FDP INSTITUTE by CAIA

Webinar

Leveraging Large Language Models in Finance

Welcome

We will begin promptly at 11 AM ET.

If you are unable to hear the speakers, please let us know in the chat box. You may enter your questions in the Q&A, we will address them at the end of the presentation. You can find a copy of the slide deck and recording of this webinar: www.fdpinstitute.org/webinars

FDP INSTITUTE®

Financial Data Professional Institute

FDP Institute provides world class training and education to financial professionals to meet the accelerating needs of digital transformation in the industry.

Introductions

Avi Patel, CAIA, FDP
Head of Marketing &
Data Science,
Fulton Bank

Raul Salles de Padua
Founder & Principal
Data Scientist,
Quod Analytics

Dr. Hossein Kazemi, CFA
Senior Advisor,
CAIA Association &
FDP Institute

Today’s Topic:

Leveraging Large Language Models (LLMs) in Finance

Leveraging Large Language Models (LLMs) in Finance

Avi Patel, FDP, CAIA

Raul Salles de Padua

Contact info & resources

Avi Patel:

Email: avipatel68@gmail.com

LinkedIn: https://www.linkedin.com/in/avipatel/

YouTube: https://www.youtube.com/@AviPatel68

Github: https://github.com/mktaop

Most recent research, 'Word Embeddings for Banking': https://arxiv.org/abs/2306.01807

Raul Salles de Padua:

Email: raul.padua@iese.net

LinkedIn: https://www.linkedin.com/in/raulpadua/

YouTube: https://www.youtube.com/@sallespadua

Github: https://github.com/raul-padua

HF Spaces: https://huggingface.co/raul-padua

Most recent research, 'GPT-3 Models Are Few Shot Financial Reasoners': https://arxiv.org/abs/2307.13617

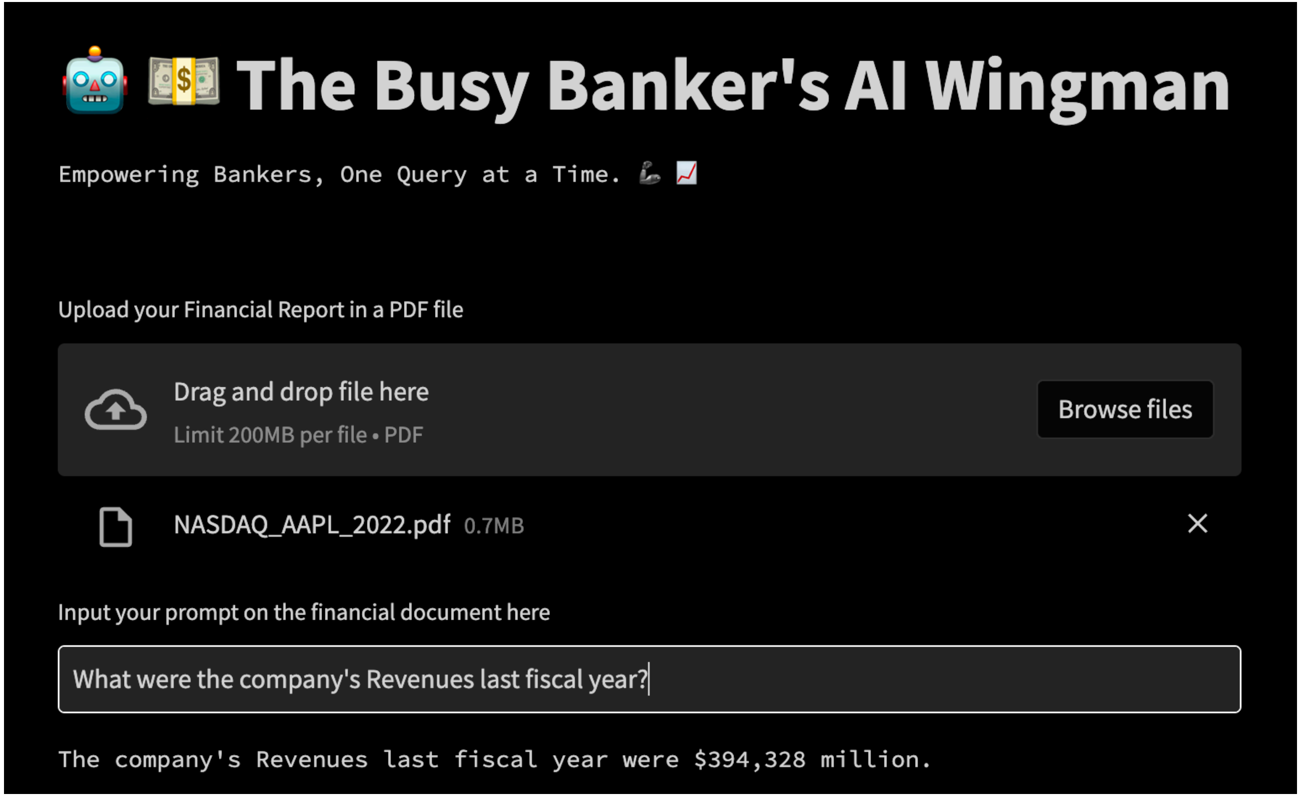
Agenda

1. Build up to Large Language Models (LLMs)
2. LLMs use cases and lifecycle
3. Prompt Engineering & Configuration Parameters
4. MLOps in GenAI Prompt Engineered Systems

Agenda

1. Build up to Large Language Models (LLMs)
2. High-level overview in the Transformers architecture
3. LLMs use cases and lifecycle
4. Prompt Engineering & Configuration Parameters

Real world application with financial documents



https://www.youtube.com/watch?v=Q5wAmPT7lxs

Build up to LLMs- Vector Space Models

From static embeddings to contextual embeddings …on large corpora…but fail to draw out the domain-specific relationship/semantics

Couple of results from word embeddings experimentation from my paper

Table 4: Evaluation of Models

Model	Dimensions	Spearman Correlation
Dervied Embeddings		
LSA[1]	100	0.44
LSA	500	0.49
LSA + Autoencoder	300	0.51
GloVe	100	0.32
GloVe	300	0.42
Word2Vec	300	0.33
LSTM	300	0.26
Embeddings Lookup		
GloVe	300	0.41
Word2Vec	300	0.30
BERT	768	0.27
ADA-002	1536	0.34

