June ‘23 I started a self funded summer project to develop a “SSE_GPT” prototype using data from one of my courses
- Final goal is to create an automatic tutor for SSE students
- The course I picked runs in multiple section each year, being one of the most popular at SSE, with about 150 students/year
- There is a relatively large amount of data on this course, coming from transcripts and teaching material. Transcripts cover about 10 full semester
- Another valuable source of data is the collection of emails from/to students with questions (from the students) and answers (from the instructor and/or the TA). This dataset has been used as a stage 1 in the finding answers
- The team is composed by 2 of my PhD students - Shiyu Yuan and Balaji Rao – 2 Masters students (Kunal Pradeep Gandhi and Naveen Mathews Renji) a Postgraduate (Amineh Zadbood) and myself
- There are also 3 high school students working on different aspects of the use of LLMs, that could be integrated down the road with the main project
- The prototype is working, but still with a lot of errors. We are continuing the developments, expecting a stable prototype by the end of year. Results will likely be far less sensational as in leading LLMs, due to the limited resource we have, compared to industry