CPN Junior GenAI Knowledge Sharing

Feng Lu

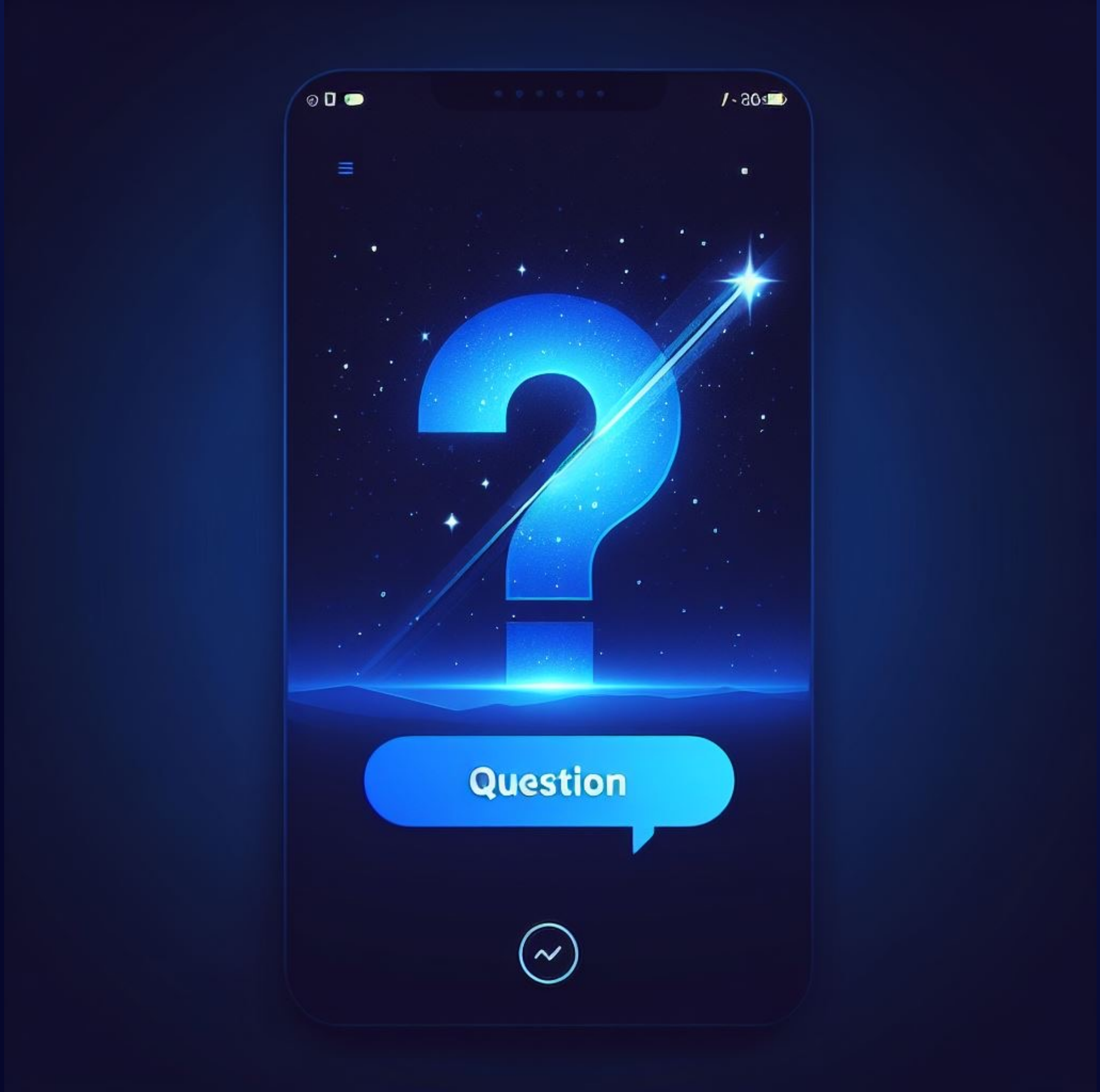
Developer, Architect and Technical Evangelist

Oslo Area, Norway

https://feng.lu/

Generative AI Project Lifecycle
aws
sopra steria
Gå sammen. Nå lenger.
Augment model and build LLM powered applications

AWS GenAI public event in Oslo, Nov 2023

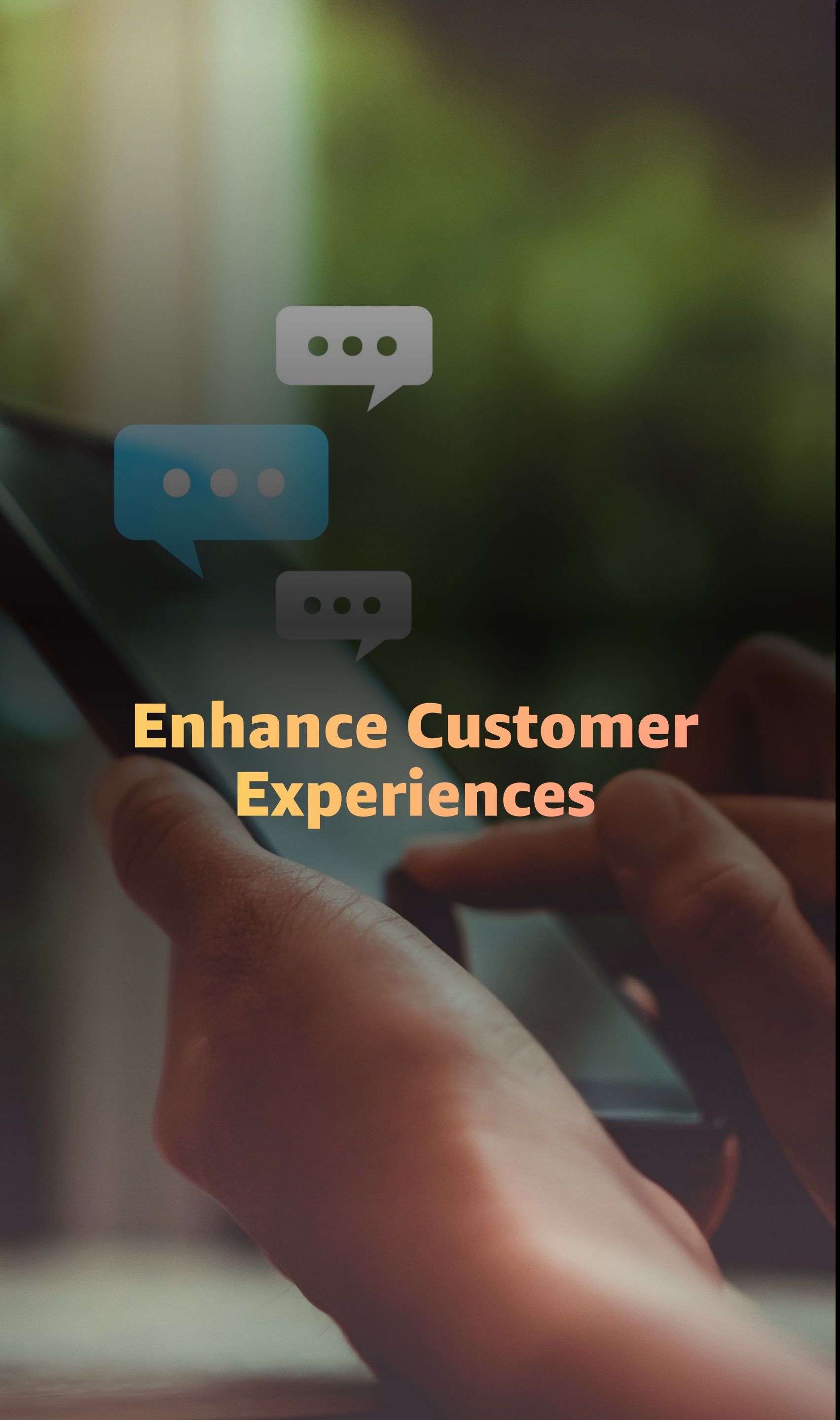

Question

Agenda

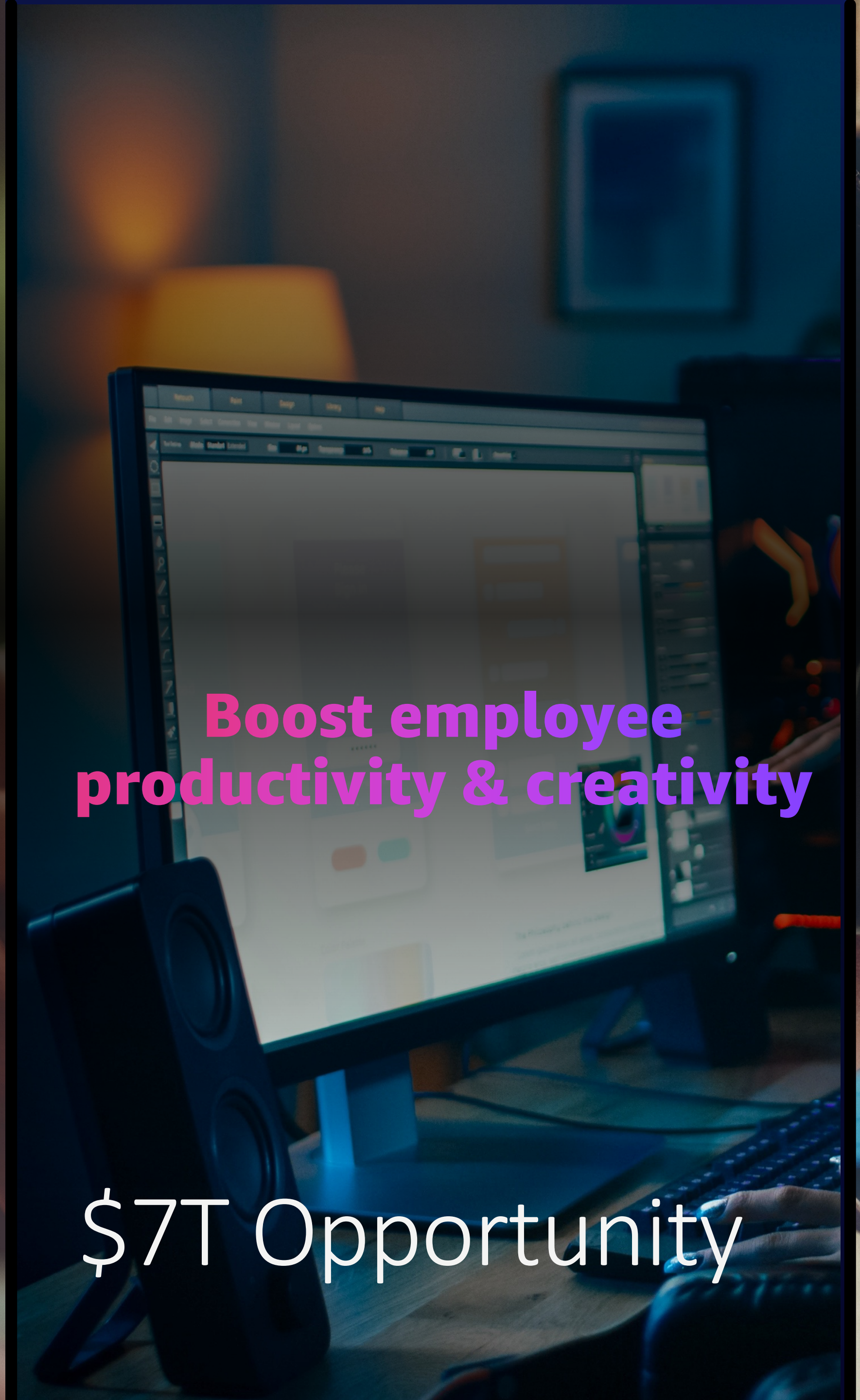
- Motivation ⬅
- Basic concepts
- Use case and patterns
- Tools and resources

Innovation can **transform industries**

GENERATIVE AI



Enhance Customer Experiences

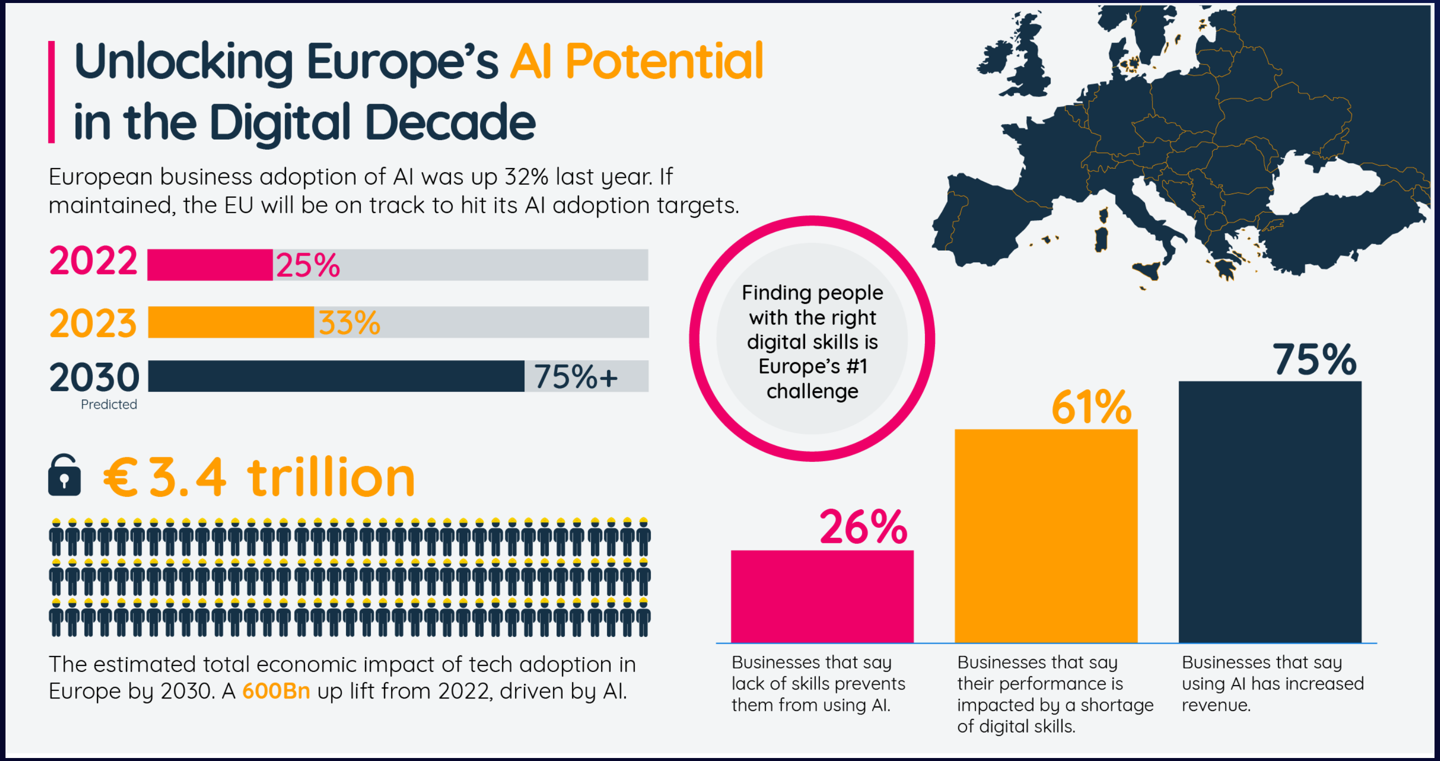


Boost employee productivity & creativity

$7T Opportunity

Optimize business processes

Motivation: Unlocking Europe’s AI Potential in the Digital Decade - report



Unlocking Europe’s AI Potential in the Digital Decade

European business adoption of AI was up 32% last year. If maintained, the EU will be on track to hit its AI adoption targets.

2022
25%
2023
33%
2030
Predicted
75%+

Finding people with the right digital skills is Europe’s #1 challenge

€ 3.4 trillion

The estimated total economic impact of tech adoption in Europe by 2030. A **600Bn** up lift from 2022, driven by AI.

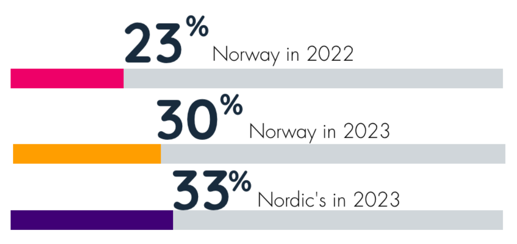
26%
61%
75%

Businesses that say lack of skills prevents them from using AI.

Businesses that say their performance is impacted by a shortage of digital skills.

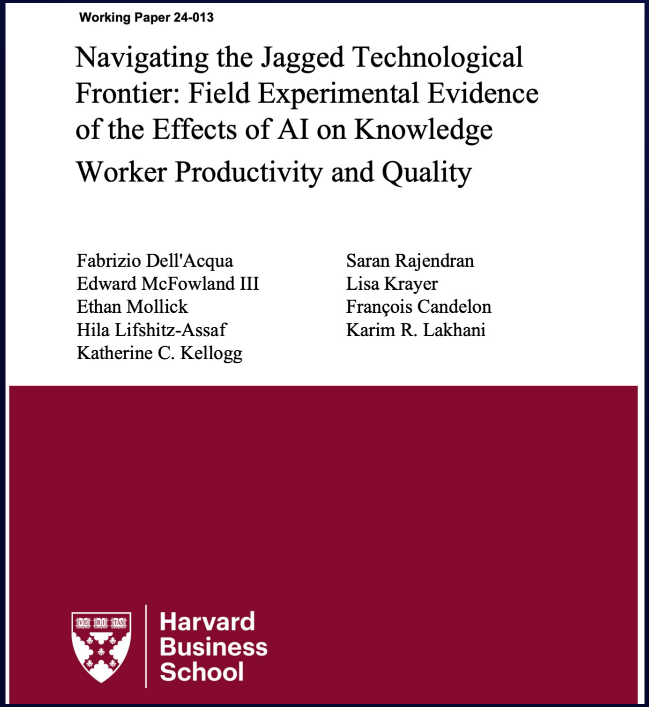
Businesses that say using AI has increased revenue.

Unlock 748 billion NOK for the Norwegian economy by 2030.