[1] Baseline model/benchmark

Table 8: Clustering results

Word-1	Word-2	Model 1	Model 2	Model 3	Model 4
home	equity	1	0	0	0
credit	report	1	1	0	0
credit	card	0	0	0	1
identity	theft	1	1	0	0
fair	credit	1	0	0	0
reporting	act	1	1	0	1
account	number	1	1	0	0
wells	fargo	1	0	0	0
credit	score	0	1	0	0
bank	america	1	0	0	0
inaccurate	information	1	0	0	0
credit	bureau	1	1	0	0
collection	agency	1	0	1	0
third	party	1	1	0	0
checking	account	0	1	1	0
please	remove	1	0	0	0
late	payments	0	0	0	0
debt	collection	1	0	0	0
bank	account	0	0	0	1
late	payment	0	0	0	0
interest	rate	1	1	0	1
payment	history	0	1	0	0
late	fees	0	0	0	0
home	loan	1	0	0	0
bank	branch	1	1	0	0
Accuracy (%)		68%	44%	8%	16%

Table:

Model 1 LSA+Autoencoder

Model 2 BERT

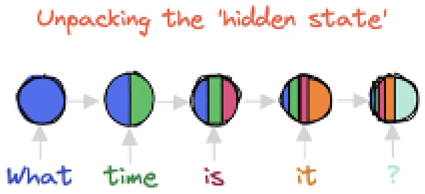
Model 3 ADA

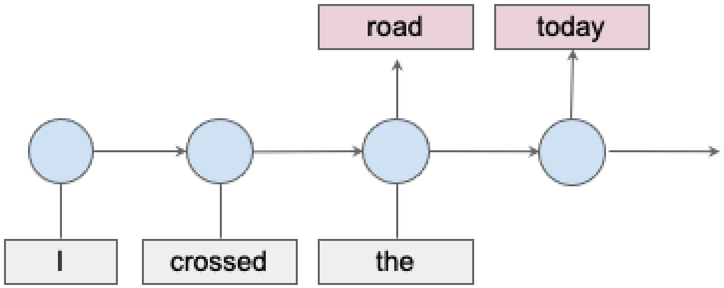
Model 4 GloVe 300d (pretrained word vectors)

Build up to LLMs- RNN Architecture

For many years, language processing was done with recurrent neural networks. Predicting each token is based on the most recent word plus the hidden state of the past. Two key problems with this:

- Not parallelizable
- Not good at long range word associations since they weigh recent words more highly





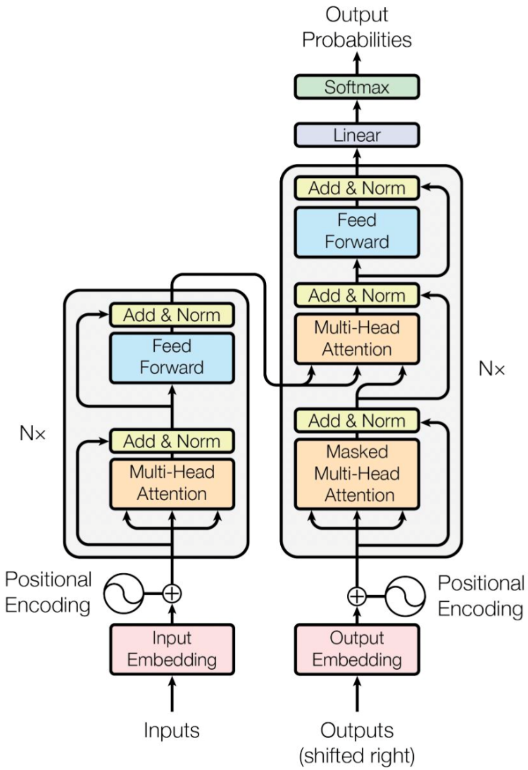
Simplified LSTM Architecture

[Source: Foundation models]

Build up to LLMs - Transformers: Attention is all you need

Encoder

Encodes inputs ("prompts") with sequences contextual understanding
E.g. Bert: MLM, PNS



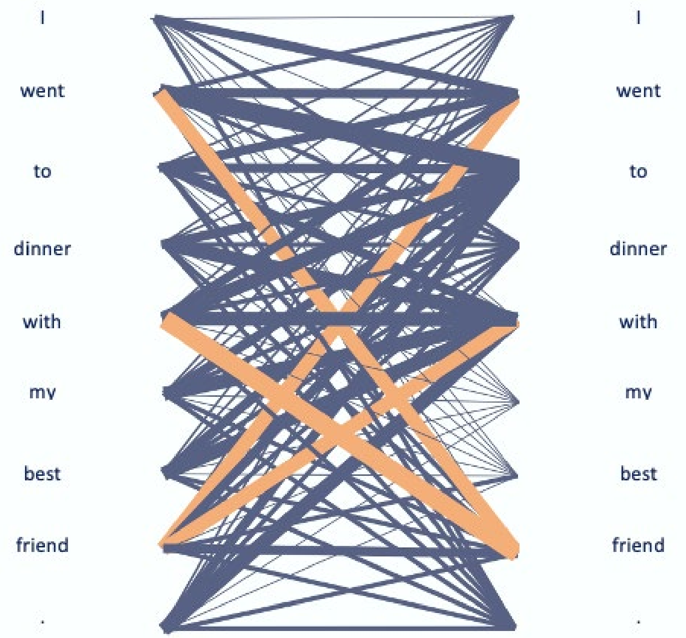
Decoder

Takes in new tokens and generates new tokens
E.g. GPT-3: Predict next token, recursive

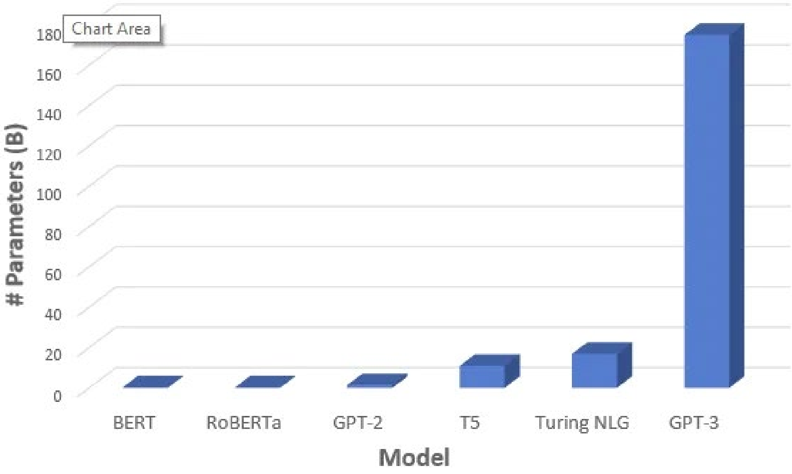
Build up to LLMs- Transformers: Attention is all you need..Learn context for each combination of words.