30% of Norwegian businesses were using at least one AI in 2023

Opportunity: Effects of AI on Knowledge Worker Productivity and Quality

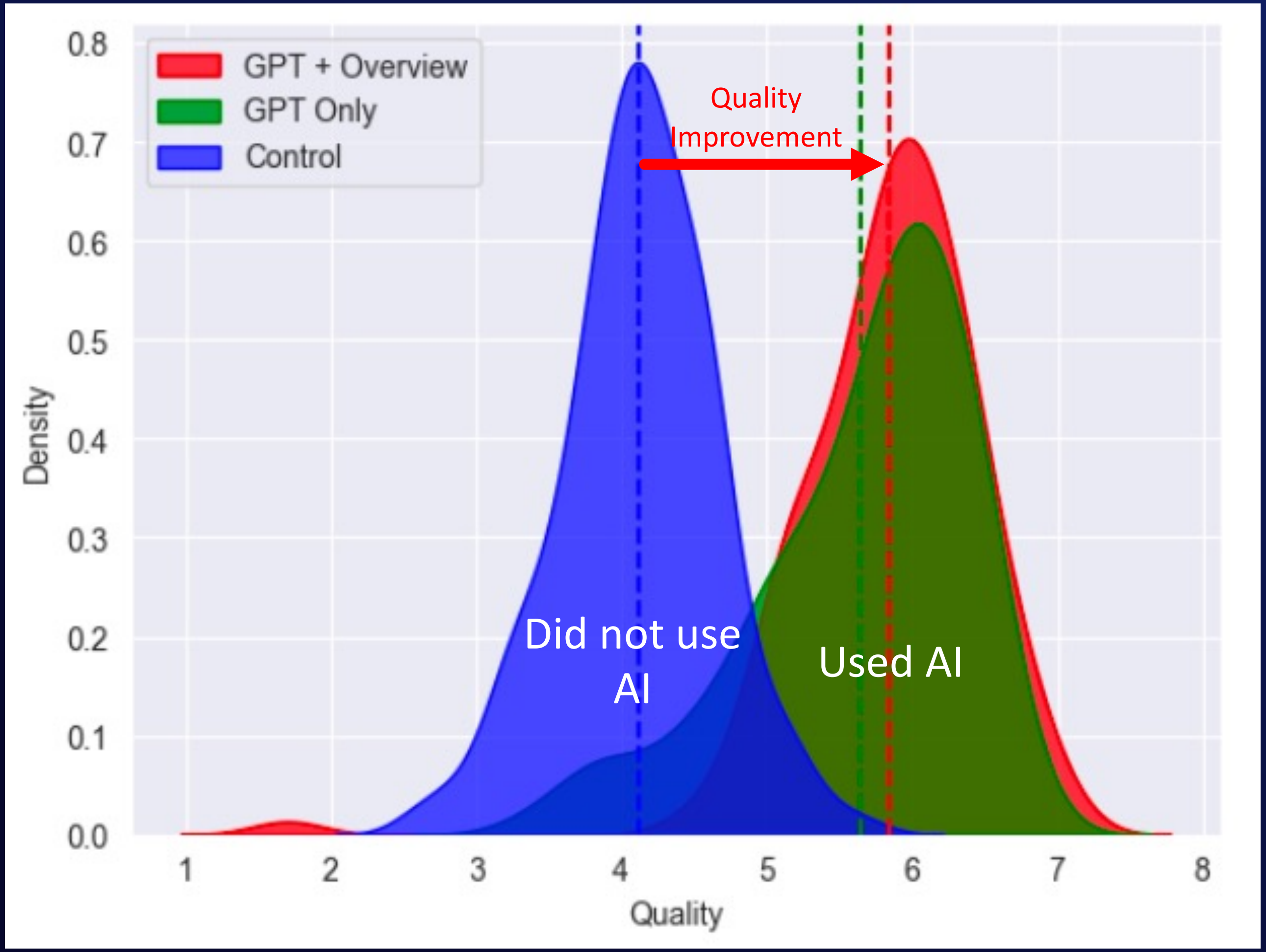


Working Paper 24-013

Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality

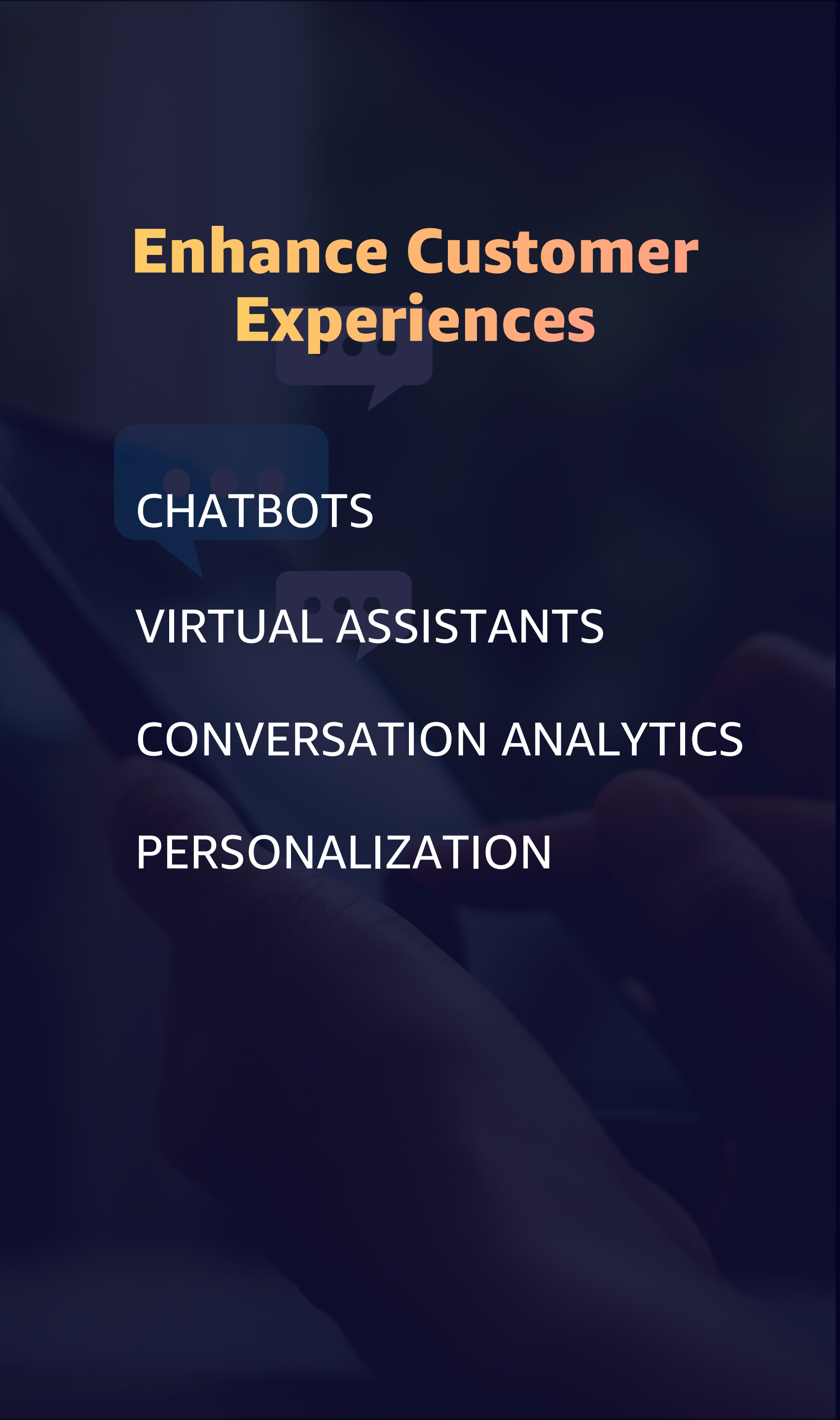
Fabrizio Dell'Acqua
Edward McFowland III
Ethan Mollick
Hila Lifshitz-Assaf
Katherine C. Kellogg
Saran Rajendran
Lisa Krayer
François Candelon
Karim R. Lakhani

Harvard Business School



Source: Dell´Acqua - Navigating the Jagged Technological Frontier

Quality improvement from using AI across 18 tasks:
12,5% more work, 26% faster, 40% higher quality



Enhance Customer Experiences

CHATBOTS

VIRTUAL ASSISTANTS

CONVERSATION ANALYTICS

PERSONALIZATION

Boost employee productivity & creativity

CONVERSATIONAL SEARCH

SUMMARIZATION

CONTENT CREATION

CODE GENERATION

DATA TO INSIGHTS

Optimize business processes

DOCUMENT PROCESSING

DATA AUGMENTATION

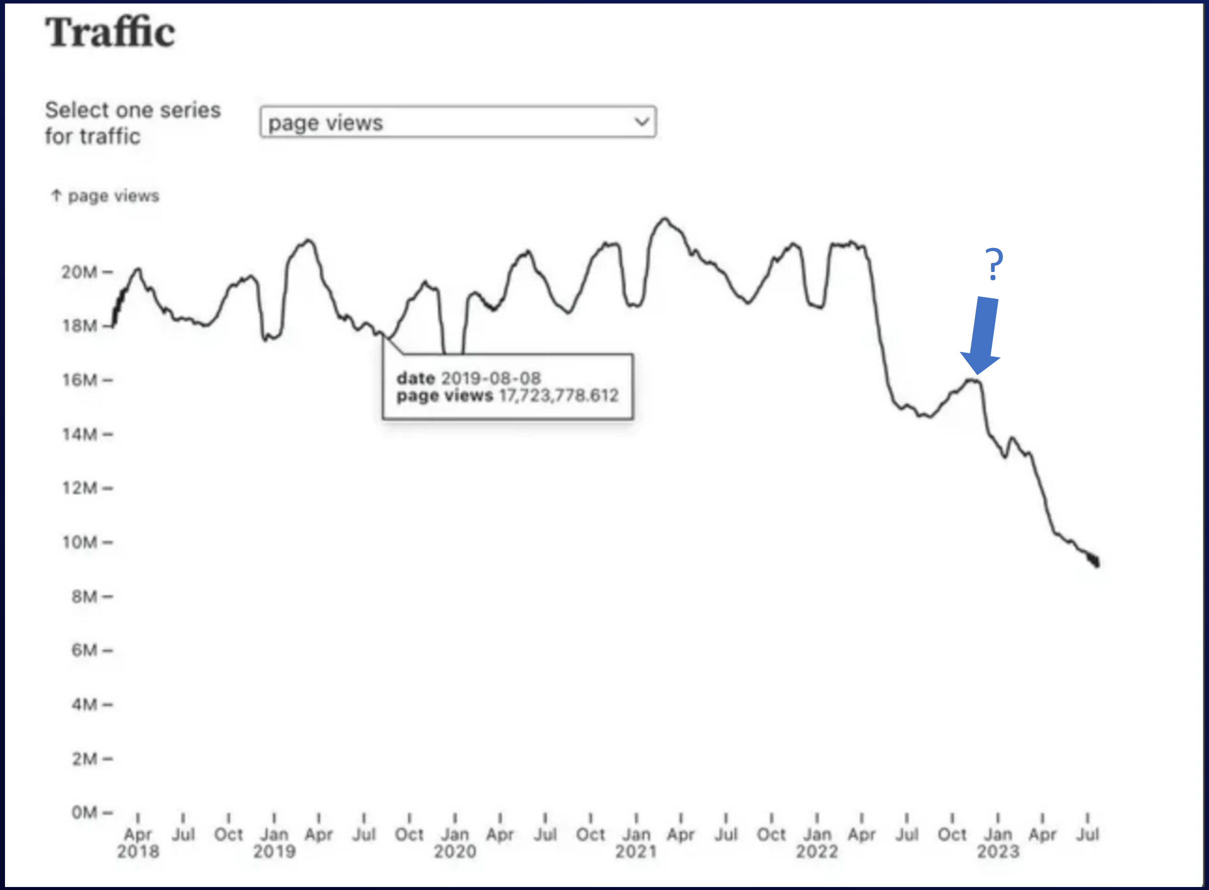
CYBERSECURITY

PROCESS OPTIMIZATION

Developers: new way of working

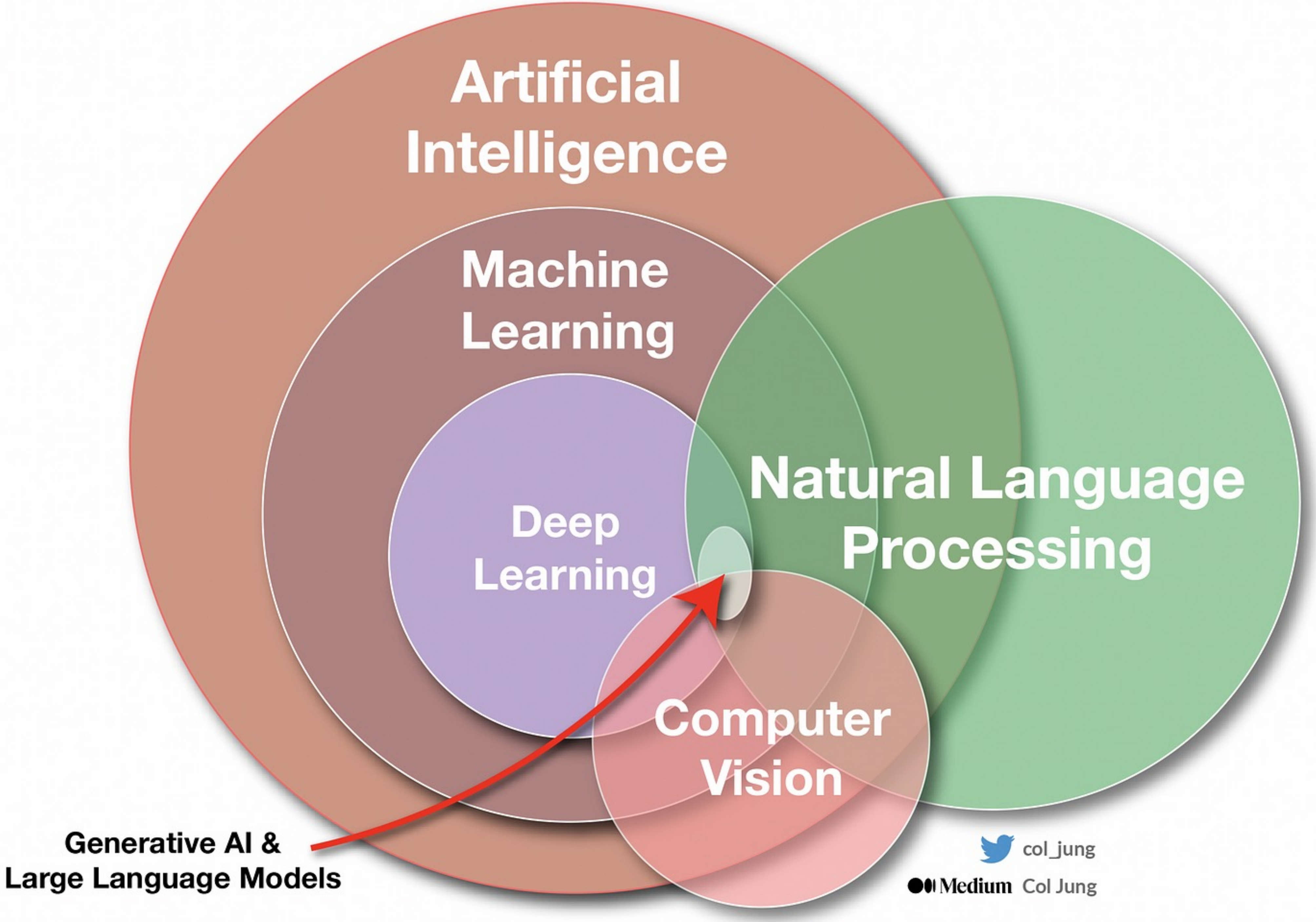
The way we interact with systems, software, search engines and tools are changing

- Developers relied on StackOverflow since 2008.
- Spend time finding code snippet to do a task or solve a bug
- If no solution – you post a question and wait for the community to answer.



Agenda

- Motivation
- Basic concepts ⬅
- Use case and patterns
- Tools and resources



Artificial Intelligence

Machine Learning

Deep Learning

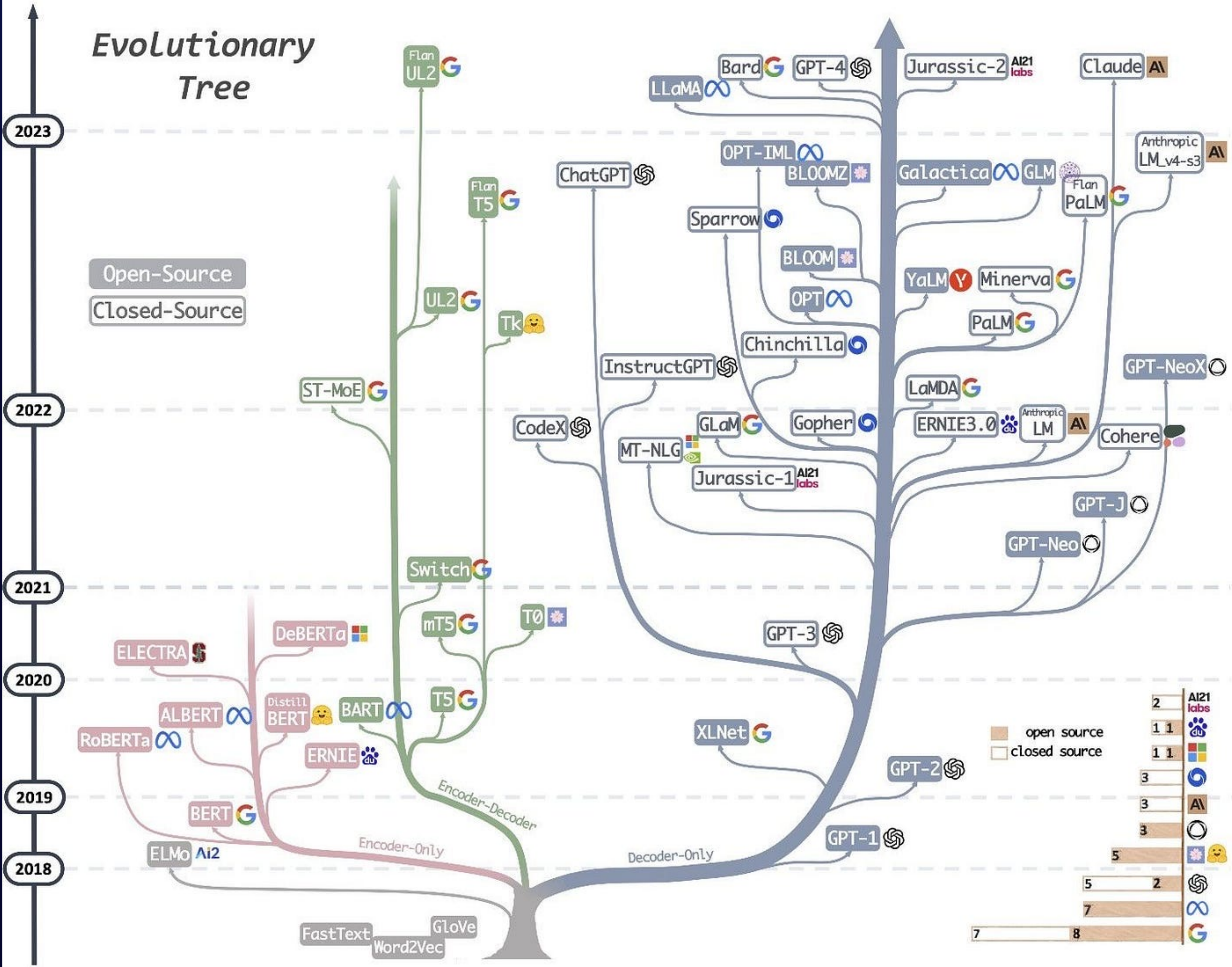
Natural Language Processing

Computer Vision

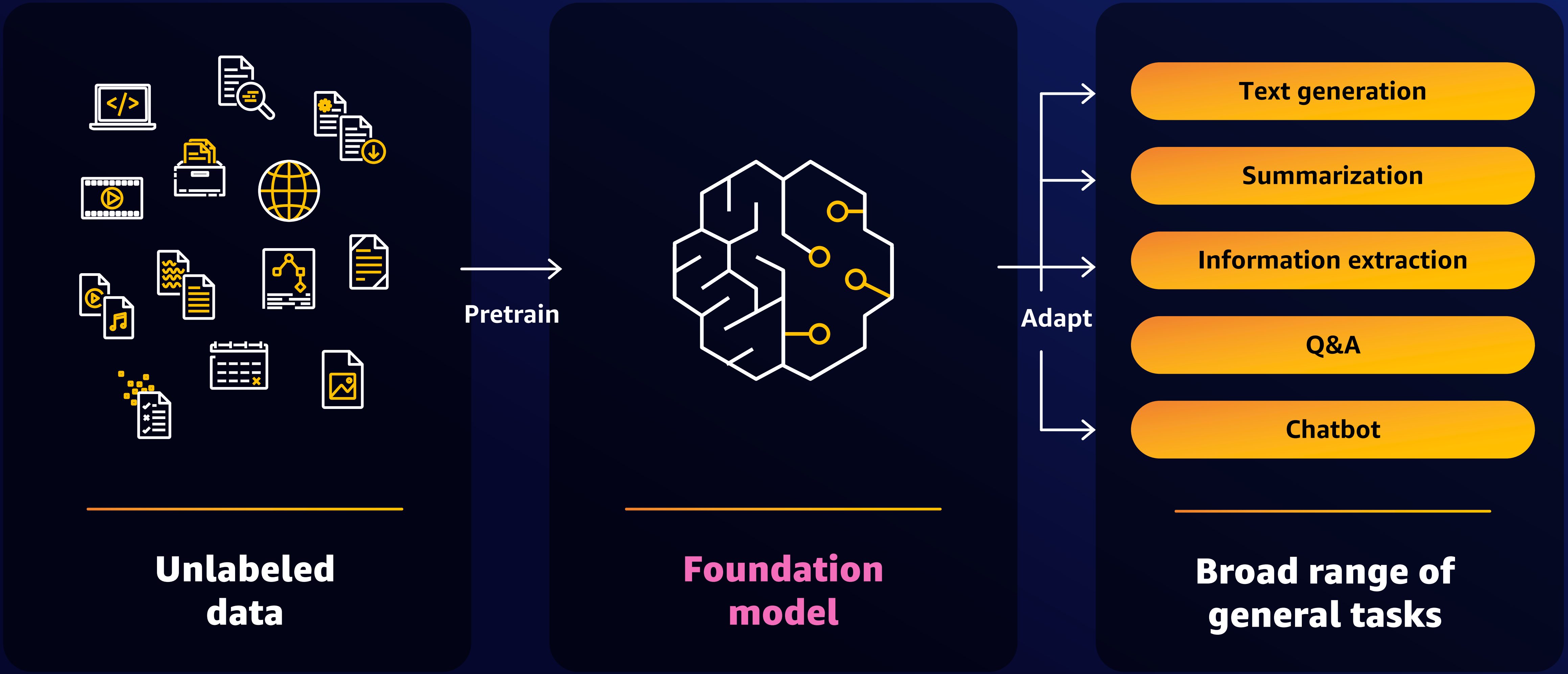
Generative AI & Large Language Models

col_jung

Medium Col Jung


Evolutionary Tree
2023
2022
2021
2020
2019
2018
Open-Source
Closed-Source
Flan UL2
Flan T5
UL2
Tk
ST-MoE
Switch
mT5
T0
T5
BART
Encoder-Decoder
ELECTRA
DeBERTa
ALBERT
Distill BERT
RoBERTa
ERNIE
BERT
ELMo
Ai2
Encoder-Only
FastText
GloVe
Word2Vec
LLaMA
Bard
GPT-4
Jurassic-2
AI21 labs
Claude
A\
ChatGPT
OPT-IML
BLOOMZ
Galactica
GLM
Flan PaLM
Anthropic LM_v4-s3
Sparrow
BLOOM
YaLM
Minerva
OPT
PaLM
Chinchilla
InstructGPT
GPT-NeoX
LaMDA
CodeX
GLaM
Gopher
ERNIE3.0
Anthropic LM
Cohere
MT-NLG
Jurassic-1
GPT-J
GPT-Neo
GPT-3
XLNet
GPT-2
GPT-1
Decoder-Only
open source
closed source
2
1 1
1 1
3
3
3
5
5 2
7
7 8

How does Generative AI work?



Unlabeled data

Pretrain

Foundation model

Adapt

- Text generation
- Summarization
- Information extraction
- Q&A
- Chatbot

Broad range of general tasks

Transformers

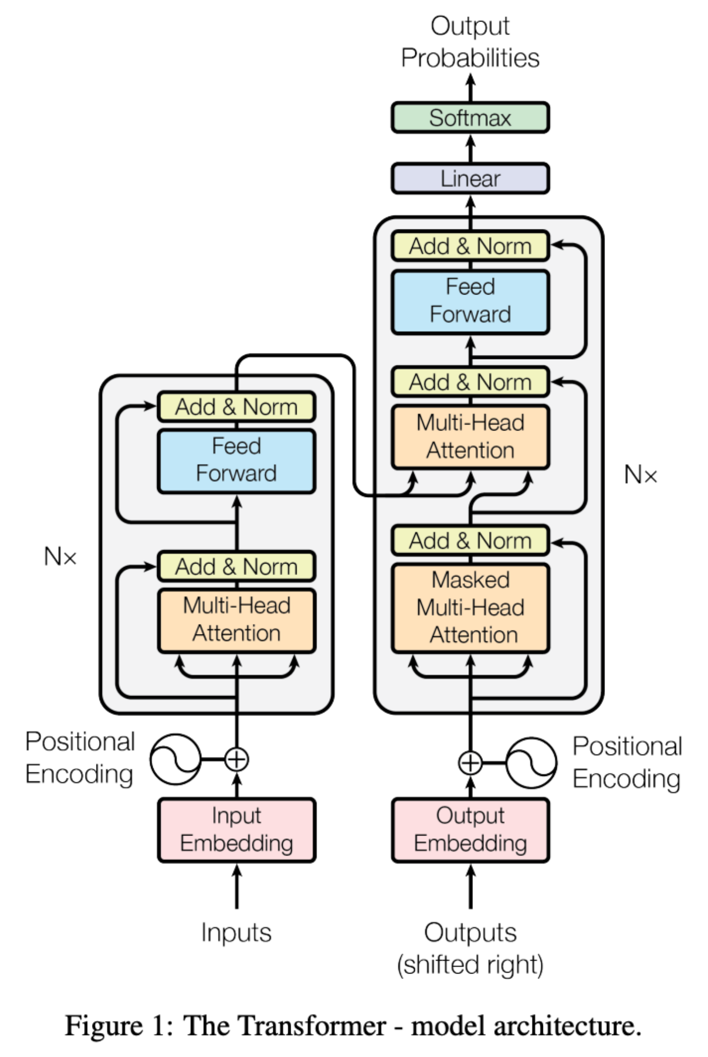
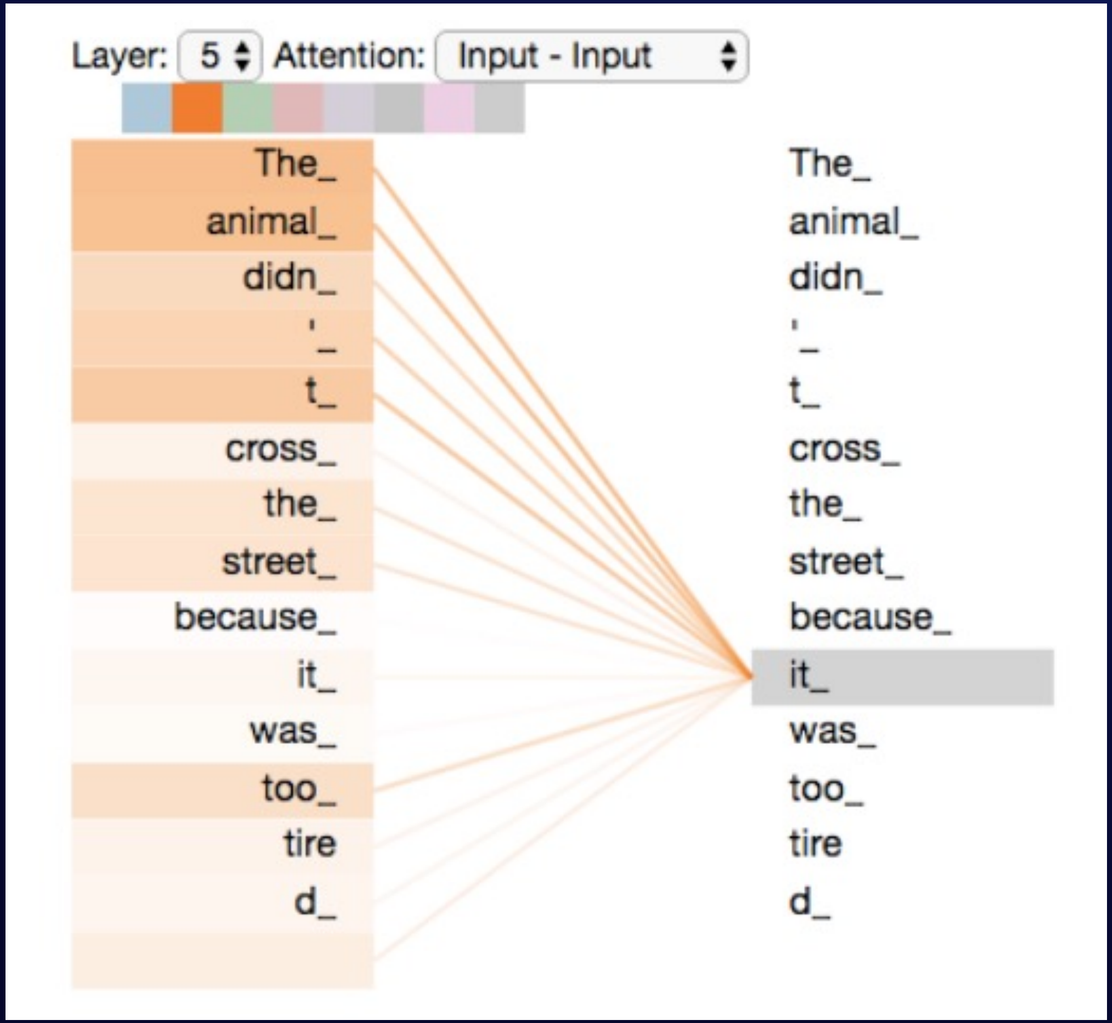


Figure 1: The Transformer - model architecture.

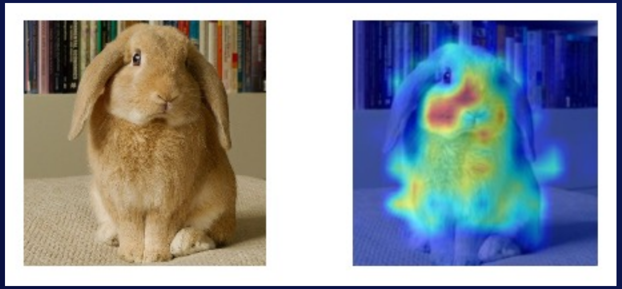
Source: Vaswani et al (2017): Attention Is All You Need

Attention Mechanism

“The animal didn´t cross the street because it was too tired”



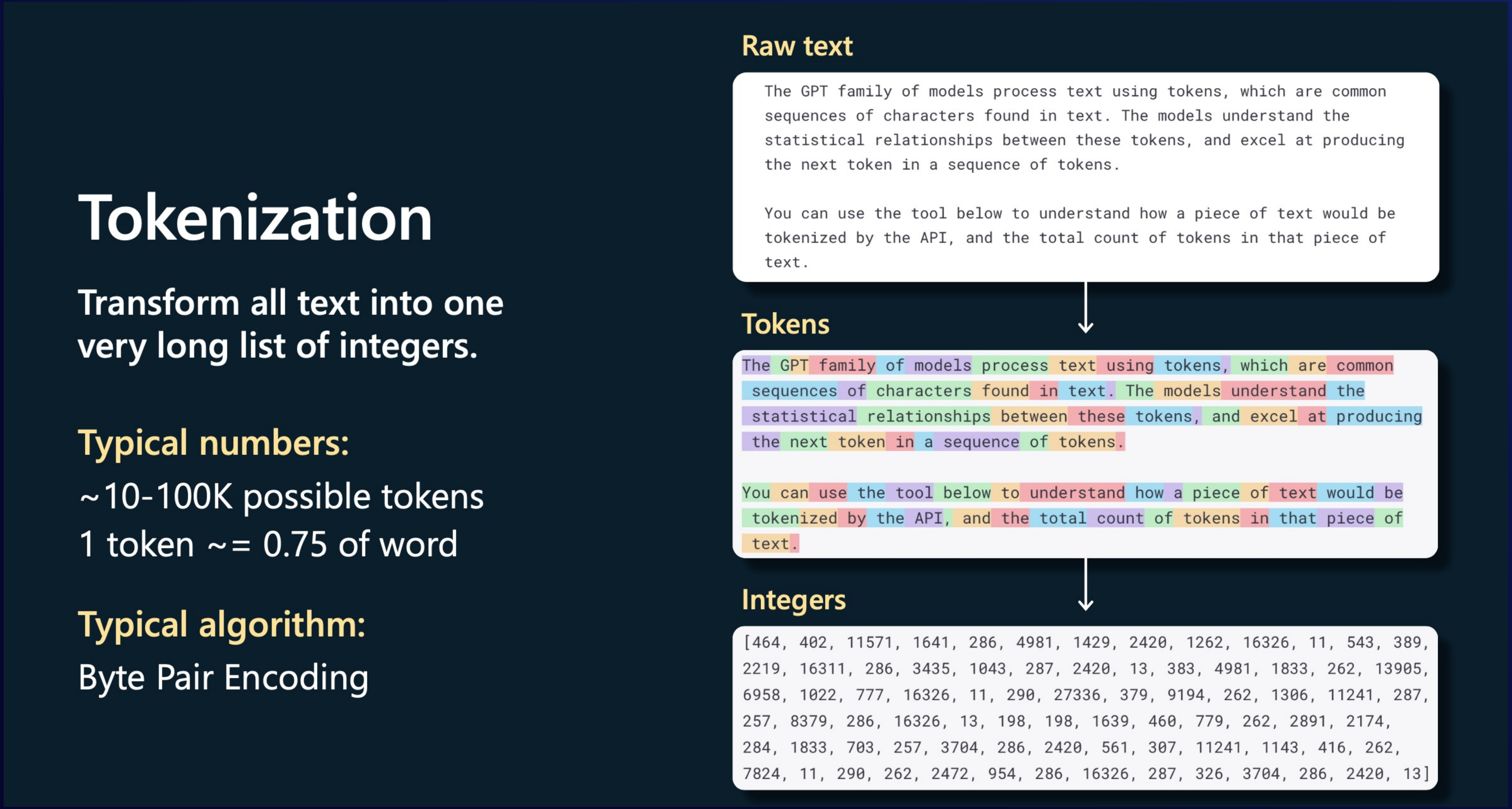
Layer: 5 Attention: Input - Input



Visual Transformer Attention Map

Source: Illustrated Transformer

Tokenization



Tokenization

Transform all text into one very long list of integers.

Typical numbers:

~10-100K possible tokens
1 token ~= 0.75 of word

Typical algorithm:

Byte Pair Encoding

Raw text

The GPT family of models process text using tokens, which are common sequences of characters found in text. The models understand the statistical relationships between these tokens, and excel at producing the next token in a sequence of tokens.

You can use the tool below to understand how a piece of text would be tokenized by the API, and the total count of tokens in that piece of text.

Tokens

The GPT family of models process text using tokens, which are common sequences of characters found in text. The models understand the statistical relationships between these tokens, and excel at producing the next token in a sequence of tokens.

You can use the tool below to understand how a piece of text would be tokenized by the API, and the total count of tokens in that piece of text.

Integers

```
[464, 402, 11571, 1641, 286, 4981, 1429, 2420, 1262, 16326, 11, 543, 389,
2219, 16311, 286, 3435, 1043, 287, 2420, 13, 383, 4981, 1833, 262, 13905,
6958, 1022, 777, 16326, 11, 290, 27336, 379, 9194, 262, 1306, 11241, 287,
257, 8379, 286, 16326, 13, 198, 198, 1639, 460, 779, 262, 2891, 2174,
284, 1833, 703, 257, 3704, 286, 2420, 561, 307, 11241, 1143, 416, 262,
7824, 11, 290, 262, 2472, 954, 286, 16326, 287, 326, 3704, 286, 2420, 13]
```

Source: State of GPT

Tokenization / Embedding

- Input text is split into ‘tokens’.

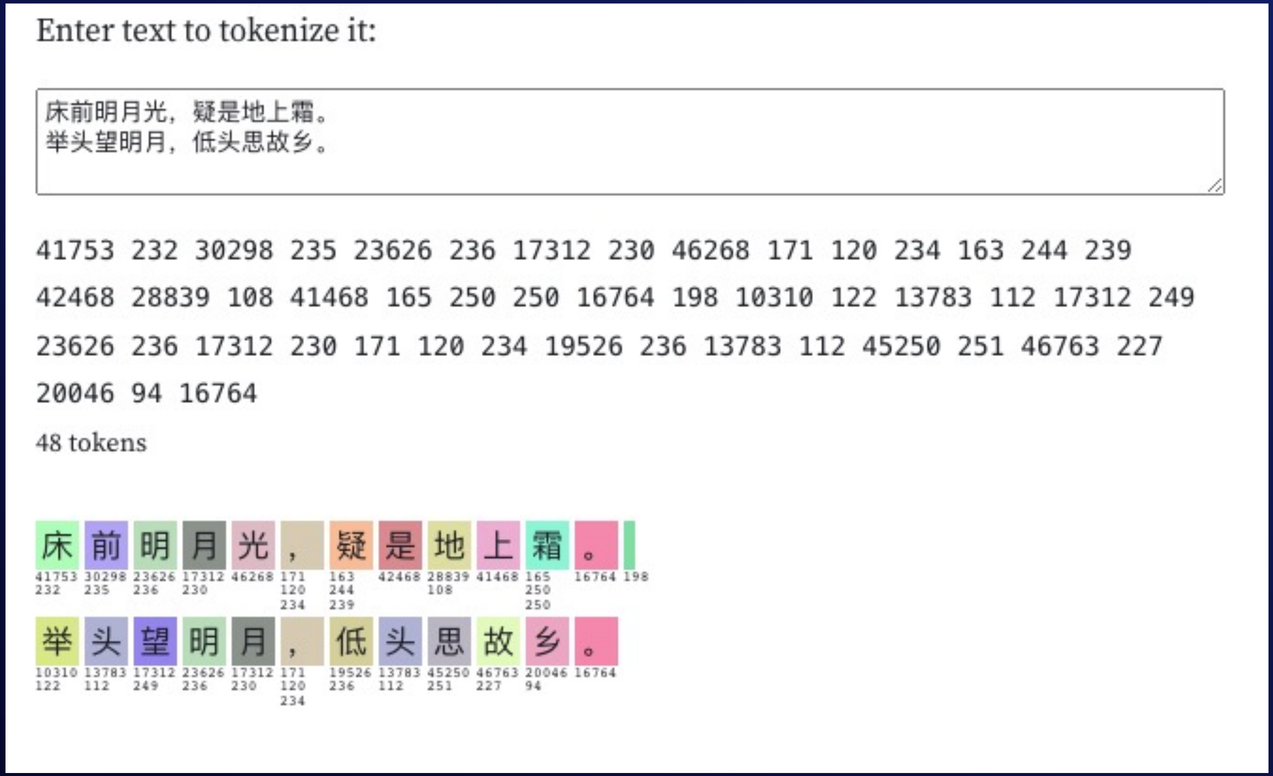
```
“The detective investigated” ->
[The_], [detective_], [invest],
[igat], [ed]
```

- The tokens are indices into the ‘vocabulary’:

```
[The_] [detective_] [invest]
[igat] [ed_] -> [3 721 68 1337 42]
```

- Each vocab entry corresponds to a learned dense vector:

```
[3 721 68 1337 42] -> [ [0.123, -
5.234, ...], [...], [...], [...],
[...] ]
```



Enter text to tokenize it:

床前明月光，疑是地上霜。
举头望明月，低头思故乡。

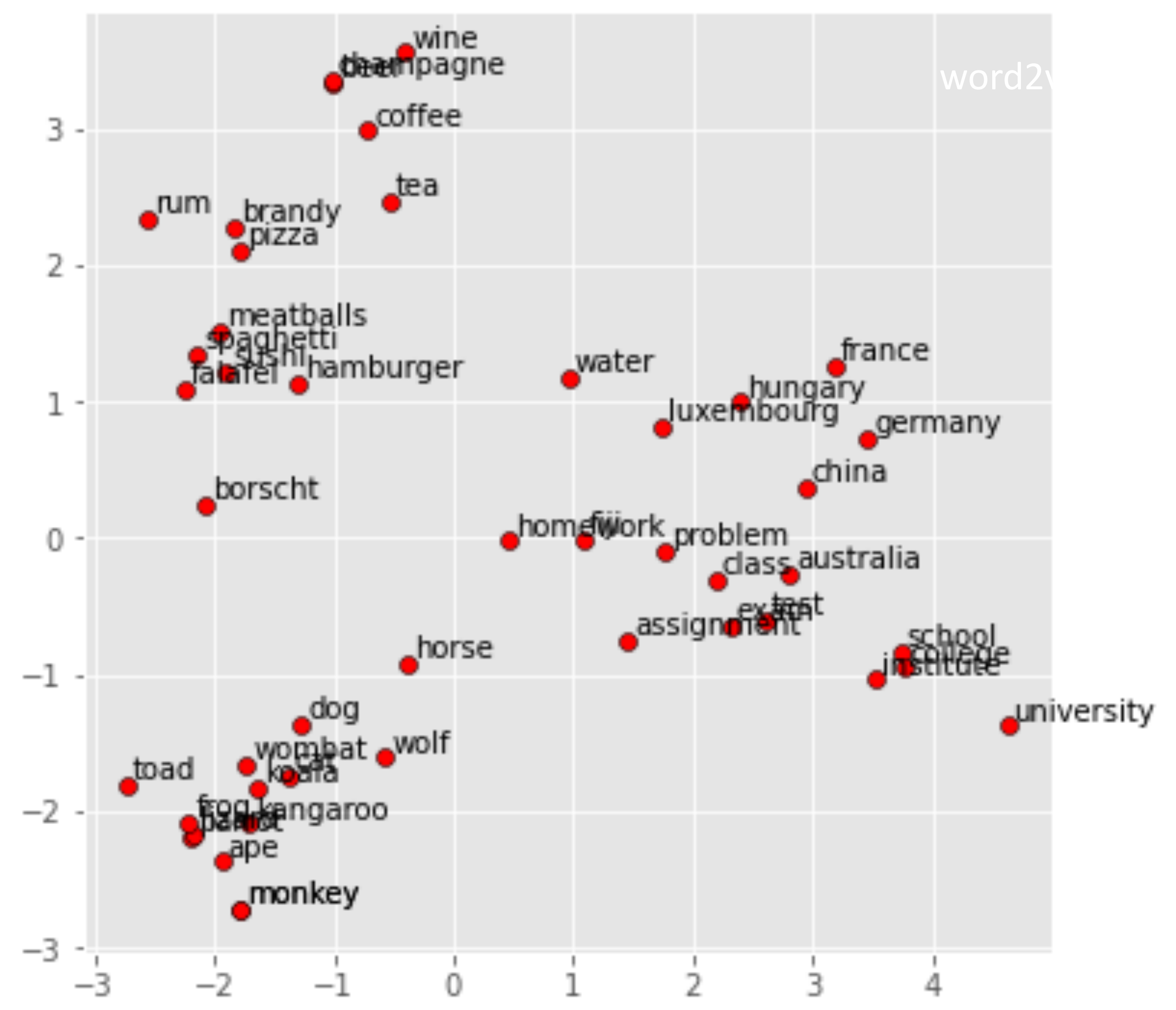
```
41753 232 30298 235 23626 236 17312 230 46268 171 120 234 163 244 239
42468 28839 108 41468 165 250 250 16764 198 10310 122 13783 112 17312 249
23626 236 17312 230 171 120 234 19526 236 13783 112 45250 251 46763 227
20046 94 16764
```

48 tokens

床	前	明	月	光	，	疑	是	地	上	霜	。	
41753 232	30298 235	23626 236	17312 230	46268	171 120 234	163 244 239	42468	28839 108	41468	165 250 250	16764	198