Typical RNN

Transformers

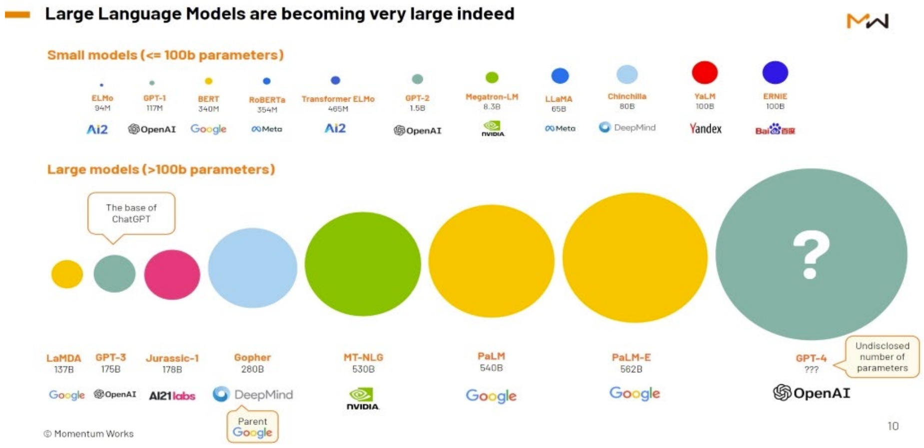


Build up to LLMs- Size Matters



GPT-3 by far was the largest model, in 2020

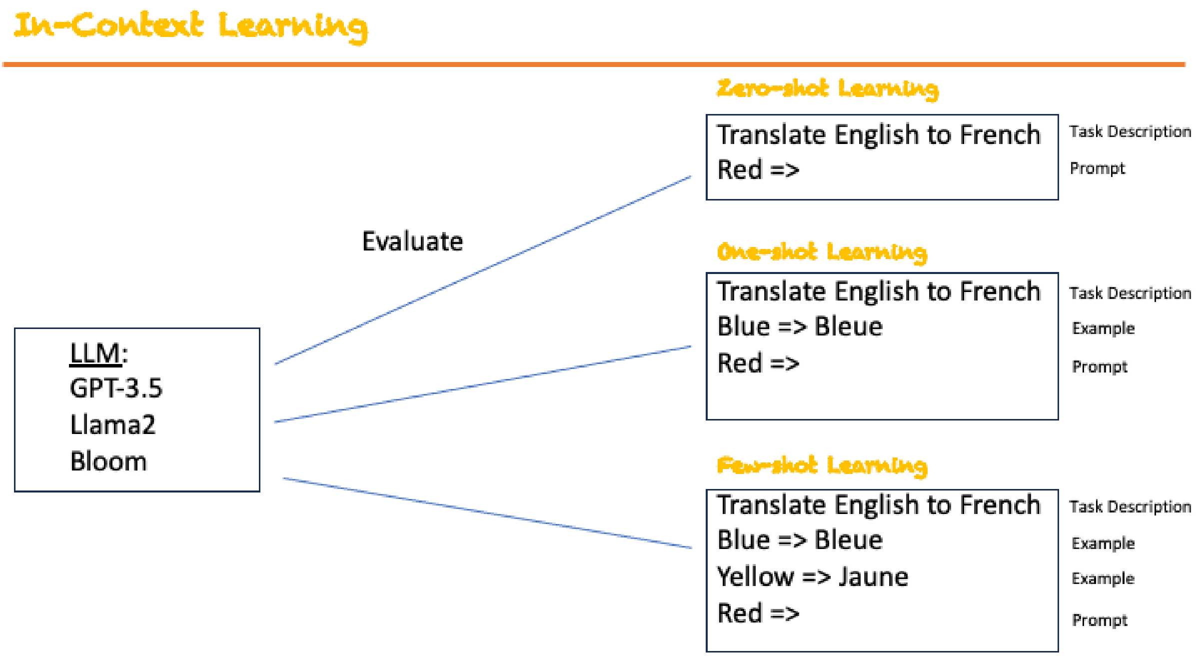
Source: https://towardsdatascience.com/gpt-3-the-new-mighty-language-model-from-openai-a74ff35346fc



Since then, just an explosion in scaling

Source: https://thelowdown.momentum.asia/the-emergence-of-large-language-models-llms/

That's how we got to LLMs, scale and architecture. An emergent property of these LLMs is "in-context" learning, which was not expected. They can complete tasks that they were not explicitly trained on!



Allows us to skip the time and $ required for fine-tuning. *Using the same LLM, we can perform multiple tasks.*

Agenda

1. Build up to Large Language Models (LLMs)
2. LLMs use cases and lifecycle
3. Prompt Engineering & Configuration Parameters
4. MLOps in GenAI Prompt Engineered Systems

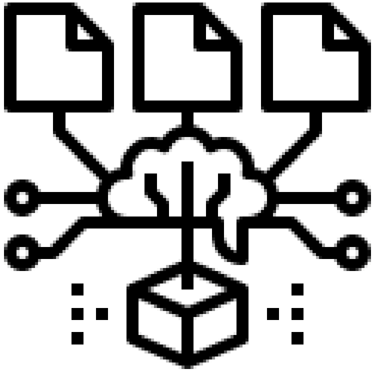
Leverage LLMs to perform tasks using your domain knowledge

Text Summarization

Question Answering

Machine Translation

EDA - Model Build w/ Prompts

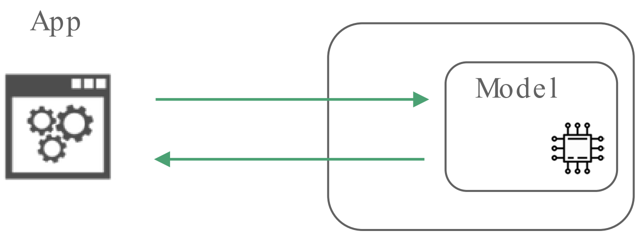


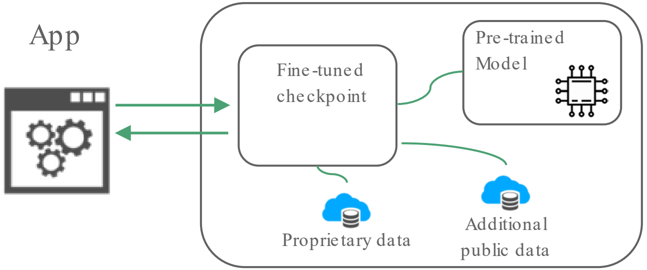
Disrupting for better, also into AI development?