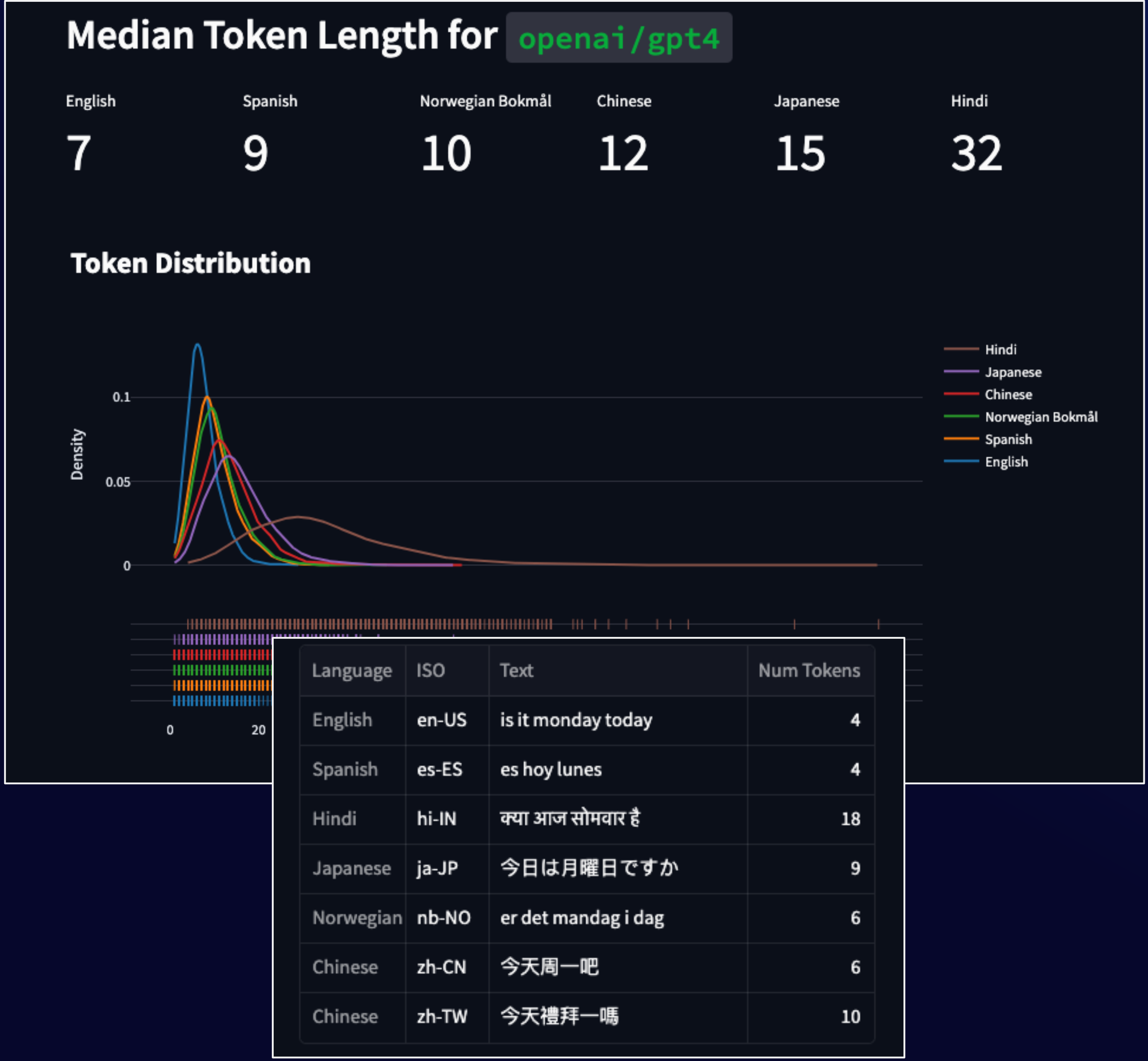
举	头	望	明	月	，	低	头	思	故	乡	。
10310 122	13783 112	17312 249	23626 236	17312 230	171 120 234	19526 236	13783 112	45250 251	46763 227	20046 94	16764

Embedding Concept



word2v

Amount of tokens needed are different across languages and models

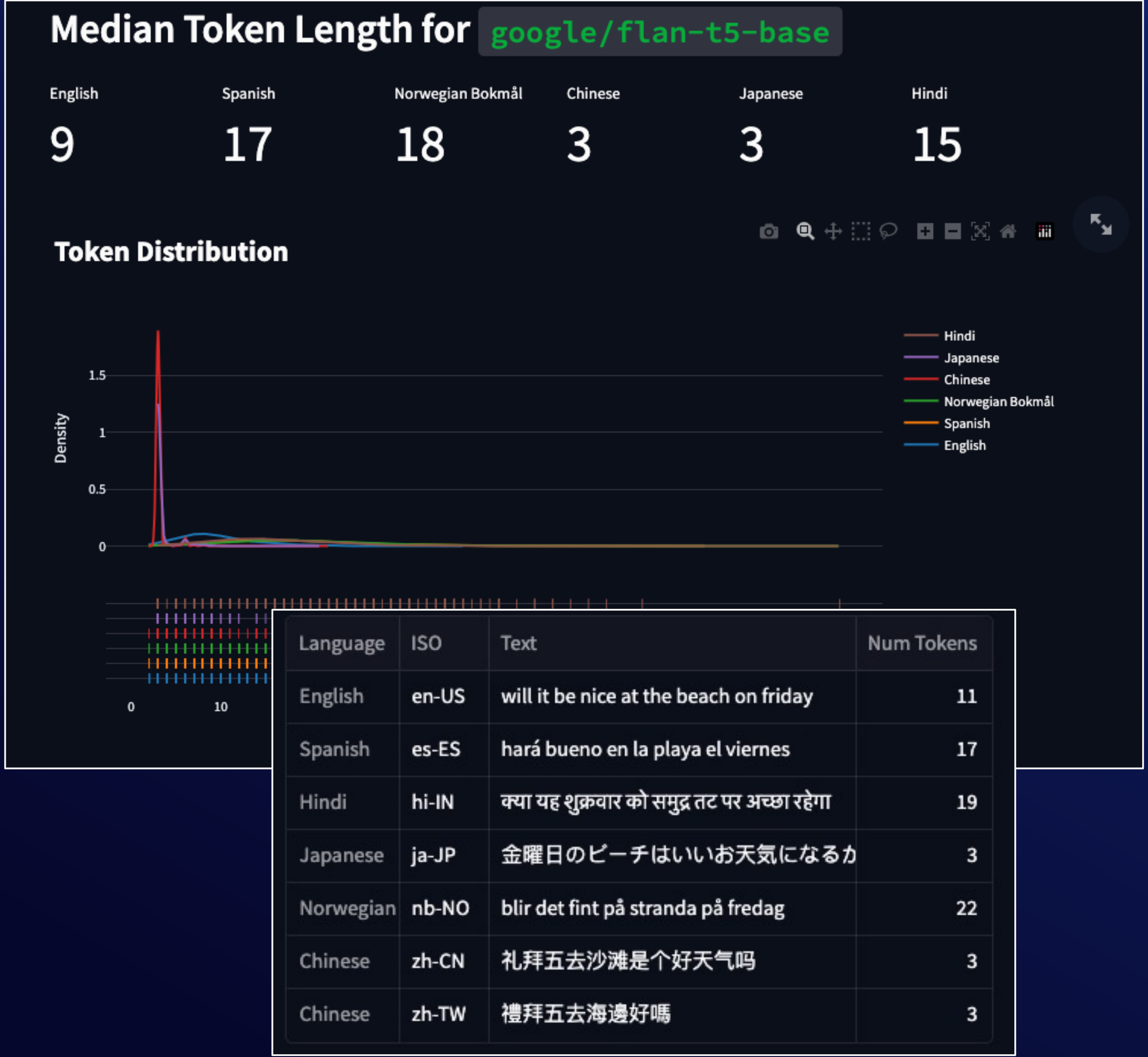


Median Token Length for `openai/gpt4`

English	Spanish	Norwegian Bokmål	Chinese	Japanese	Hindi
7	9	10	12	15	32

Token Distribution

Language	ISO	Text	Num Tokens
English	en-US	is it monday today	4
Spanish	es-ES	es hoy lunes	4
Hindi	hi-IN	क्या आज सोमवार है	18
Japanese	ja-JP	今日は月曜日ですか	9
Norwegian	nb-NO	er det mandag i dag	6
Chinese	zh-CN	今天周一吧	6
Chinese	zh-TW	今天禮拜一嗎	10



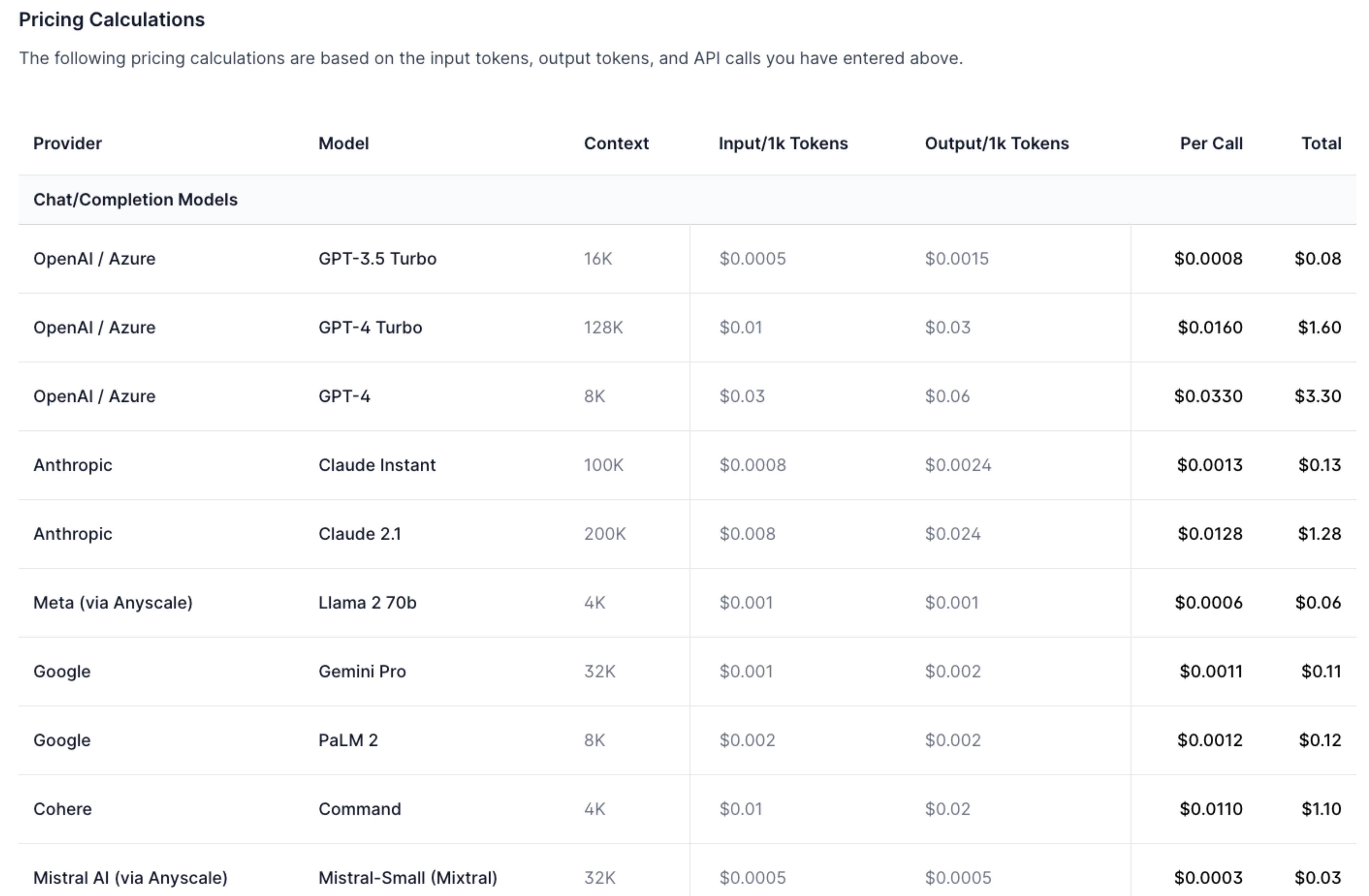
Median Token Length for `google/flan-t5-base`

English	Spanish	Norwegian Bokmål	Chinese	Japanese	Hindi
9	17	18	3	3	15

Token Distribution

Language	ISO	Text	Num Tokens
English	en-US	will it be nice at the beach on friday	11
Spanish	es-ES	hará bueno en la playa el viernes	17
Hindi	hi-IN	क्या यह शुक्रवार को समुद्र तट पर अच्छा रहेगा	19
Japanese	ja-JP	金曜日のビーチはいいお天気になるか	3
Norwegian	nb-NO	blir det fint på stranda på fredag	22
Chinese	zh-CN	礼拜五去沙滩是个好天气吗	3
Chinese	zh-TW	禮拜五去海邊好嗎	3

https://huggingface.co/spaces/yenniejun/tokenizers-languages



Pricing Calculations

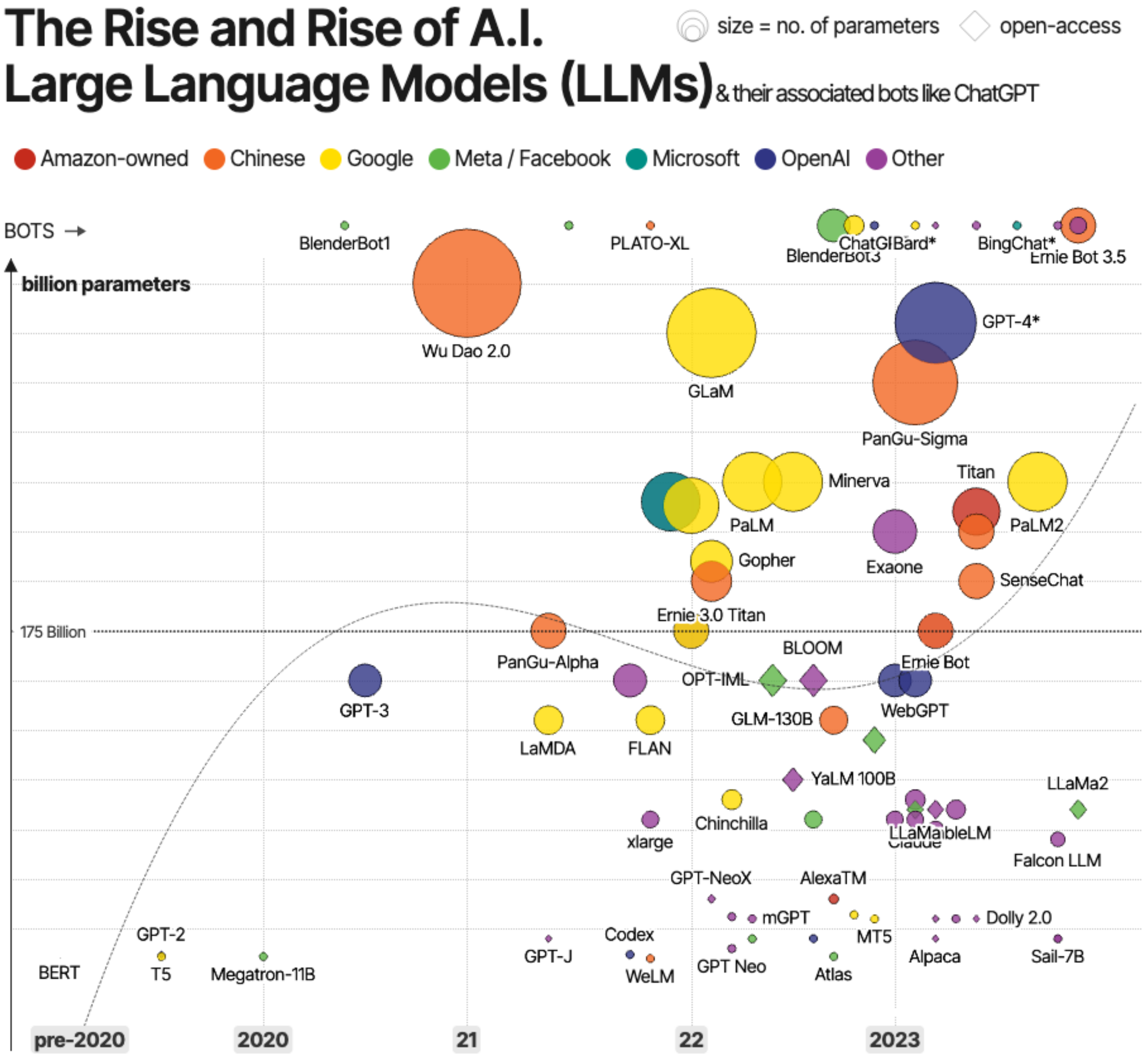
The following pricing calculations are based on the input tokens, output tokens, and API calls you have entered above.

Provider	Model	Context	Input/1k Tokens	Output/1k Tokens	Per Call	Total
Chat/Completion Models						
OpenAI / Azure	GPT-3.5 Turbo	16K	$0.0005	$0.0015	$0.0008	$0.08
OpenAI / Azure	GPT-4 Turbo	128K	$0.01	$0.03	$0.0160	$1.60
OpenAI / Azure	GPT-4	8K	$0.03	$0.06	$0.0330	$3.30
Anthropic	Claude Instant	100K	$0.0008	$0.0024	$0.0013	$0.13
Anthropic	Claude 2.1	200K	$0.008	$0.024	$0.0128	$1.28
Meta (via Anyscale)	Llama 2 70b	4K	$0.001	$0.001	$0.0006	$0.06
Google	Gemini Pro	32K	$0.001	$0.002	$0.0011	$0.11
Google	PaLM 2	8K	$0.002	$0.002	$0.0012	$0.12
Cohere	Command	4K	$0.01	$0.02	$0.0110	$1.10
Mistral AI (via Anyscale)	Mistral-Small (Mixtral)	32K	$0.0005	$0.0005	$0.0003	$0.03

https://docsbot.ai/tools/gpt-openai-api-pricing-calculator

Foundational Models Zoo

Llama, Alpaca, Vicuna, Guanaco …



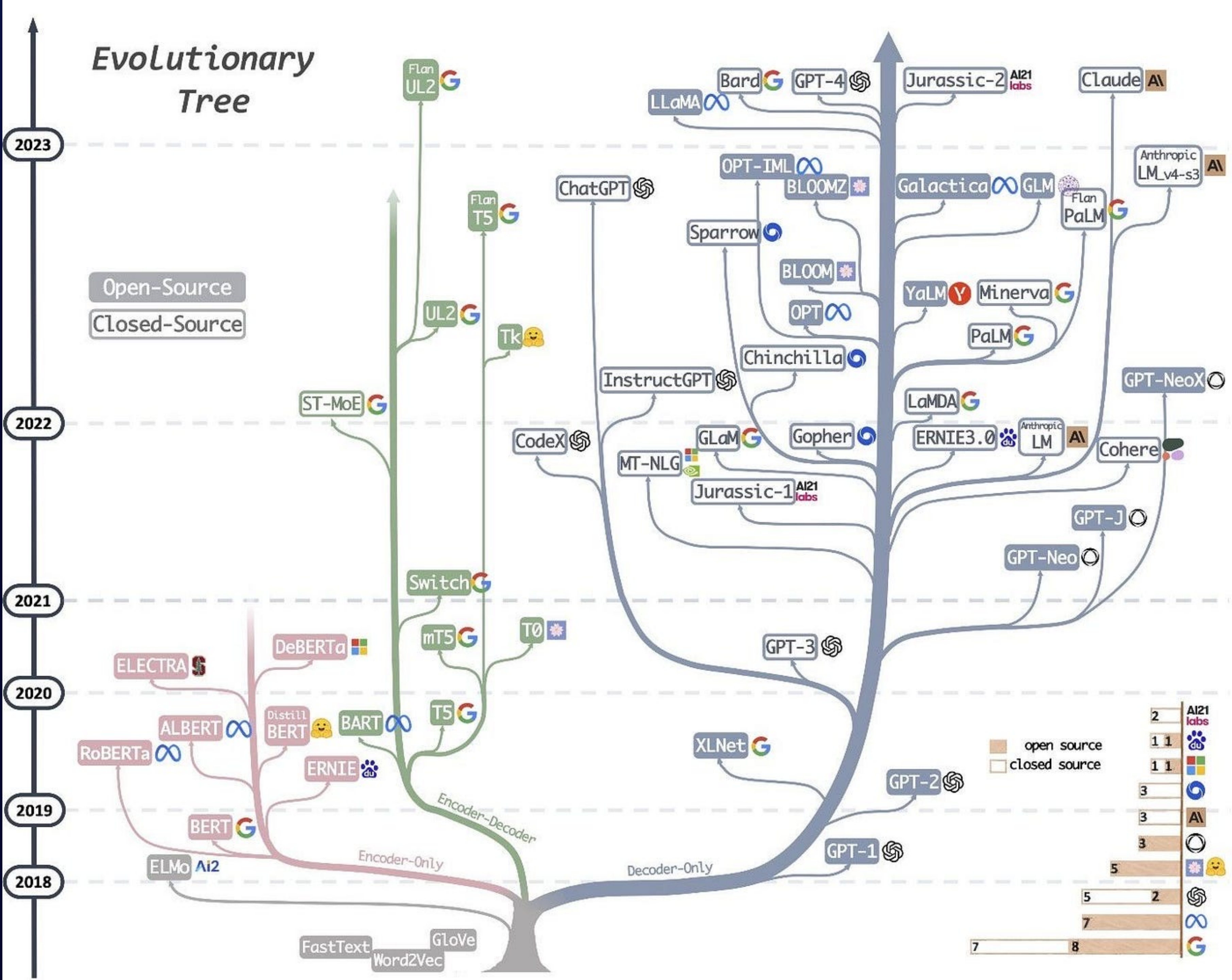
The Rise and Rise of A.I. Large Language Models (LLMs) & their associated bots like ChatGPT