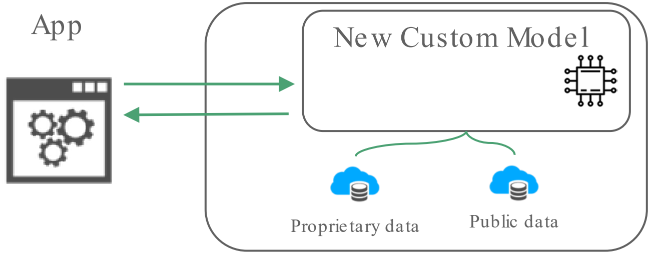
Supervised Learning	~1-2 months	Get labeled data
	3 months	Train a model on that data
	~2-3 months	Deploy & Call model in apps
Prompt Engineering	~1-2 days	Define prompt
	< 1 day	Call model

There are mainly three ways to leverage adoption of LLMs' Gen AI capabilities…

Approach	How to…	Effort/ Cost
1. Prompt Engineering	• Commercial/ Open Source API calls • Instruct LLM through prompts to pass role and context • In-Context prompts: zero to few-shot	Low
2. Fine-tuning	• Instruction-led fine-tuning of pre-trained LLM with additional examples related to the task	Costly
3. Build your own	• Create and Train a new model from scratch	Very Costly







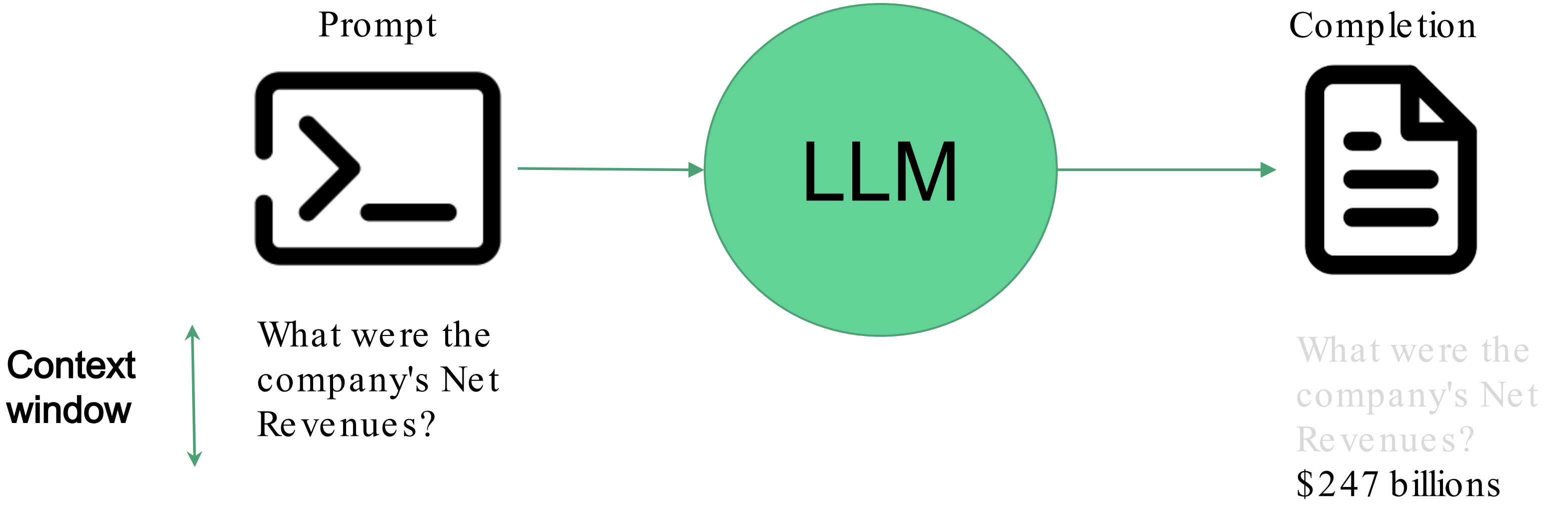
… as they are setting stones within LLMOps lifecycle

Scope & Constraints	Model Selection	Tailor your model	Deploy into apps integration
• Define application & Use case	• Choose model ○ **pretrained** ○ build your own pretrained model	• **Prompt Engineering** • **Fine-tuning** • Align with human feedback	• Deploy for inference • Build LLM application • Monitor

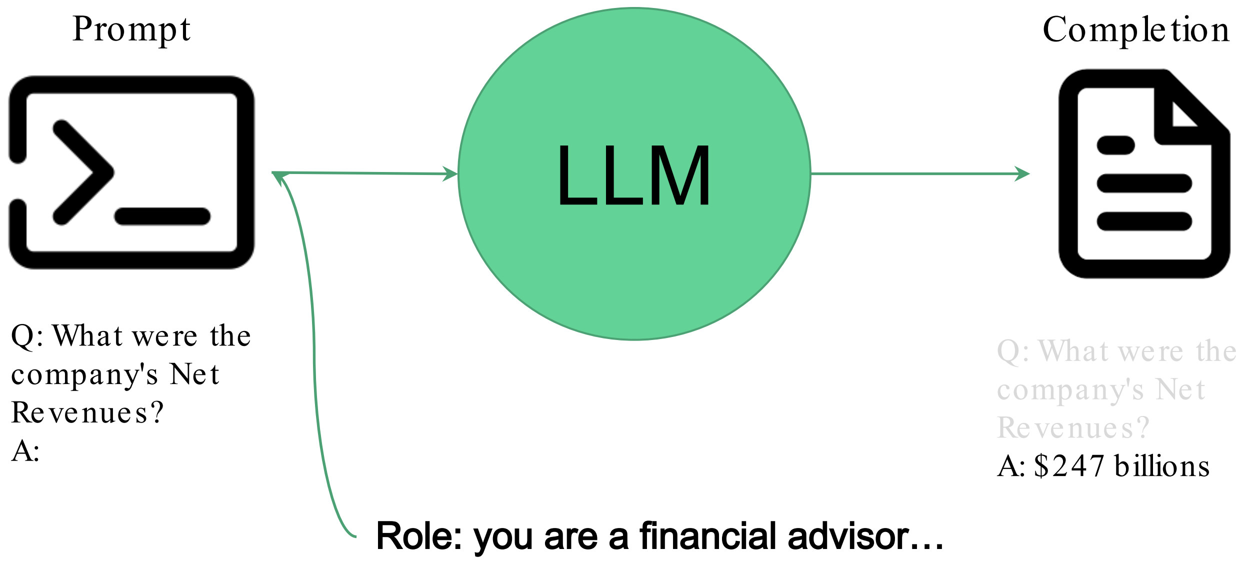
Agenda

1. Build up to Large Language Models (LLMs)
2. LLMs use cases and lifecycle
3. Prompt Engineering & LLM's instruction led finetuning
4. MLOps in GenAI Prompt Engineered Systems

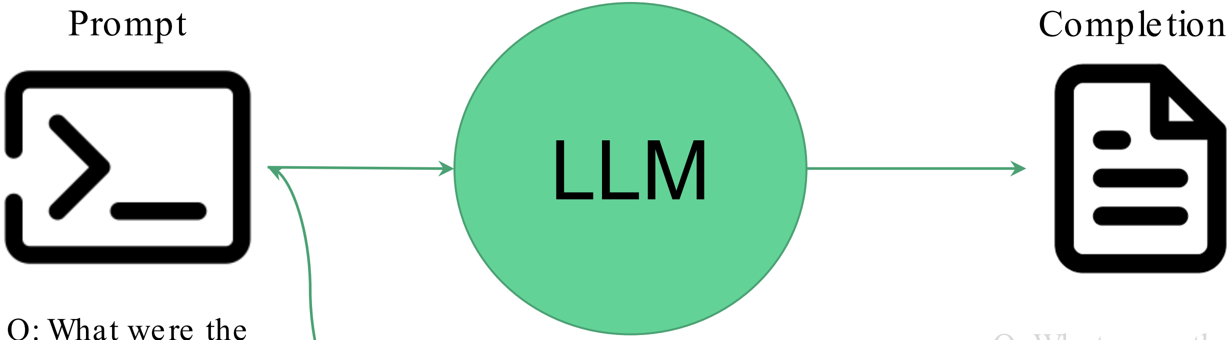
What is prompt engineering?