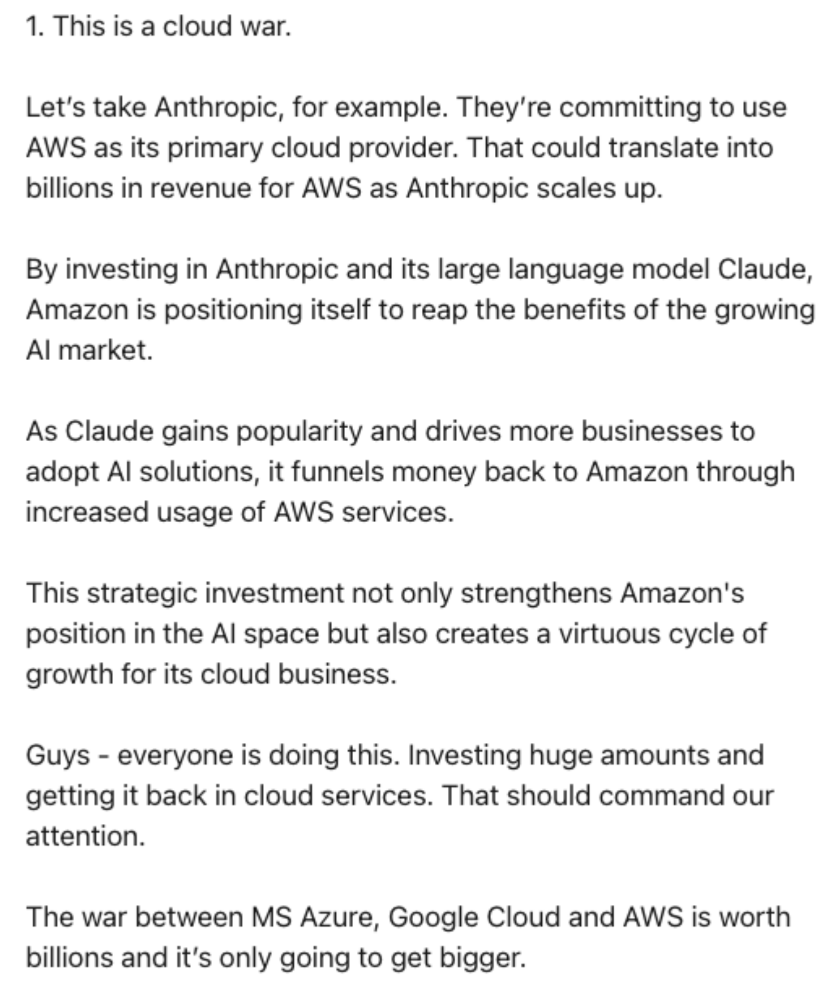
David McCandless, Tom Evans, Paul Barton

Information is Beautiful // UPDATED 27th Jul 23

source: news reports, LifeArchitect.ai

* = parameters undisclosed // see the data


Evolutionary Tree
2023
2022
2021
2020
2019
2018
Open-Source
Closed-Source
Flan UL2
Flan T5
UL2
Tk
ST-MoE
Switch
mT5
T0
T5
BART
Encoder-Decoder
ELECTRA
DeBERTa
ALBERT
Distill BERT
RoBERTa
ERNIE
BERT
ELMo
Ai2
Encoder-Only
FastText
Word2Vec
GloVe
Decoder-Only
LLaMA
Bard
GPT-4
Jurassic-2
AI21 labs
Claude
A\
ChatGPT
OPT-IML
BLOOMZ
Galactica
GLM
Anthropic LM_v4-s3
Flan PaLM
Sparrow
BLOOM
YaLM
Minerva
OPT
PaLM
Chinchilla
InstructGPT
GPT-NeoX
LaMDA
CodeX
GLaM
Gopher
ERNIE3.0
Anthropic LM
Cohere
MT-NLG
Jurassic-1
GPT-J
GPT-Neo
GPT-3
XLNet
GPT-2
GPT-1
open source
closed source
2
1 1
1 1
3
3
3
5
5 2
7
7 8



1. This is a cloud war.

Let's take Anthropic, for example. They're committing to use AWS as its primary cloud provider. That could translate into billions in revenue for AWS as Anthropic scales up.

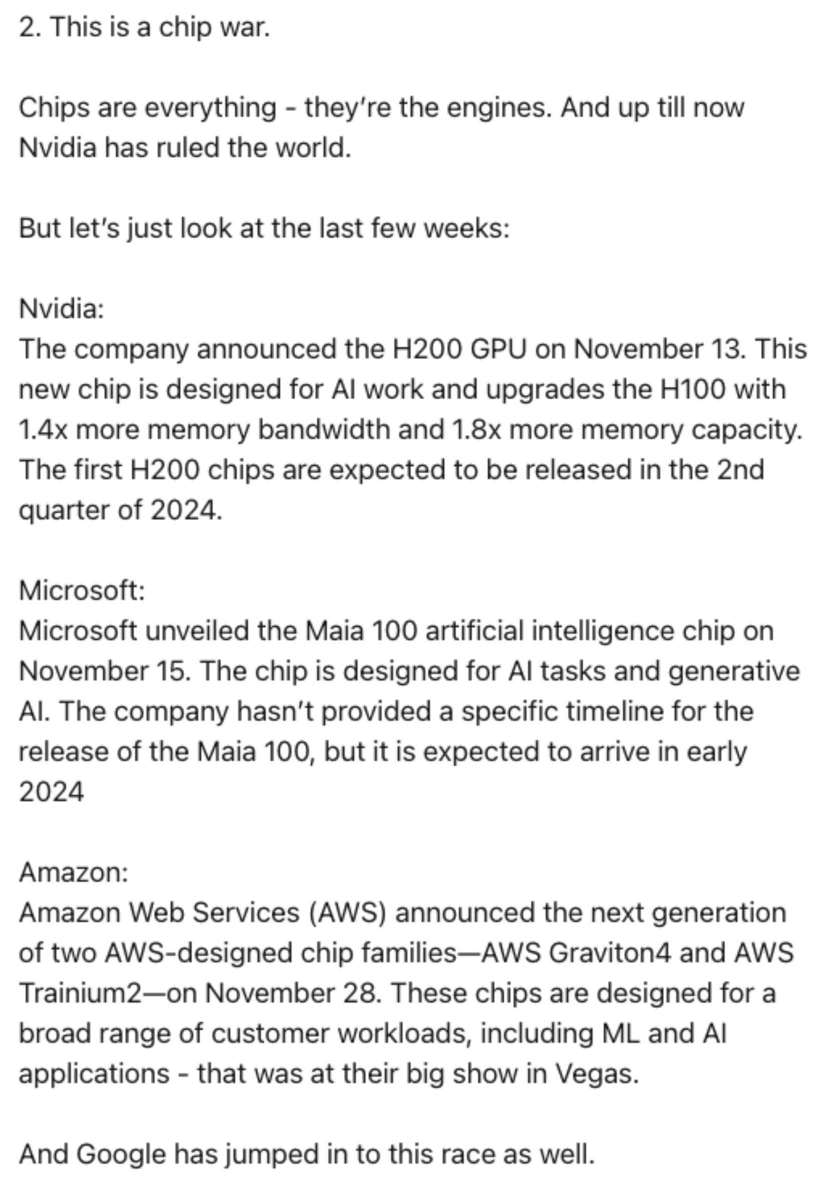
By investing in Anthropic and its large language model Claude, Amazon is positioning itself to reap the benefits of the growing AI market.

As Claude gains popularity and drives more businesses to adopt AI solutions, it funnels money back to Amazon through increased usage of AWS services.

This strategic investment not only strengthens Amazon's position in the AI space but also creates a virtuous cycle of growth for its cloud business.

Guys - everyone is doing this. Investing huge amounts and getting it back in cloud services. That should command our attention.

The war between MS Azure, Google Cloud and AWS is worth billions and it's only going to get bigger.



2. This is a chip war.

Chips are everything - they're the engines. And up till now Nvidia has ruled the world.

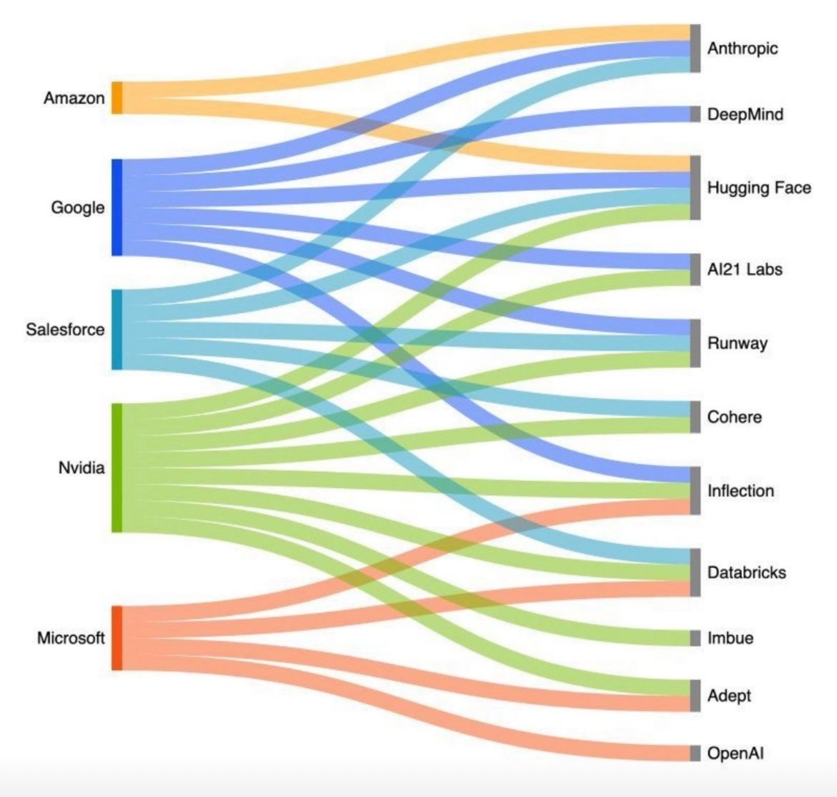
But let's just look at the last few weeks:

Nvidia:
The company announced the H200 GPU on November 13. This new chip is designed for AI work and upgrades the H100 with 1.4x more memory bandwidth and 1.8x more memory capacity. The first H200 chips are expected to be released in the 2nd quarter of 2024.

Microsoft:
Microsoft unveiled the Maia 100 artificial intelligence chip on November 15. The chip is designed for AI tasks and generative AI. The company hasn't provided a specific timeline for the release of the Maia 100, but it is expected to arrive in early 2024

Amazon:
Amazon Web Services (AWS) announced the next generation of two AWS-designed chip families—AWS Graviton4 and AWS Trainium2—on November 28. These chips are designed for a broad range of customer workloads, including ML and AI applications - that was at their big show in Vegas.

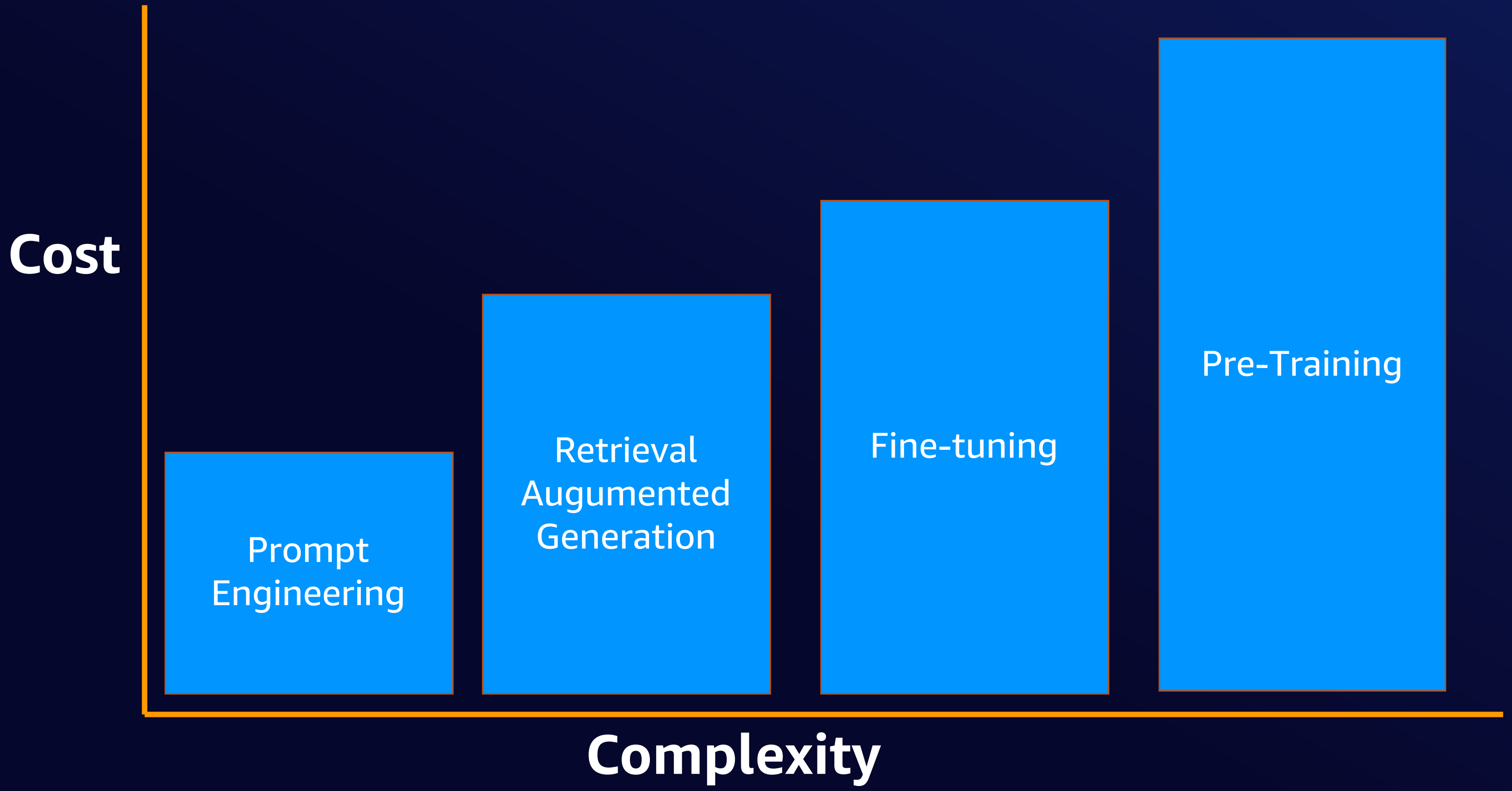
And Google has jumped in to this race as well.



Agenda

- Motivation
- Basic concepts
- Use case and patterns ⇐
- Tools and resources

The Era of LLMs



Cost

Prompt Engineering

Retrieval Augumented Generation

Fine-tuning

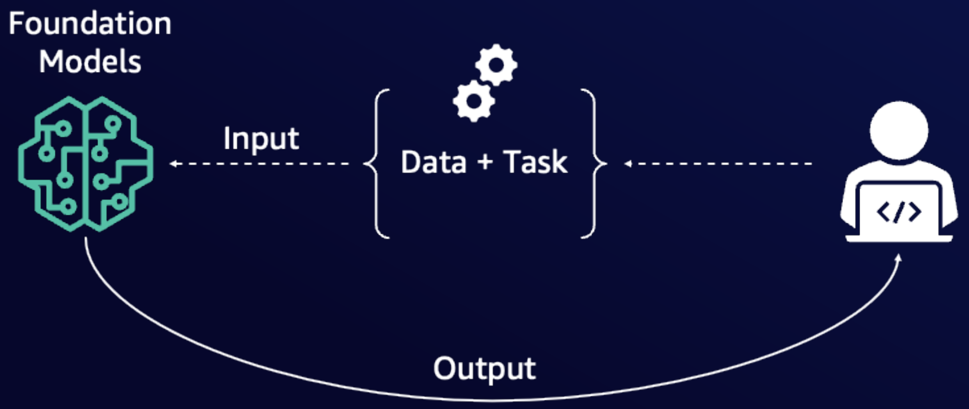
Pre-Training

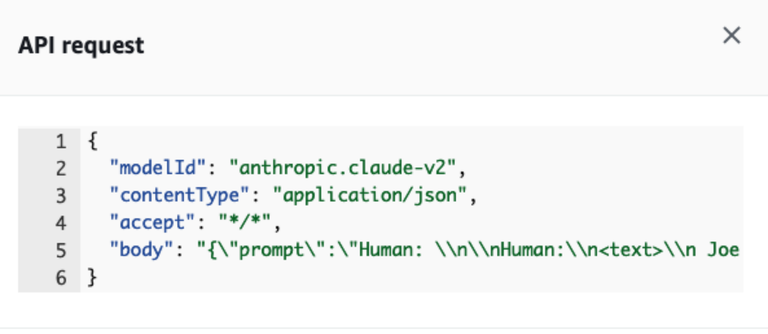
Complexity

Prompt Engineering 101

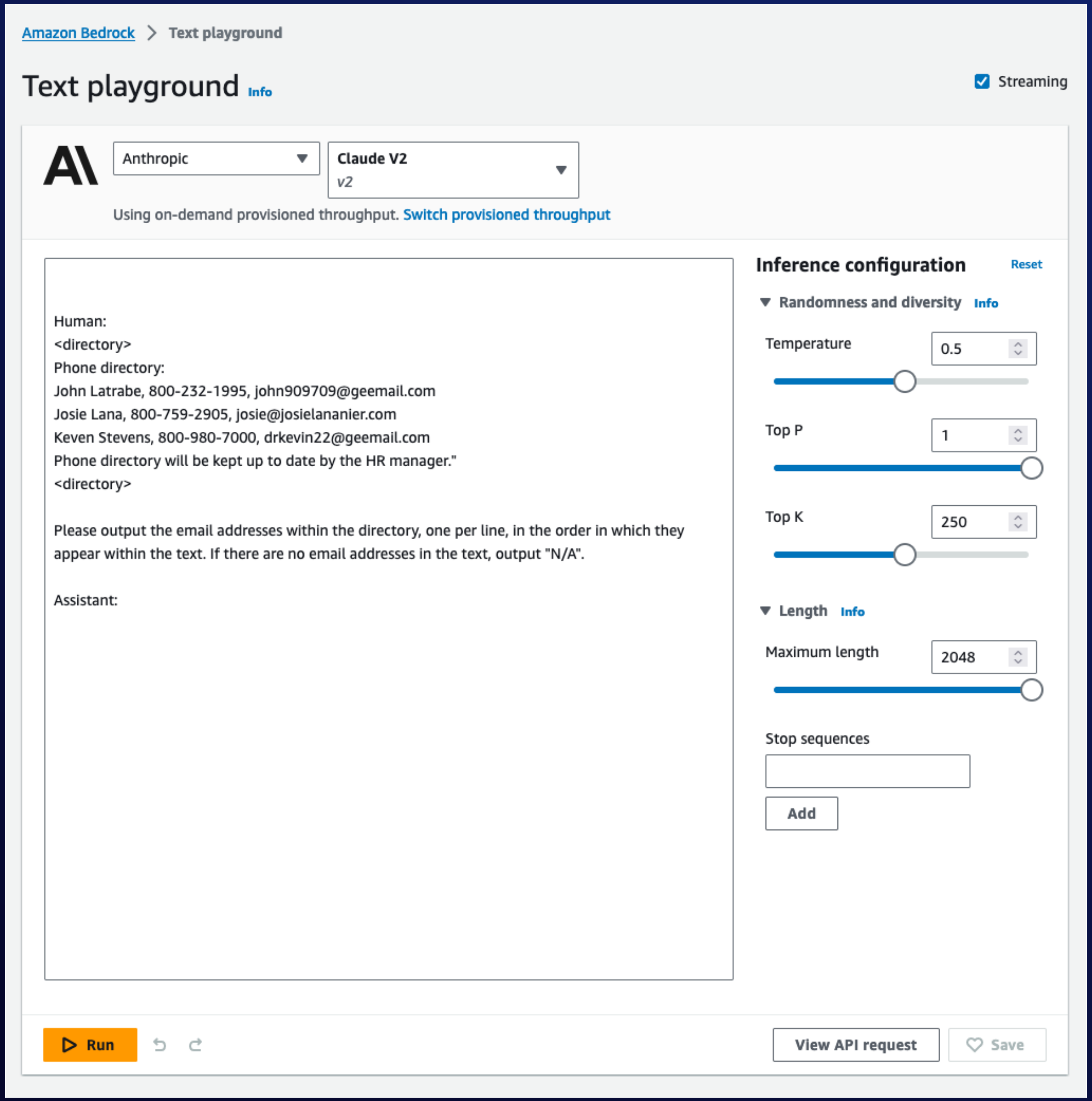
PROMPT	+	CONTEXT	+	FM MODEL	=	OUTPUT
Task description or instruction to model Indicate Output Format		Additional/external company specific information to steer the model output		LLM Model for text, image, translation, music, video, etc		Final output generated by the LLM model
Task		**Example or Context**		**Foundation Model**		**Task result**

1. Instruction Prompting




API request

```
{
  "modelId": "anthropic.claude-v2",
  "contentType": "application/json",
  "accept": "*/*",
  "body": "{\"prompt\":\"Human: \\n\\nHuman:\\n<text>\\n Joe
}
```



Text playground

Anthropic
Claude V2
v2

Using on-demand provisioned throughput. Switch provisioned throughput

Human:
<directory>
Phone directory:
John Latrabe, 800-232-1995, john909709@geemail.com
Josie Lana, 800-759-2905, josie@josielananier.com
Keven Stevens, 800-980-7000, drkevin22@geemail.com
Phone directory will be kept up to date by the HR manager."
<directory>

Please output the email addresses within the directory, one per line, in the order in which they appear within the text. If there are no email addresses in the text, output "N/A".

Assistant:

Inference configuration

▼ Randomness and diversity

Temperature: 0.5
Top P: 1
Top K: 250

▼ Length

Maximum length: 2048

Stop sequences

2. Few Shot Prompting

Zero Shot

Input

Tell me sentiment of this statement:
I loved the pizza at that Italian pizzeria

Output

The statement expresses a positive sentiment towards a pizza restaurant.

Few Shot

Input

Best Pakistani restaurant in Zurich: Positive
New York stinks, don’t go there: Negative
The talk was on Generative AI: Neutral
This is a rip-off, store not recommended:

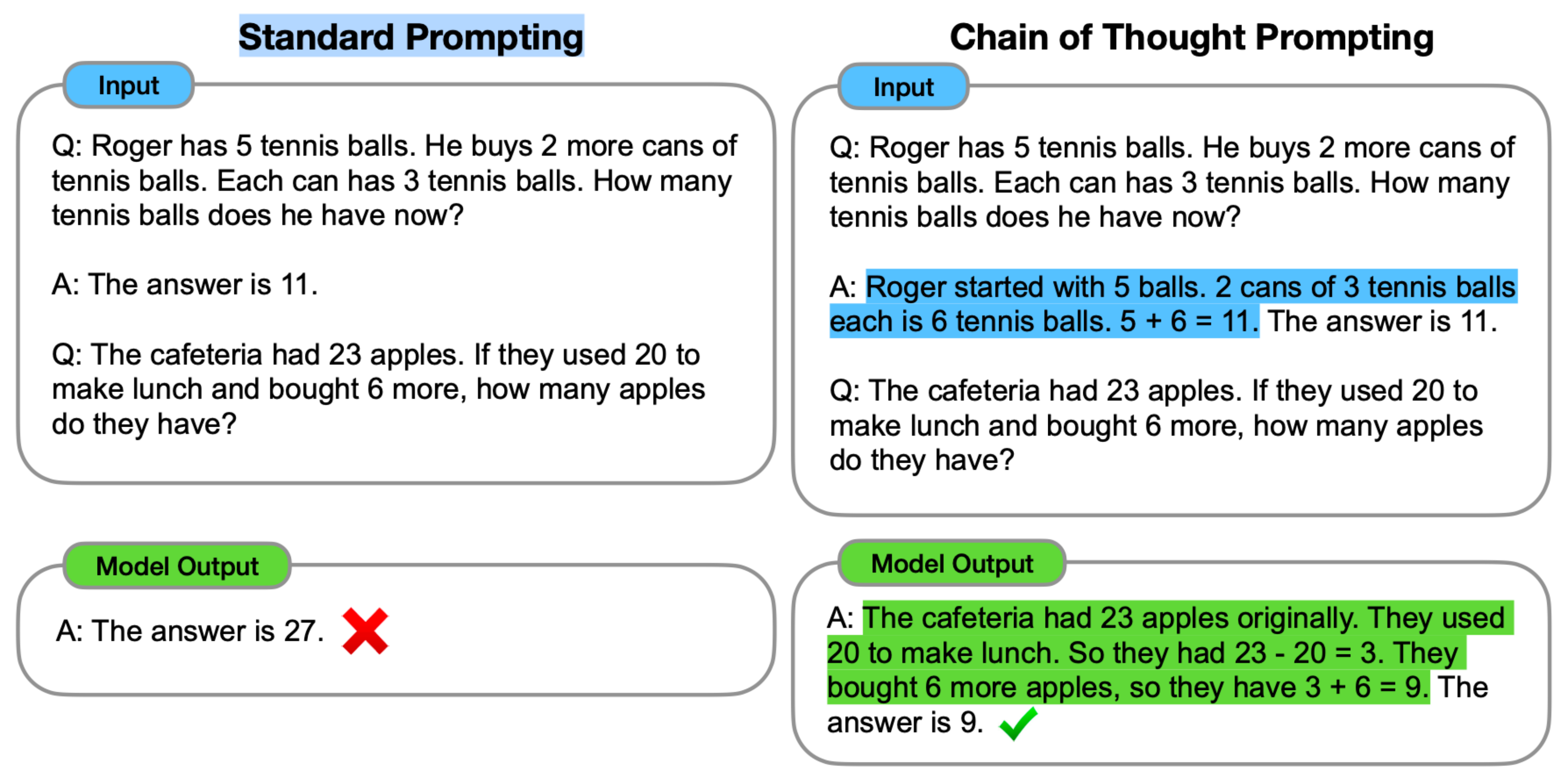
Output

Negative

3. Chain-of-Thought (CoT) Prompting

Effective with:
- Arithmetic
- Common Sense
- Symbolic reasoning

Works best with ~100B parameter model(s)



Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ✗

Chain of Thought Prompting

Input

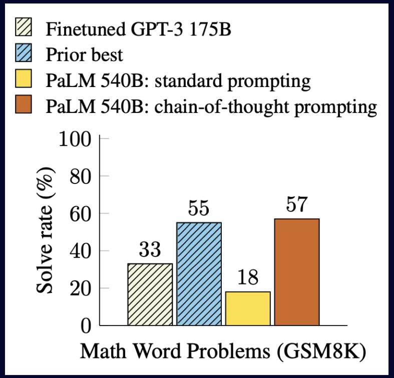
Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

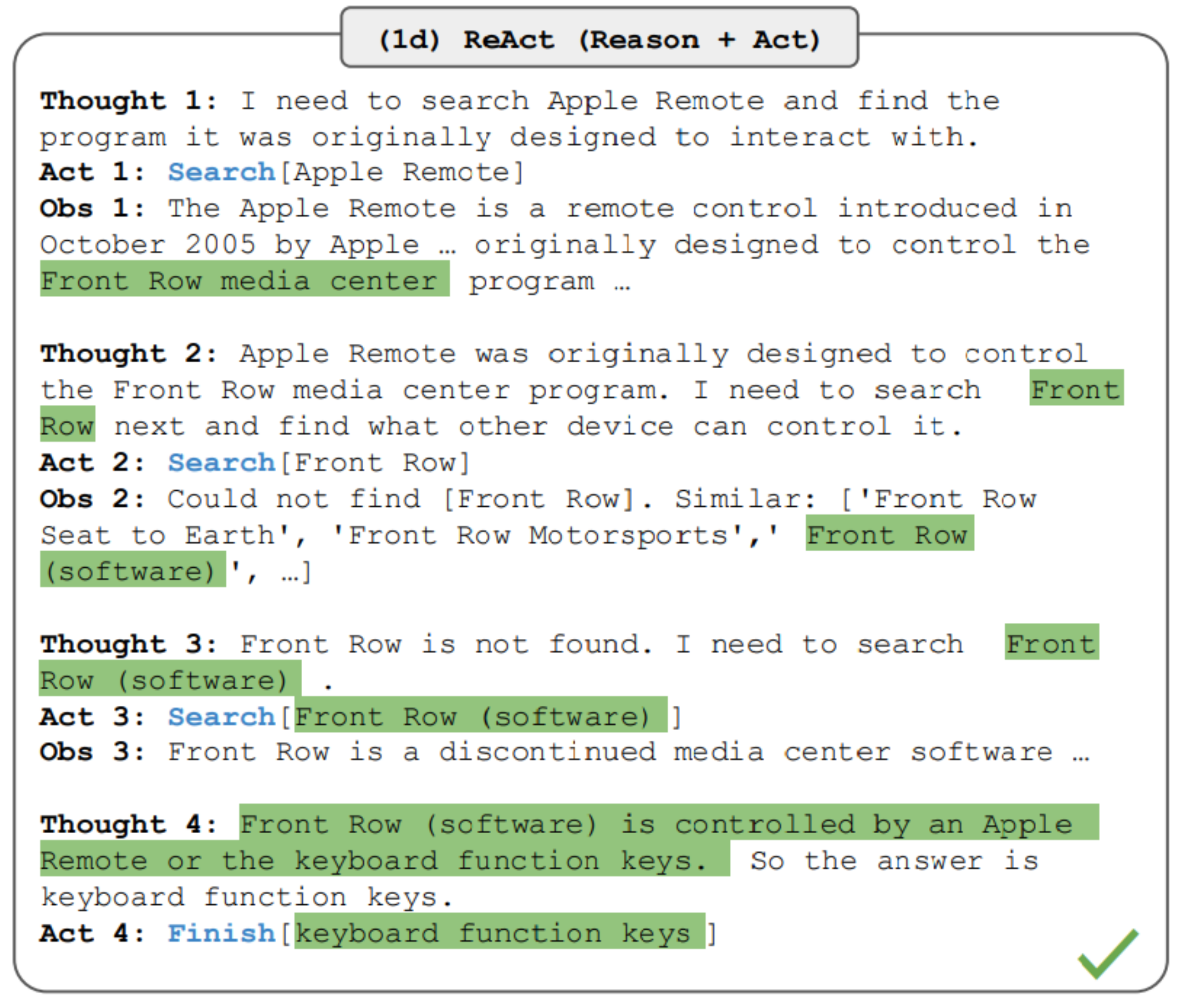
A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. ✓



* https://learnprompting.org/docs/intermediate/chain_of_thought

4. ReAct (Reasoning + Action) Prompting

- Combines reasoning and acting with LLMs.
- Prompts LLMs to generate verbal reasoning traces and actions for a task.



(1d) ReAct (Reason + Act)

Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with.
Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple … originally designed to control the Front Row media center program …

Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.
Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports',' Front Row (software) ', …]

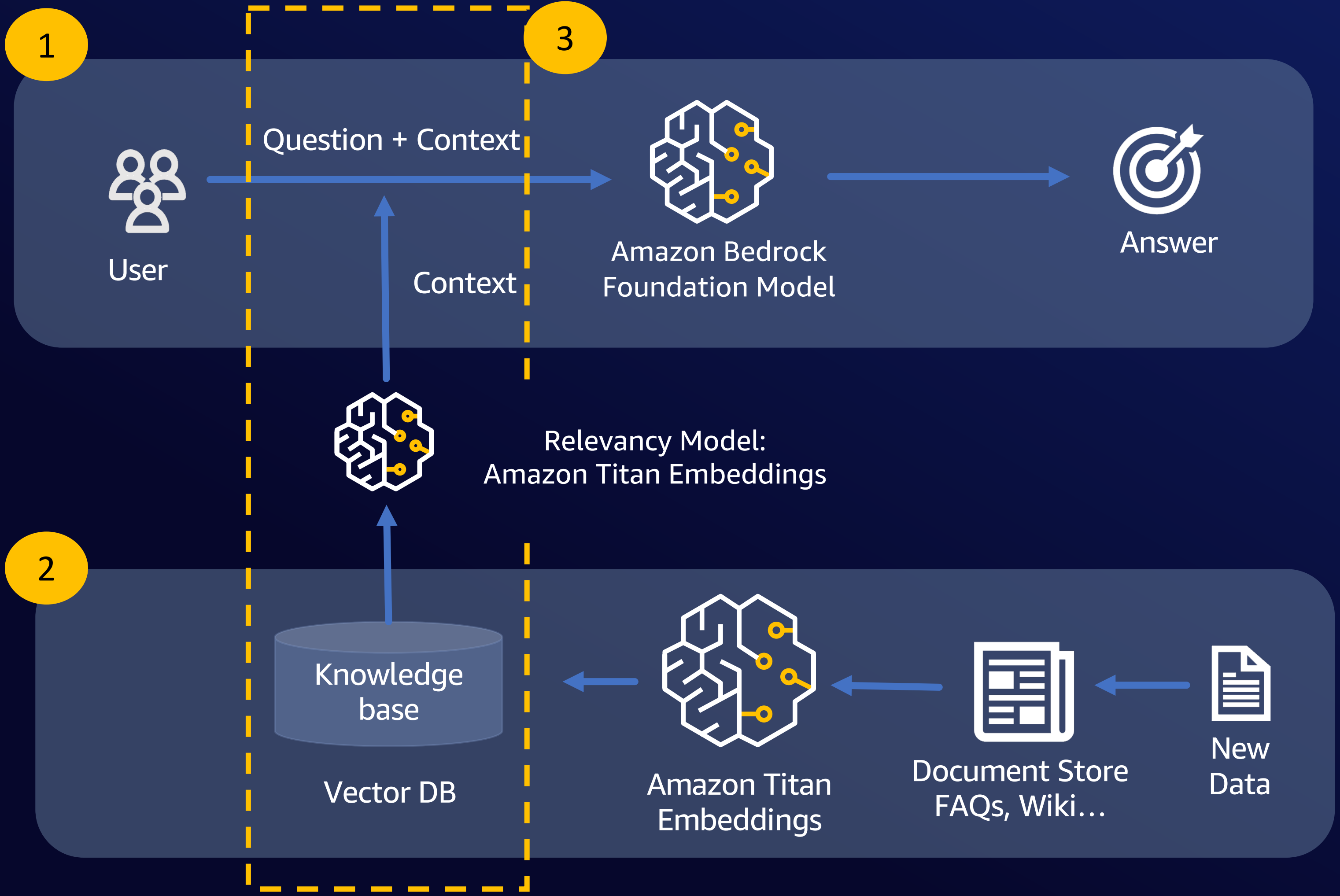
Thought 3: Front Row is not found. I need to search Front Row (software) .
Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software …

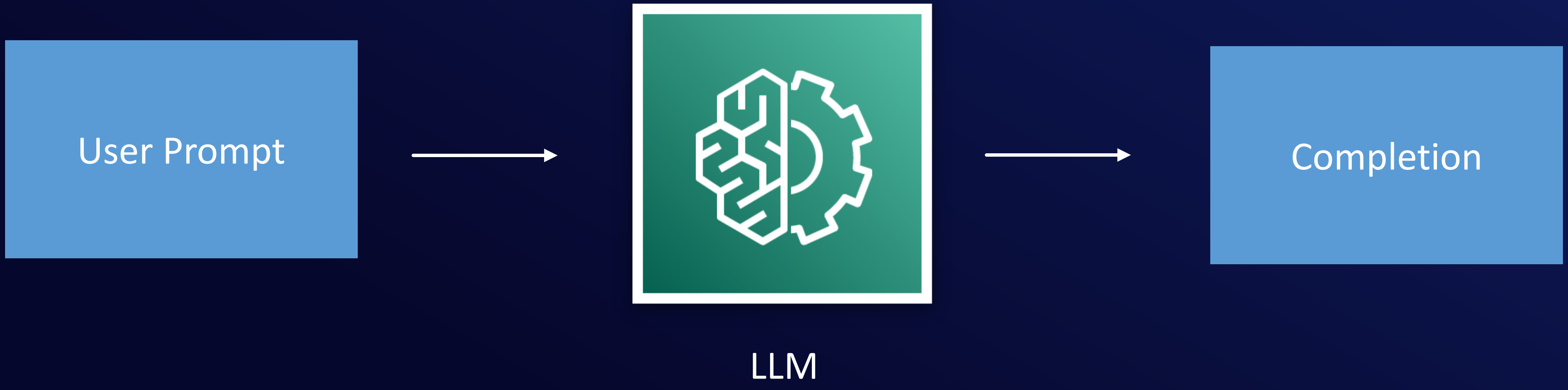
Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.
Act 4: Finish[keyboard function keys]

✓

Image Source: Yao et al., 2022

5. Retrieval Augmented Generation (RAG)



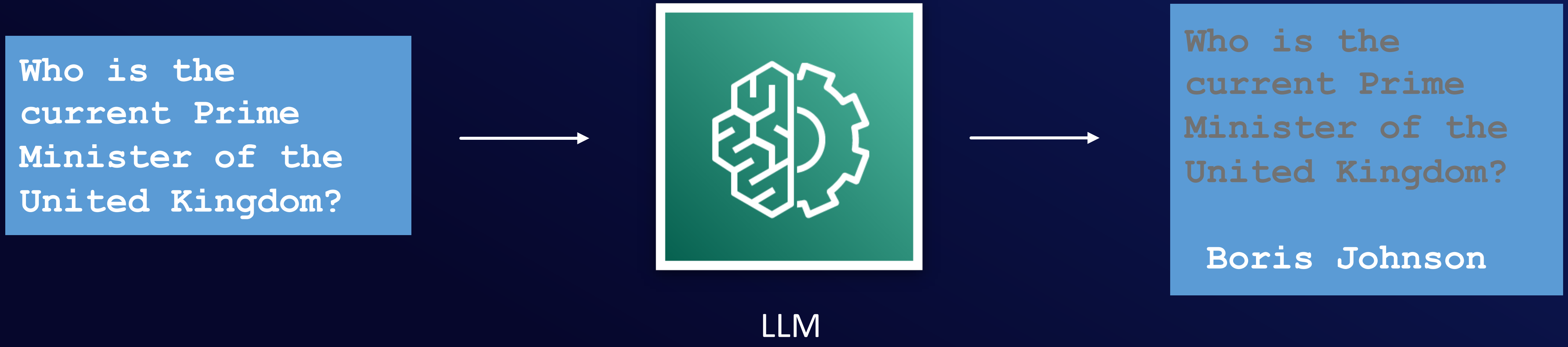


User Prompt

LLM

Completion

Knowledge cut-offs in LLMs



`Who is the current Prime Minister of the United Kingdom?`

LLM

`Who is the current Prime Minister of the United Kingdom?`

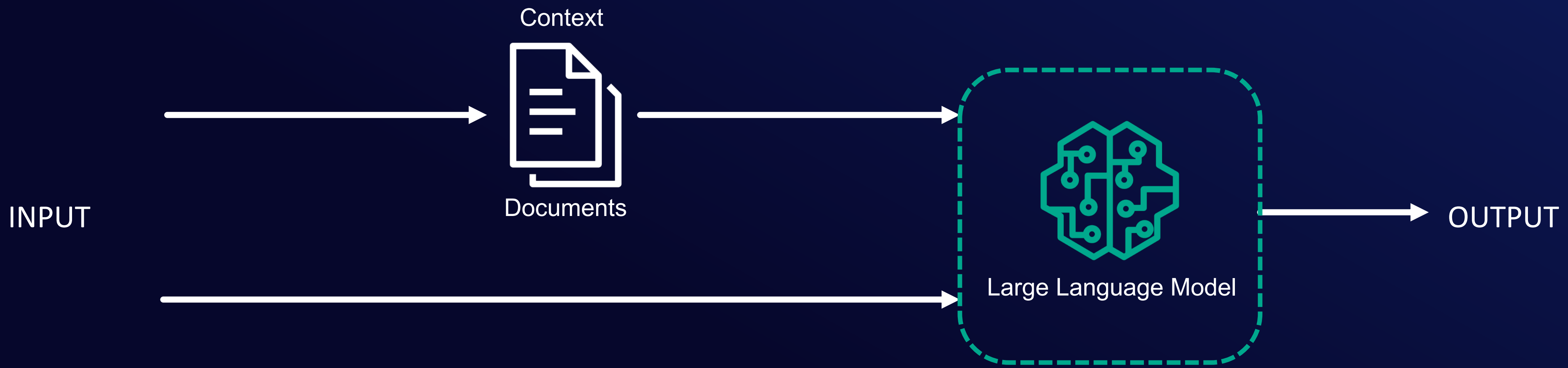
`Boris Johnson`

(Example from Deeplearning.ai course)