Zero-shot prompt engineering



One-shot prompt engineering



Q: What were the company's Net Revenues?
A: $247 billions

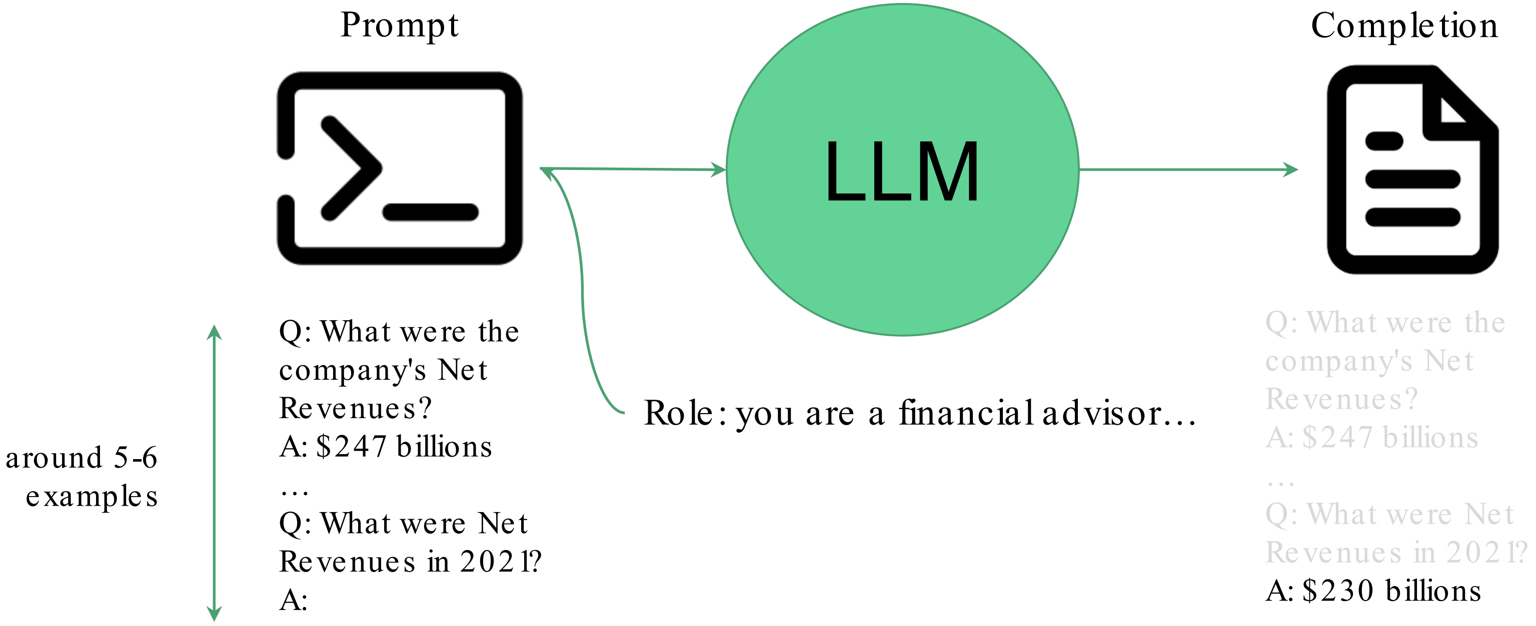
Q: What were the company's Net Profits?
A:

Role: you are a financial advisor…

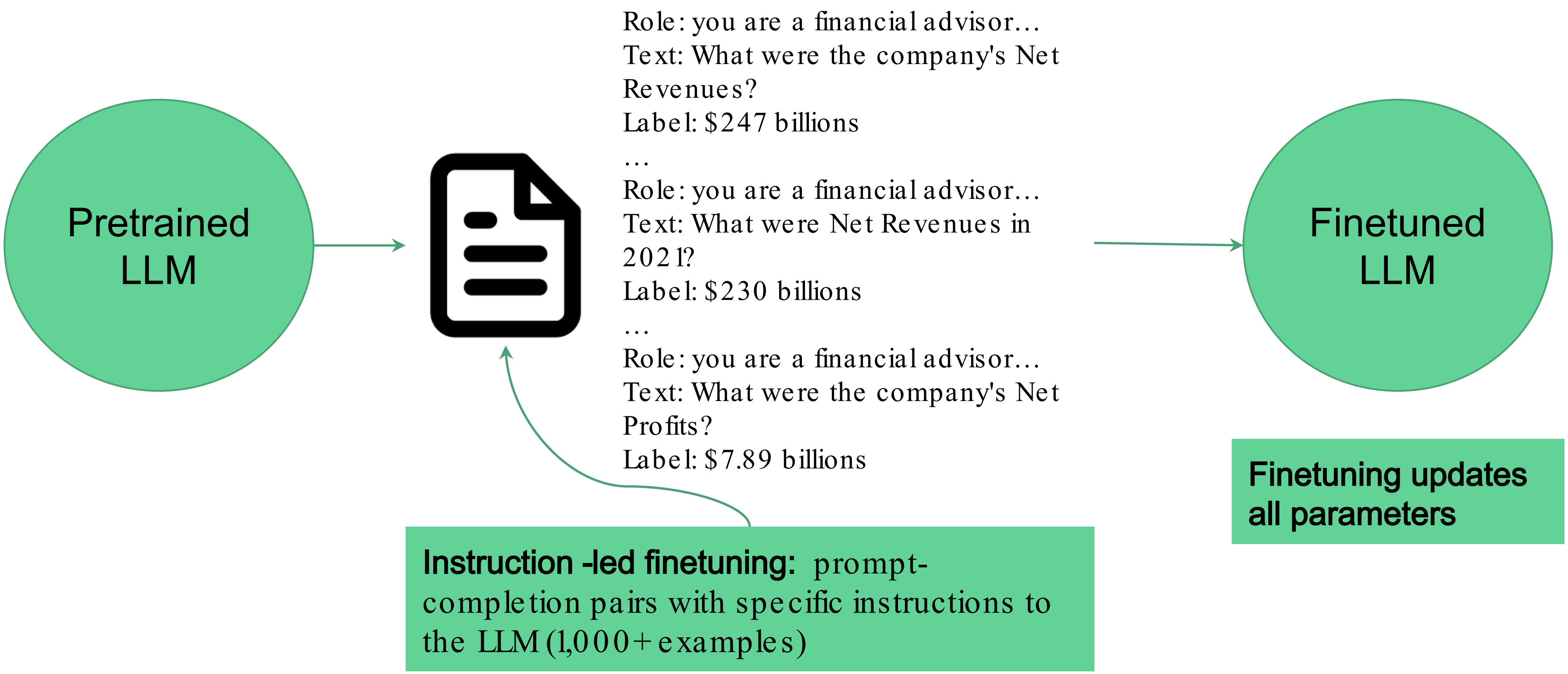
Q: What were the company's Net Revenues?
A: $247 billions