Retrieval-Augmented Generation (RAG) – High-level

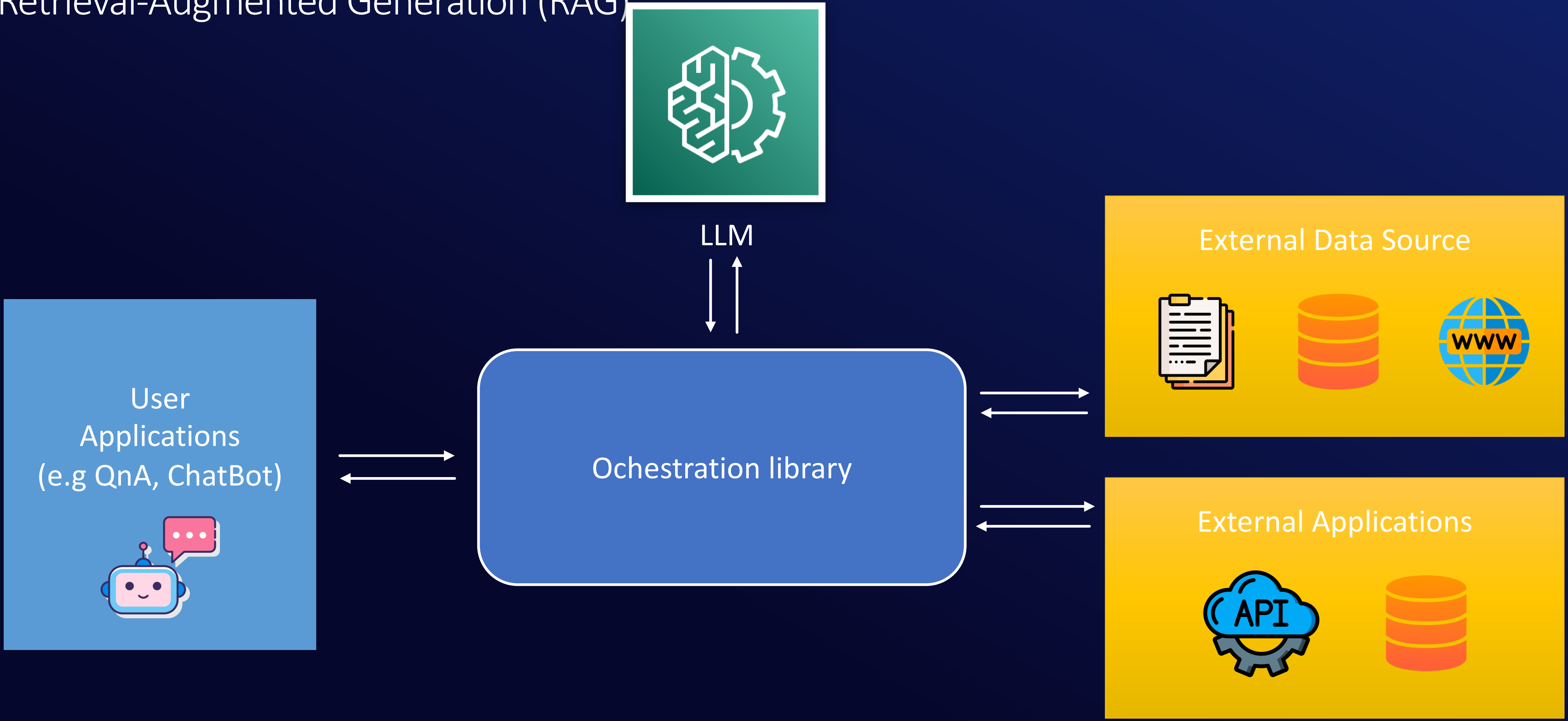
Benefits

- Overcome knowledge cut-off
- Enhanced Contextual Understanding
- Improved Factuality (Less Hallucinations)
- Domain Adaptability



https://arxiv.org/pdf/2005.11401.pdf

LLM-powered applications with Retrieval-Augmented Generation (RAG)



LLM

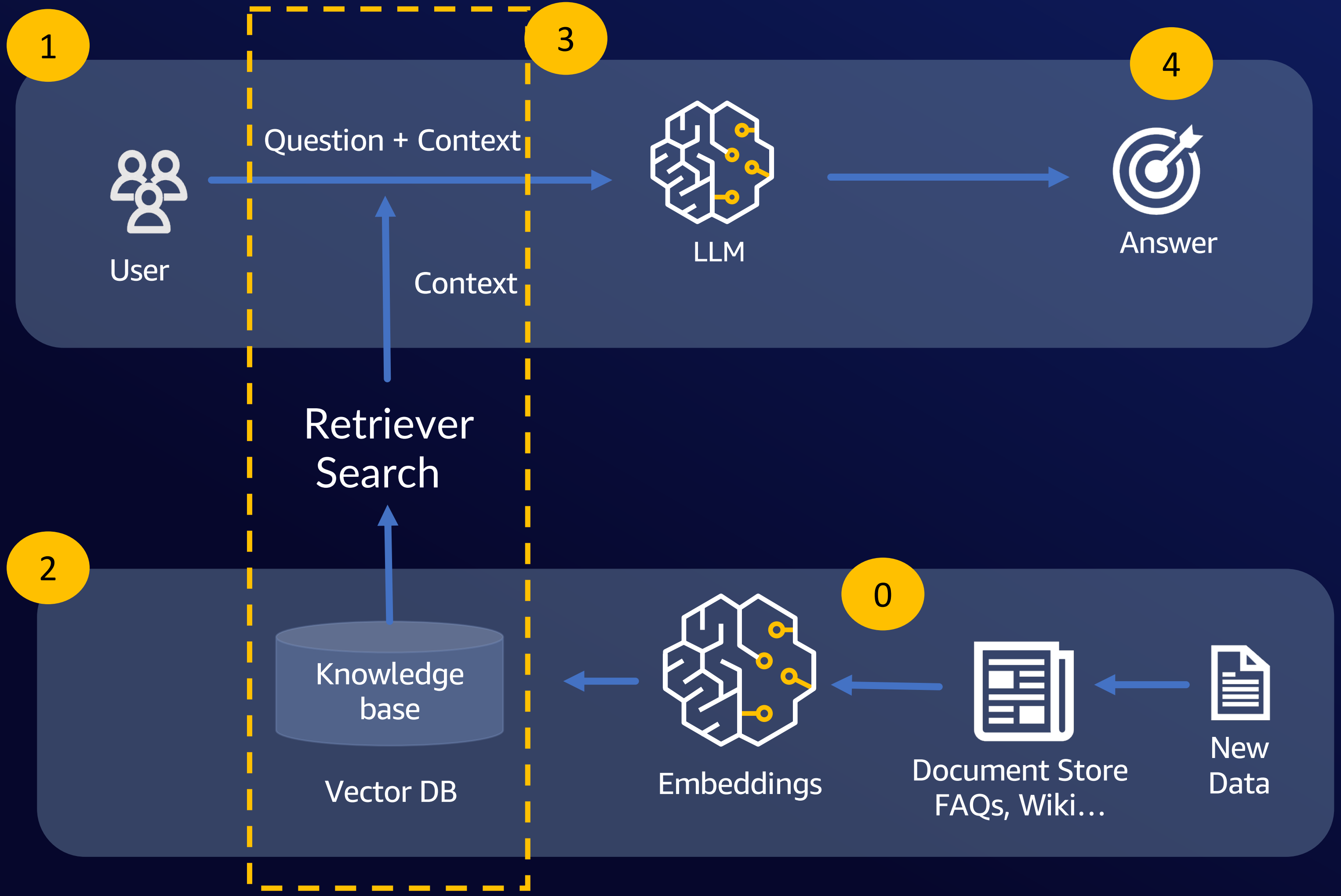
User Applications (e.g QnA, ChatBot)

Ochestration library

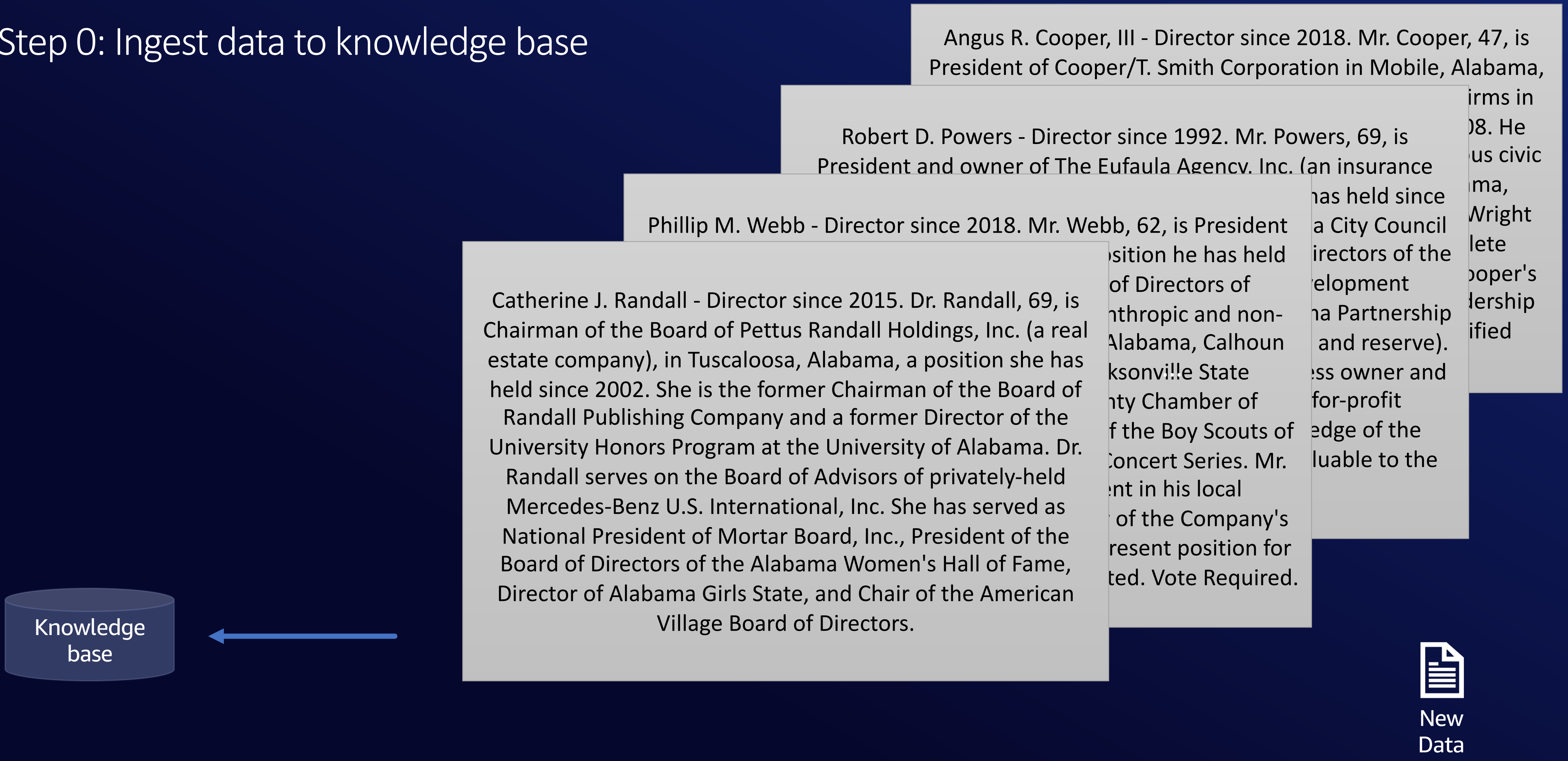
External Data Source

External Applications

5. Retrieval Augmented Generation (RAG)



Step 0: Ingest data to knowledge base



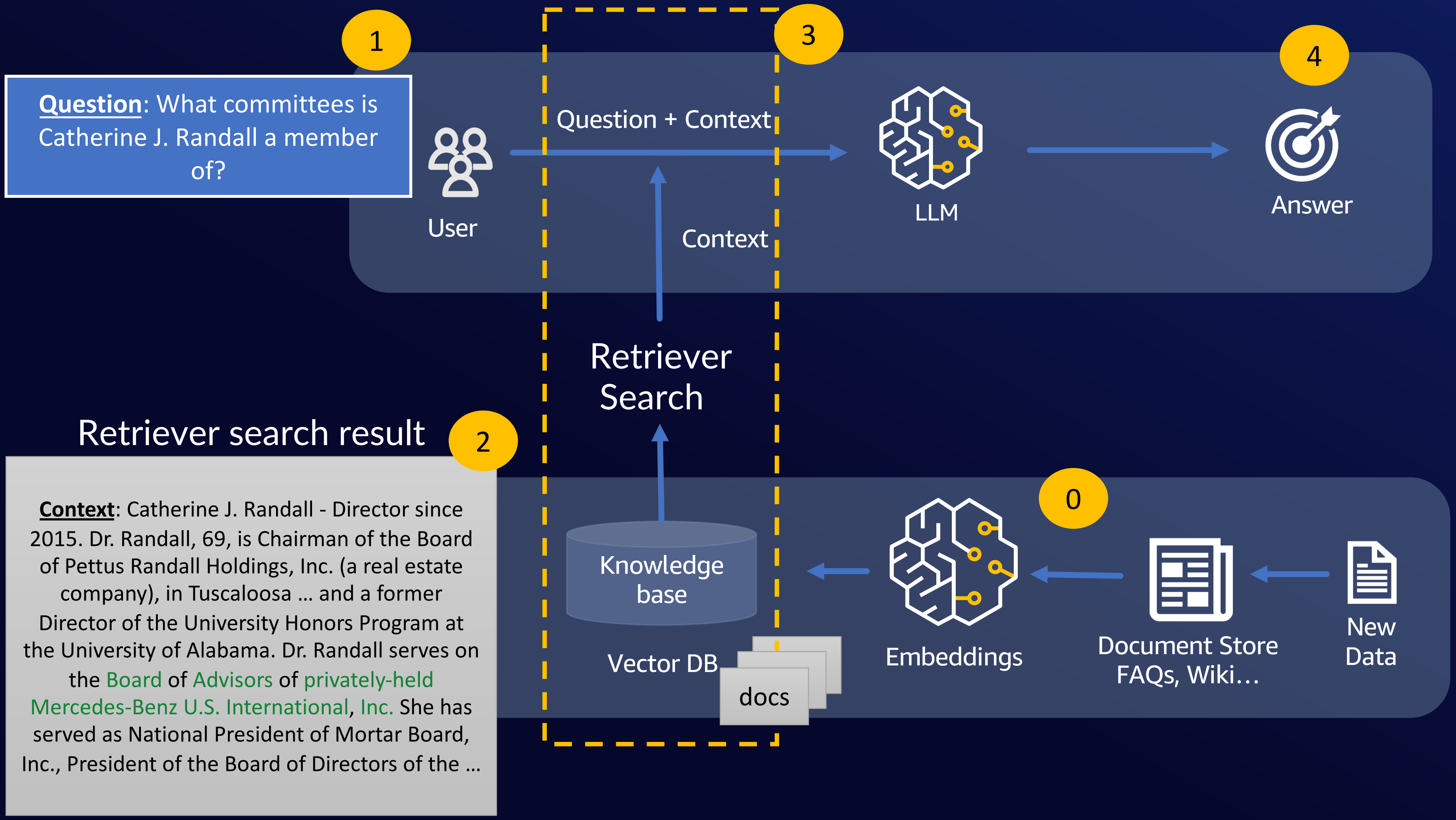
Angus R. Cooper, III - Director since 2018. Mr. Cooper, 47, is
President of Cooper/T. Smith Corporation in Mobile, Alabama,
irms in
8. He
us civic
ma,
Wright
lete
oper's
dership
ified

Robert D. Powers - Director since 1992. Mr. Powers, 69, is
President and owner of The Eufaula Agency, Inc. (an insurance
as held since
a City Council
irectors of the
elopment
a Partnership
and reserve).
ss owner and
for-profit
edge of the
luable to the

Phillip M. Webb - Director since 2018. Mr. Webb, 62, is President
sition he has held
of Directors of
nthropic and non-
Alabama, Calhoun
ksonville State
nty Chamber of
f the Boy Scouts of
Concert Series. Mr.
nt in his local
of the Company's
resent position for
ted. Vote Required.

Catherine J. Randall - Director since 2015. Dr. Randall, 69, is Chairman of the Board of Pettus Randall Holdings, Inc. (a real estate company), in Tuscaloosa, Alabama, a position she has held since 2002. She is the former Chairman of the Board of Randall Publishing Company and a former Director of the University Honors Program at the University of Alabama. Dr. Randall serves on the Board of Advisors of privately-held Mercedes-Benz U.S. International, Inc. She has served as National President of Mortar Board, Inc., President of the Board of Directors of the Alabama Women's Hall of Fame, Director of Alabama Girls State, and Chair of the American Village Board of Directors.

Retrieval Augmented Generation (RAG) - Flow



Question: What committees is Catherine J. Randall a member of?

Context: Catherine J. Randall - Director since 2015. Dr. Randall, 69, is Chairman of the Board of Pettus Randall Holdings, Inc. (a real estate company), in Tuscaloosa … and a former Director of the University Honors Program at the University of Alabama. Dr. Randall serves on the Board of Advisors of privately-held Mercedes-Benz U.S. International, Inc. She has served as National President of Mortar Board, Inc., President of the Board of Directors of the …

Step 3 and 4: LLM Prompt and Completion

3

LLM Prompt =

System prompt: "Use the following…"

+

Context: "Catherine J. Randall - Director since 2015…"

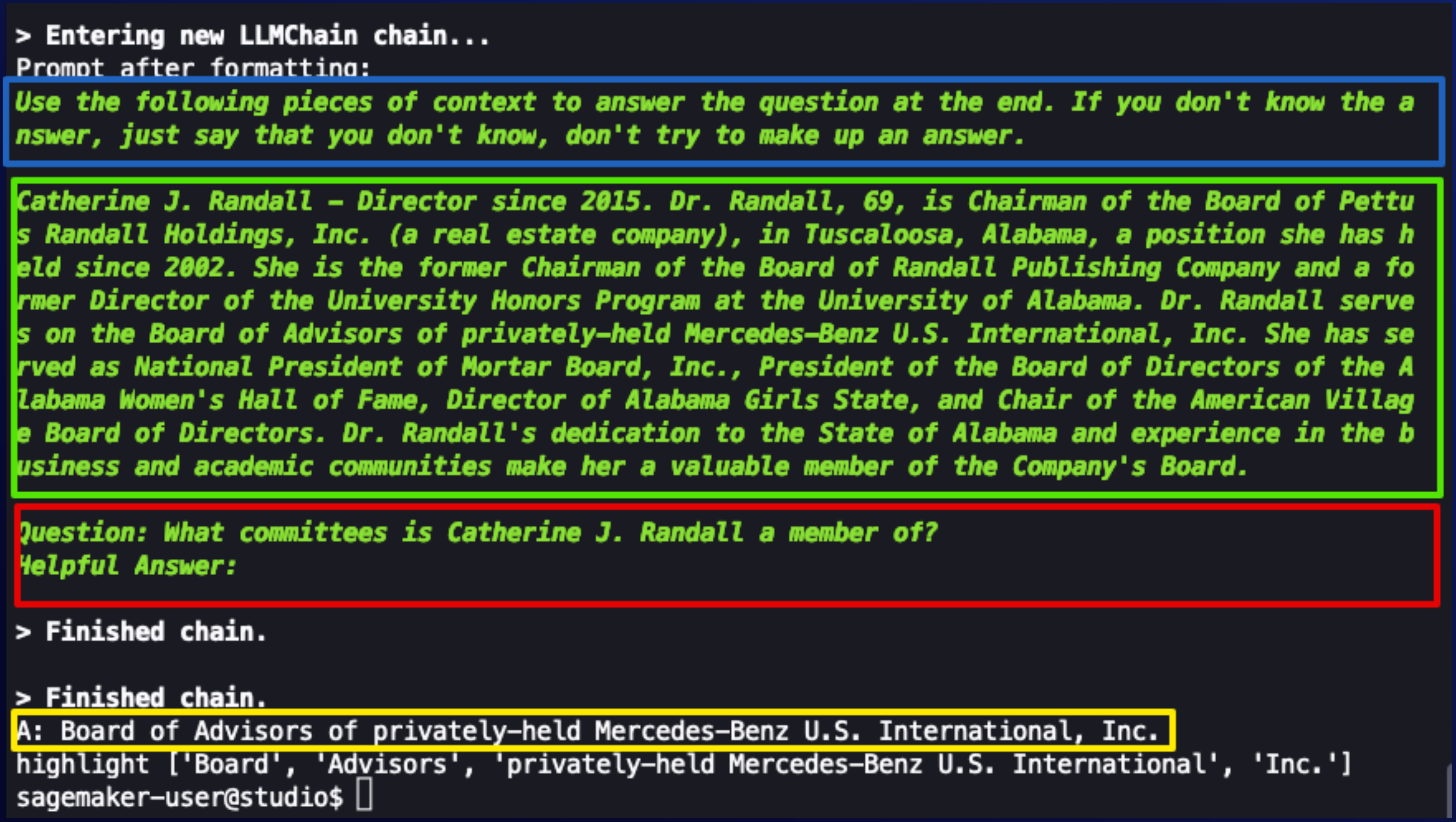
+

Question: "What committees is Catherine J. Randall a member of?"