Q: What were the company's Net Profits?
A: $7.89 billions

Few-shot prompt engineering



Fine-tuning a pretrained model with instruction



Sample prompts in instruction-led fine-tuning

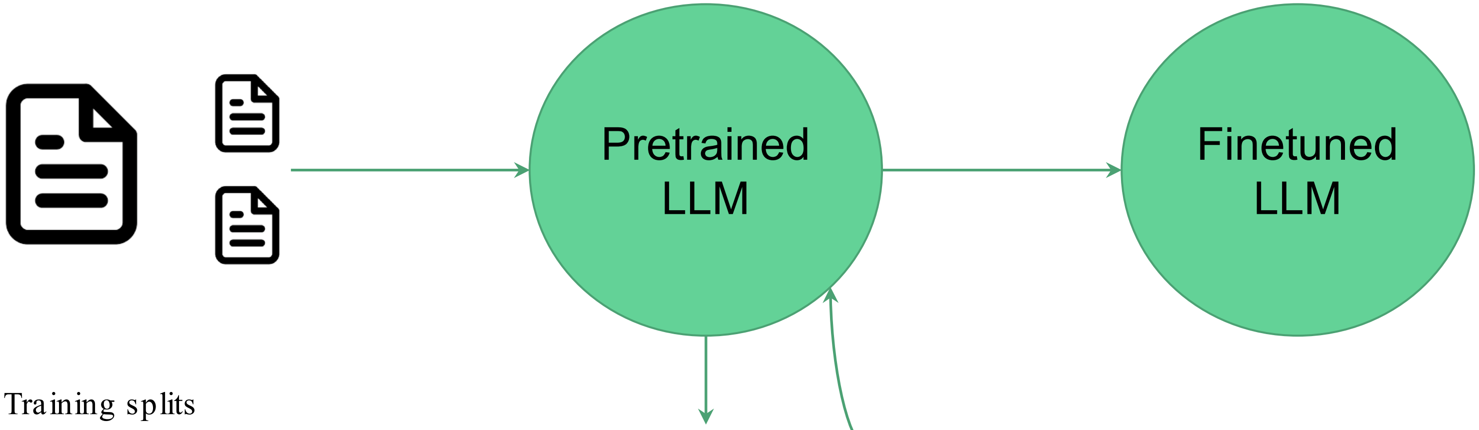
Classification / Sentiment Analysis

"Given the following news article:\n{{news_article}}\npredict the associated sentiment (positive, neutral, and negative)"

Agent: Question Answering

"""You are an excellent document analyst specialized in the financial markets and industry.\
You are great at performing the task queried in the prompt in a straightforward and easy to understand manner.\
You answer the prompts always in English, translating from the document if it's not written in English language.\
If the query is out of scope from the document answer back with {{out_scope}}.\n
Here is the prompted query:\n {{prompt}}"""

LLMs' finetuning process



Training splits

Training
Development
Test

Prompt: you are a financial advisor…
Text: What were the company's Net Revenues?
Prediction: $247 billions

Label: $247 billions

Loss

How to bring this to life in your organization?

- Sell the art of the possible to generate use-case(s) from the business line folks
  - Start with an use-case with lowest amount of risk to the firm, think POC
- Engage key stakeholders for buy-in, definitely:
  - Legal, Compliance, Information Security, Technology (IT), Risk Management
- Highlight the challenges that do exist, such as:
  - Latency, Cost, Context Window, Building Complex Chains, Model & Alignment Choice

Agenda

1. Build up to Large Language Models (LLMs)
2. LLMs use cases and lifecycle
3. Prompt Engineering & LLM's instruction led finetuning
4. High-level MLOps in GenAI Prompt Engineered Systems

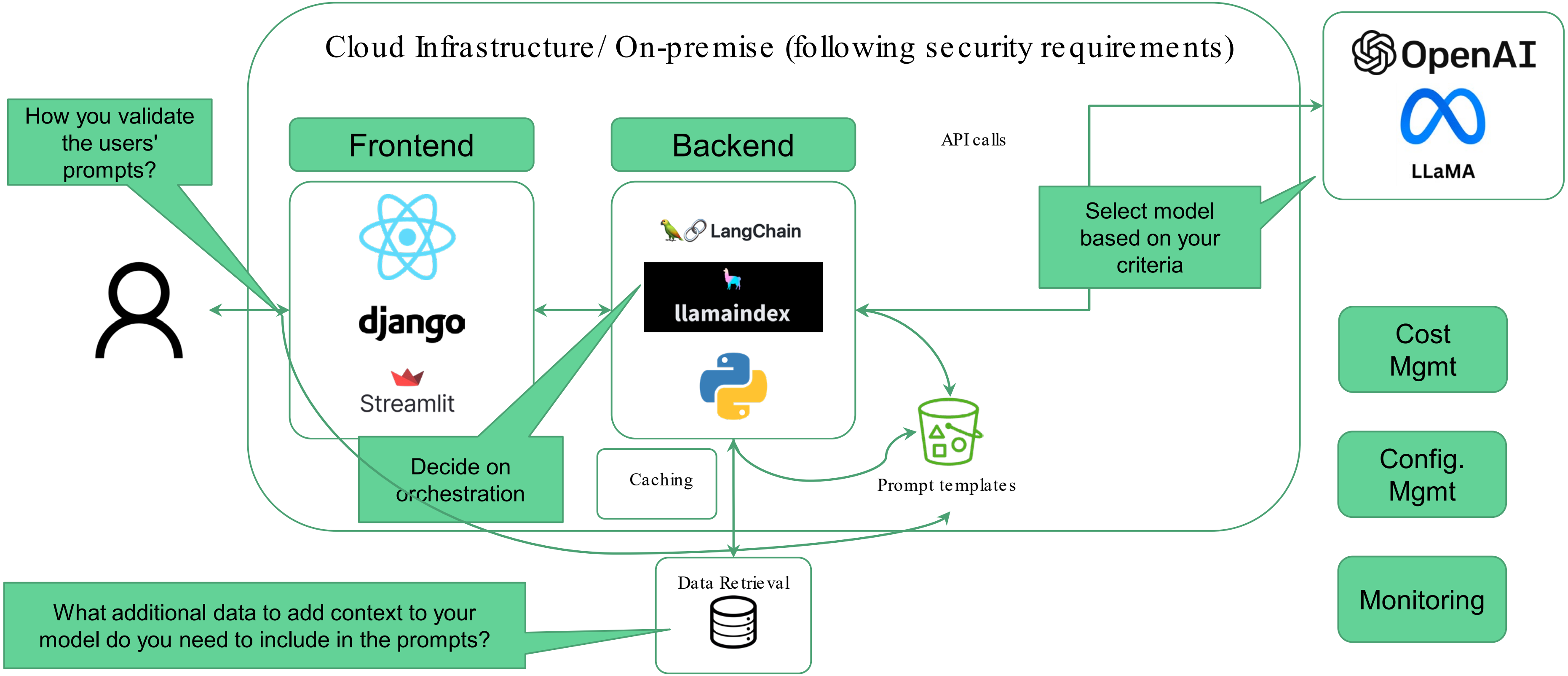
"...start with the customer experience and work backwards with the technology".

Steve Jobs

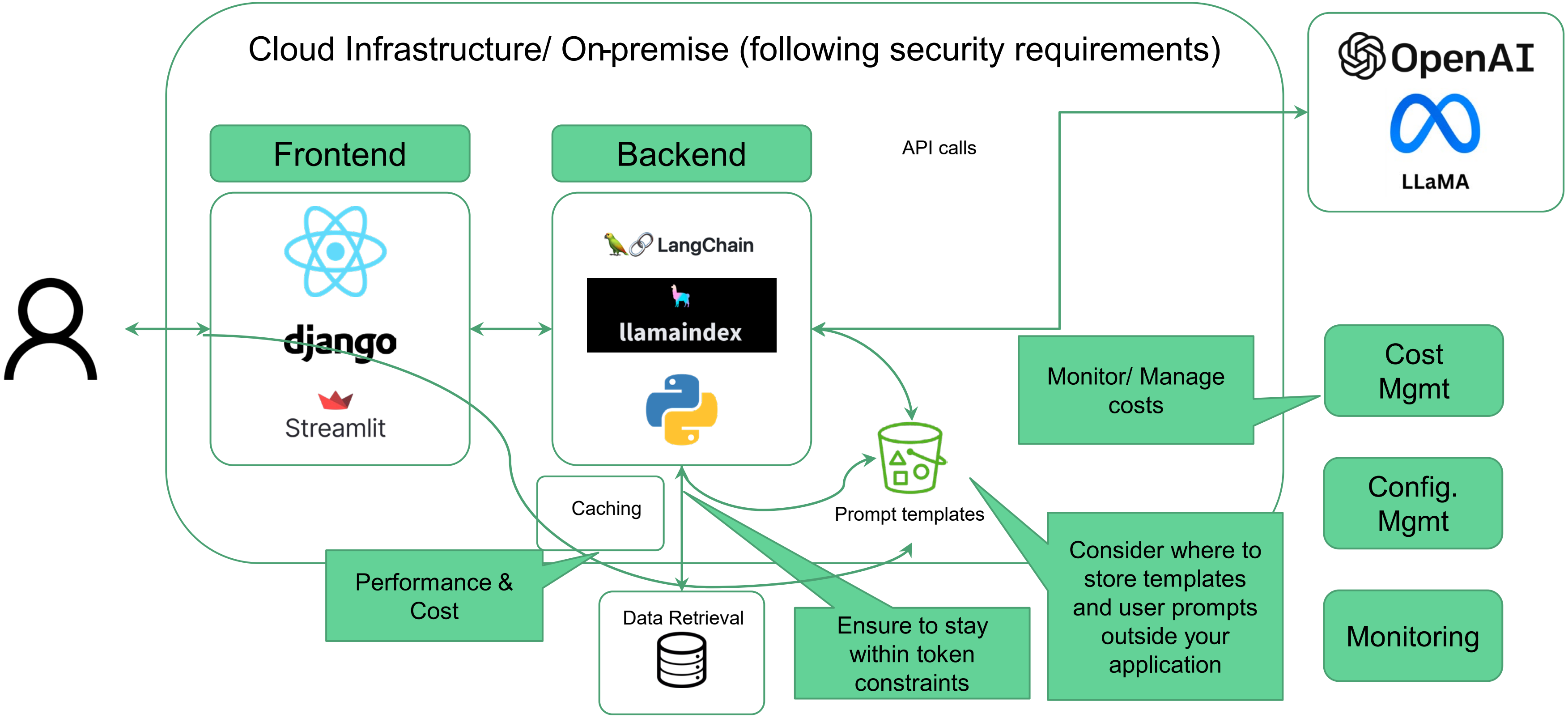
… as they are setting stones within LLMOps lifecycle

Scope & Constraints	Model Selection	Tailor your model	Deploy into apps integration
• Define application & Use case	• Choose model ○ pretrained ○ build your own pretrained model	• Prompt Engineering • Fine-tuning • Align with human feedback	• **Deploy for inference** • **Build LLM application** • **Monitor**

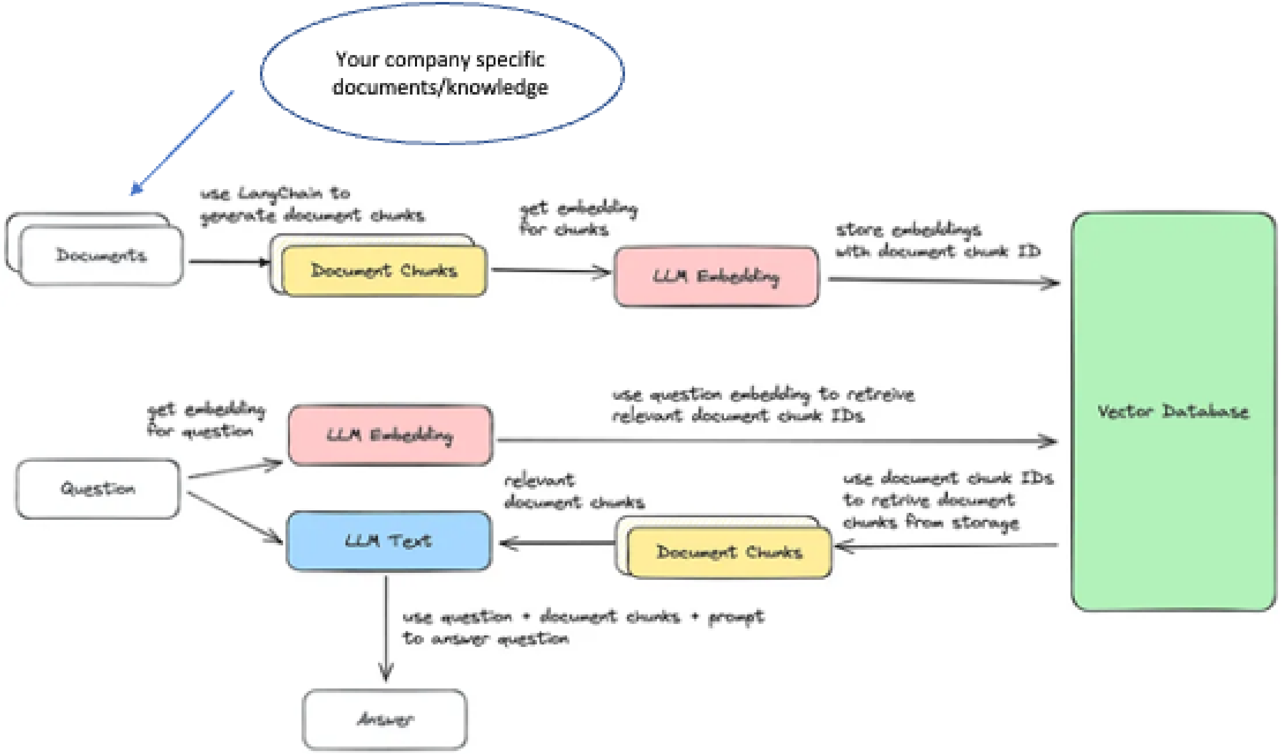
Prompt Engineered system example



Prompt Engineered system example



Q&A Example- Using domain-specific knowledge



Source: Sascha Heyer

Key takeaways

Unmeasurable "low hanging fruits" prompt engineering LLMs

You may fine-tune LLMs to specialize on your task, but keep in mind of the trade-offs

Same MLOps principles apply when LLMs serve applications in production

Questions

Q & A

Please join us for our upcoming webinar:

Register Here: https://bit.ly/3YsZIVi

Additional useful resources

Original GPT-3 Paper: Language models are few-shot learners

Attention is all you need – Transformer architecture

Interpreting Pretrained Contextualized Representations via Reductions to Static Embeddings

Word-to-Vector Paper

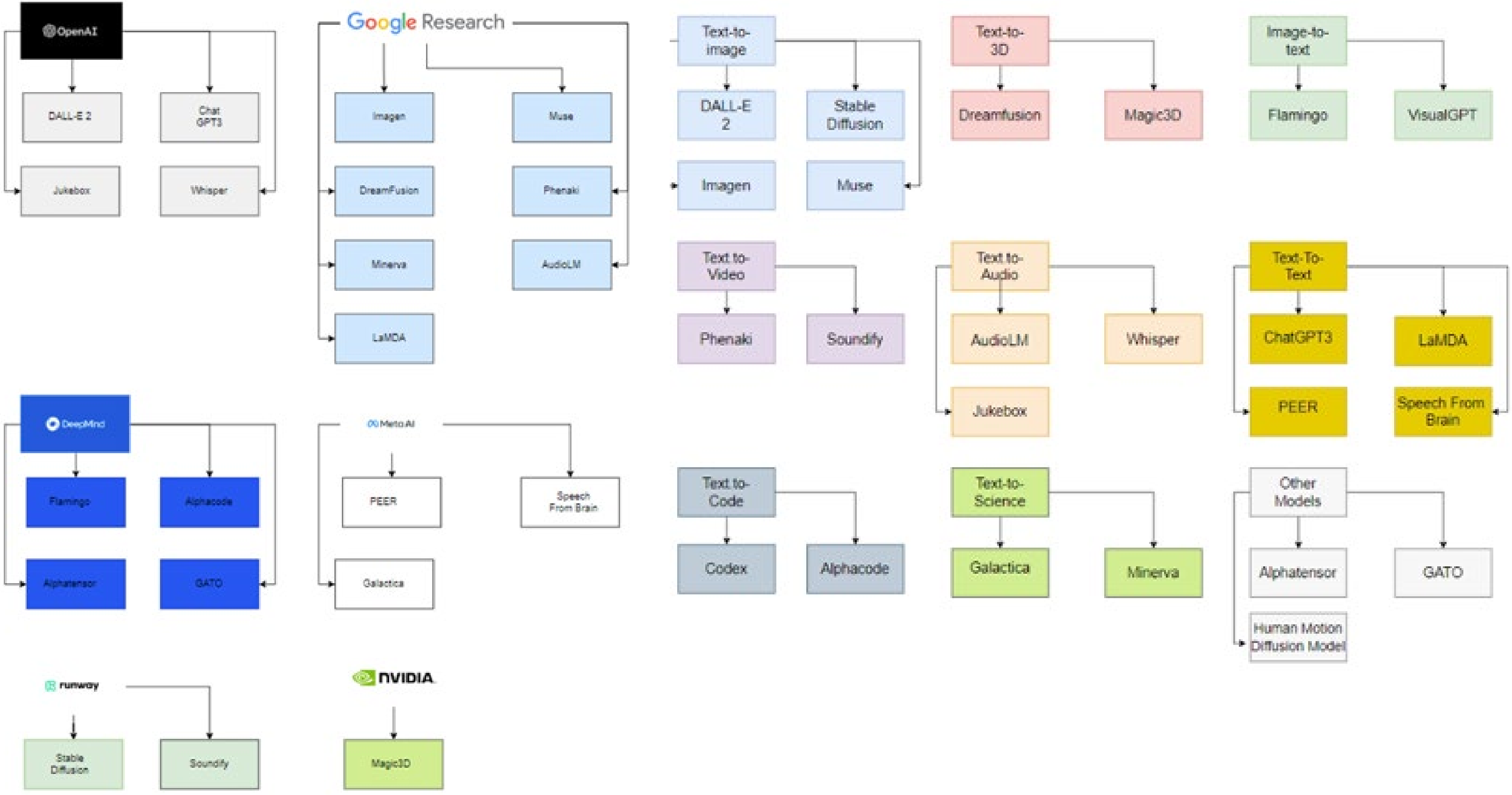
ChatGPT is not all you need. A State of the Art Review of large Generative AI models.

On the Opportunities and Risks of Foundation Models

Holistic Evaluation of Language Models

Text to powerpoint, yes you can

State of the art GenAI Models- Landscape from Jan 2023



Source: https://arxiv.org/abs/2301.04655

Thank You

Contact Us:

fdpinstitute.org

info@fdpinstitute.org

@FDPbyCAIA

linkedin.com/company/FDP Institute