4

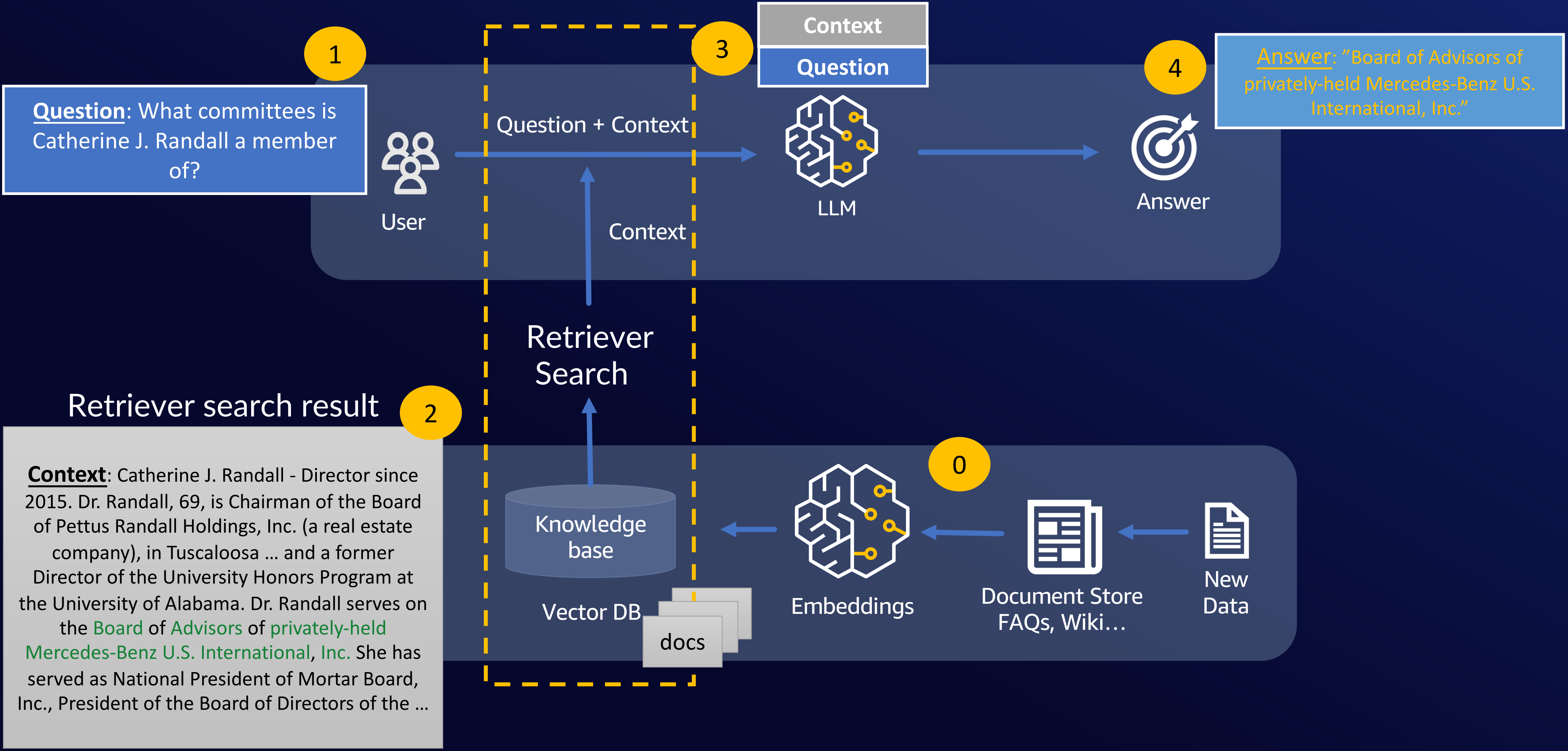
LLM Completion =

"Board of Advisors of privately-held Mercedes-Benz U.S. International, Inc."



```
> Entering new LLMChain chain...
Prompt after formatting:
Use the following pieces of context to answer the question at the end. If you don't know the a
nswer, just say that you don't know, don't try to make up an answer.

Catherine J. Randall – Director since 2015. Dr. Randall, 69, is Chairman of the Board of Pettu
s Randall Holdings, Inc. (a real estate company), in Tuscaloosa, Alabama, a position she has h
eld since 2002. She is the former Chairman of the Board of Randall Publishing Company and a fo
rmer Director of the University Honors Program at the University of Alabama. Dr. Randall serve
s on the Board of Advisors of privately-held Mercedes-Benz U.S. International, Inc. She has se
rved as National President of Mortar Board, Inc., President of the Board of Directors of the A
labama Women's Hall of Fame, Director of Alabama Girls State, and Chair of the American Villag
e Board of Directors. Dr. Randall's dedication to the State of Alabama and experience in the b
usiness and academic communities make her a valuable member of the Company's Board.

Question: What committees is Catherine J. Randall a member of?
Helpful Answer:

> Finished chain.

> Finished chain.
A: Board of Advisors of privately-held Mercedes-Benz U.S. International, Inc.
highlight ['Board', 'Advisors', 'privately-held Mercedes-Benz U.S. International', 'Inc.']
sagemaker-user@studio$ 
```

Retrieval Augmented Generation (RAG) - Flow

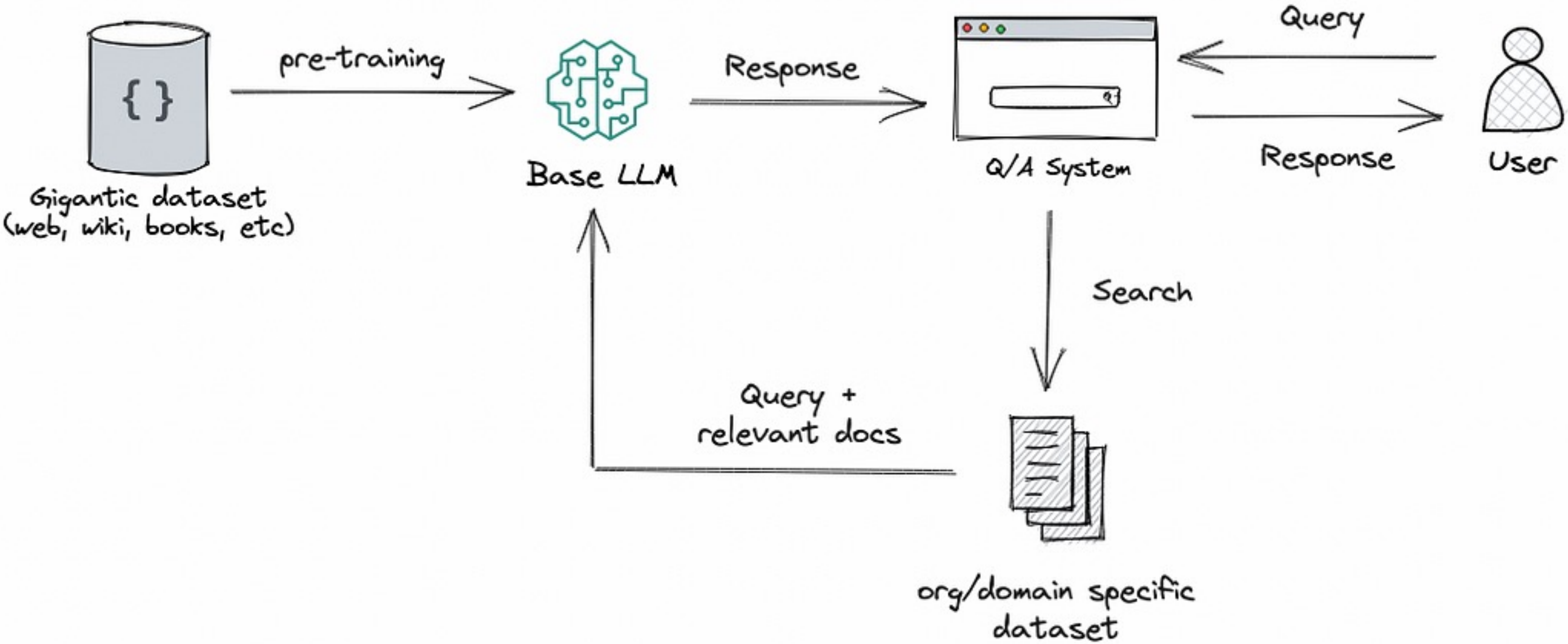


Question: What committees is Catherine J. Randall a member of?

Retriever search result

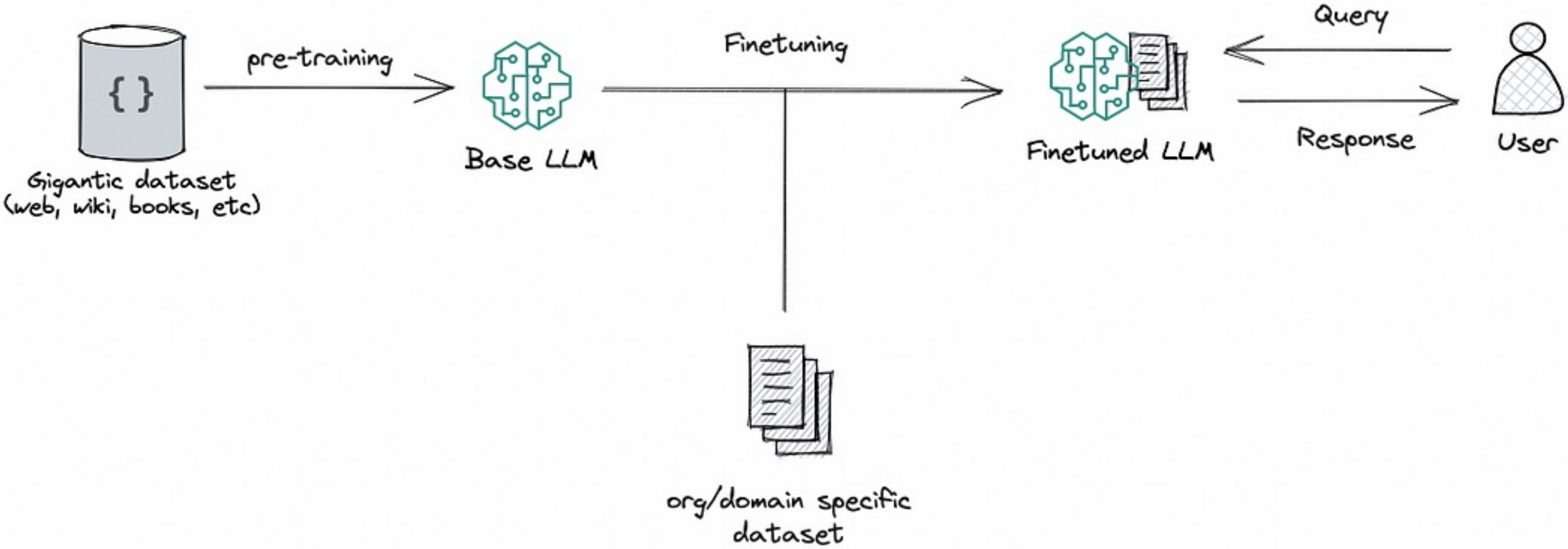
Context: Catherine J. Randall - Director since 2015. Dr. Randall, 69, is Chairman of the Board of Pettus Randall Holdings, Inc. (a real estate company), in Tuscaloosa … and a former Director of the University Honors Program at the University of Alabama. Dr. Randall serves on the Board of Advisors of privately-held Mercedes-Benz U.S. International, Inc. She has served as National President of Mortar Board, Inc., President of the Board of Directors of the …

Answer: "Board of Advisors of privately-held Mercedes-Benz U.S. International, Inc."



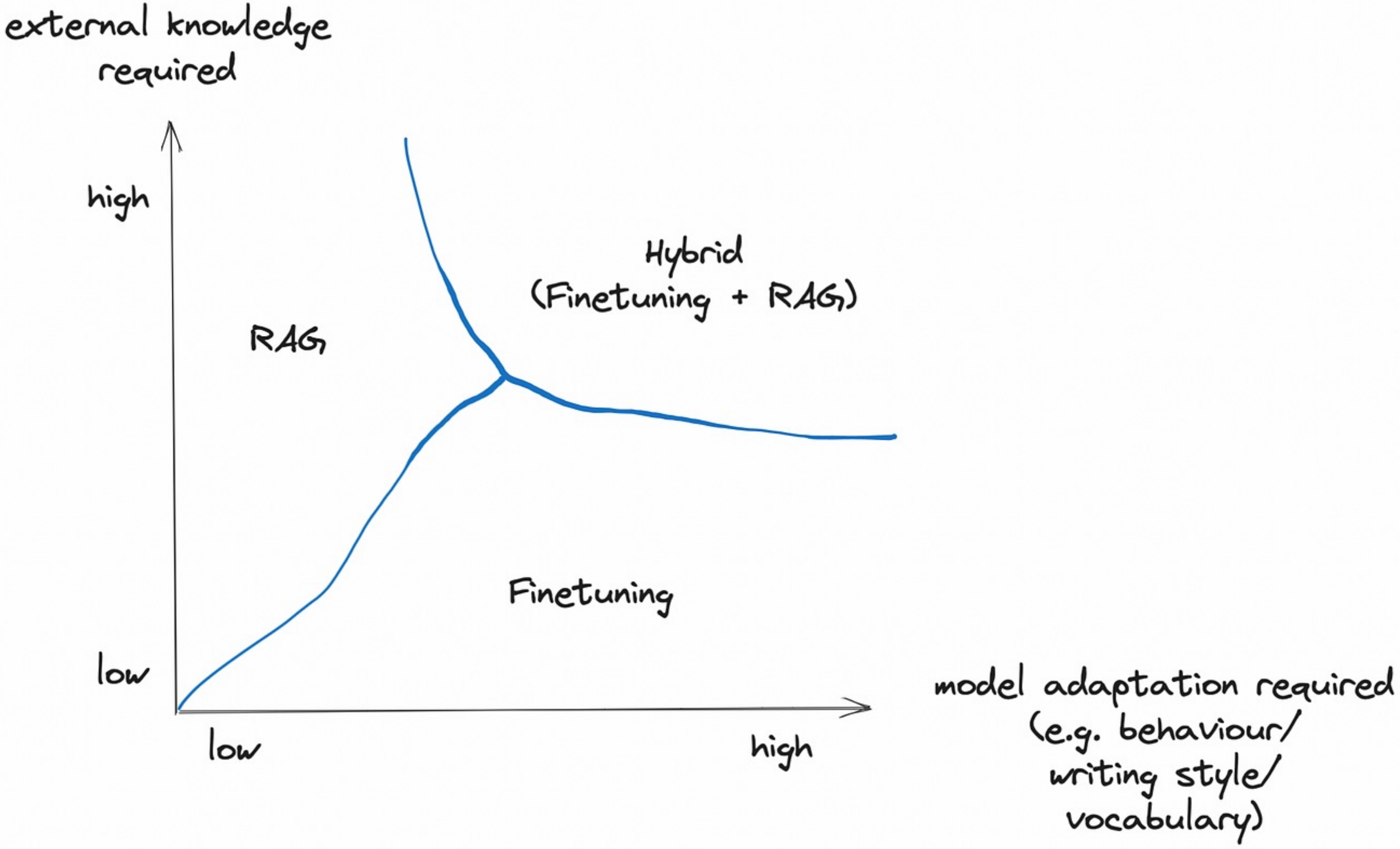
RAG helps the model to “look up” external information to improve its responses.

Source: https://towardsdatascience.com/rag-vs-finetuning-which-is-the-best-tool-to-boost-your-llm-application-94654b1eaba7



Finetuning adjusting the model’s weights based on our data, making it more tailored to our unique needs

Source: https://towardsdatascience.com/rag-vs-finetuning-which-is-the-best-tool-to-boost-your-llm-application-94654b1eaba7


external knowledge required
high
low
RAG
Hybrid (Finetuning + RAG)
Finetuning
low
high
model adaptation required (e.g. behaviour/ writing style/ vocabulary)

Fine-tuning vs RAG

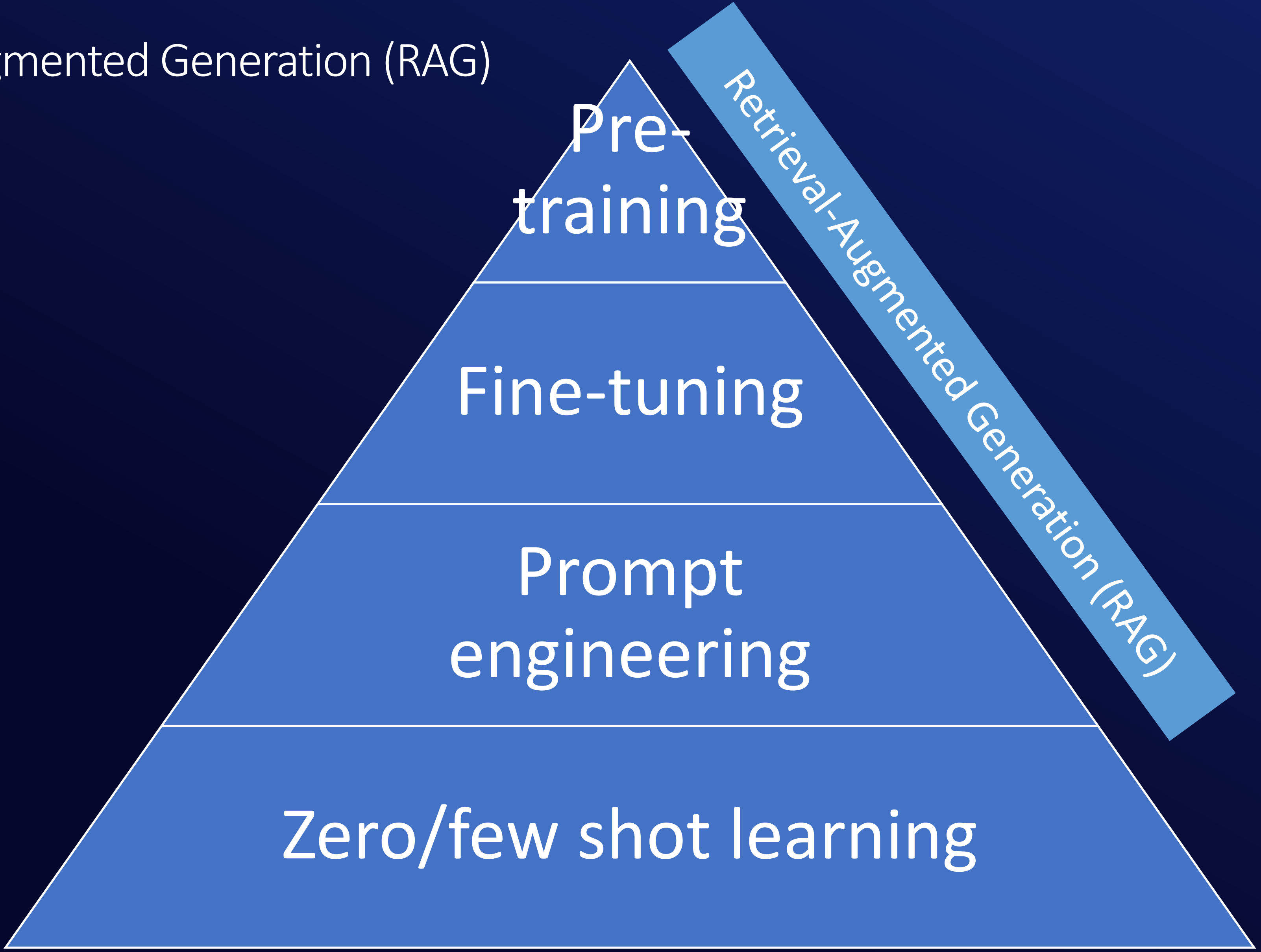
Fine-Tuning (or Retraining)

- Need to train the LLM to perform a different task e.g., generate code
- Expensive to collect training data and train the model
- Pro data science skills to optimize the objective function

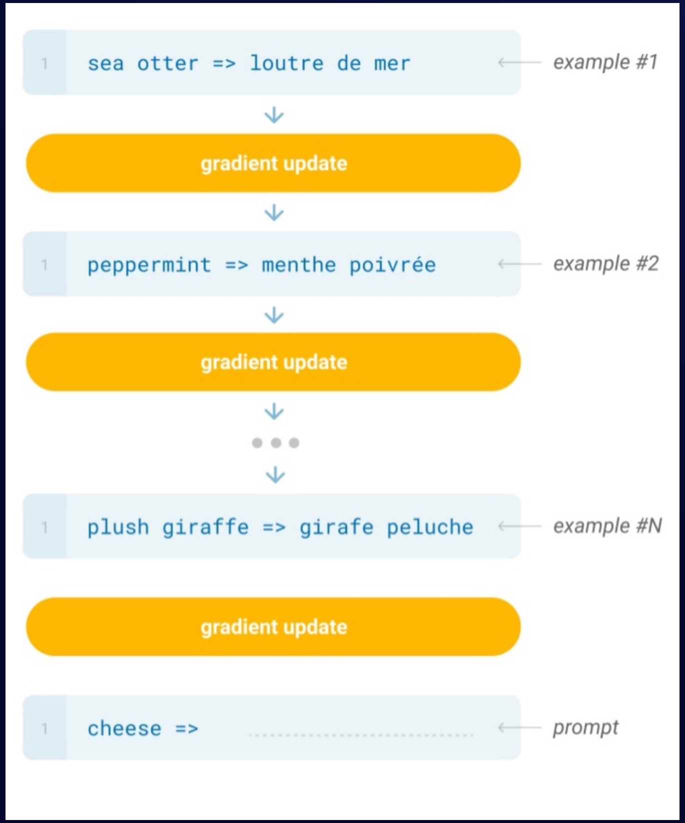
Retrieval Augmented Generation (RAG)

- Adapt the model to a new domain
- Constrain the model to existing knowledge base
- Cost-efficient as compared to fine-tuning
- Low-code / no-code skills

Retrieval-Augmented Generation (RAG)



6. Fine-tuning

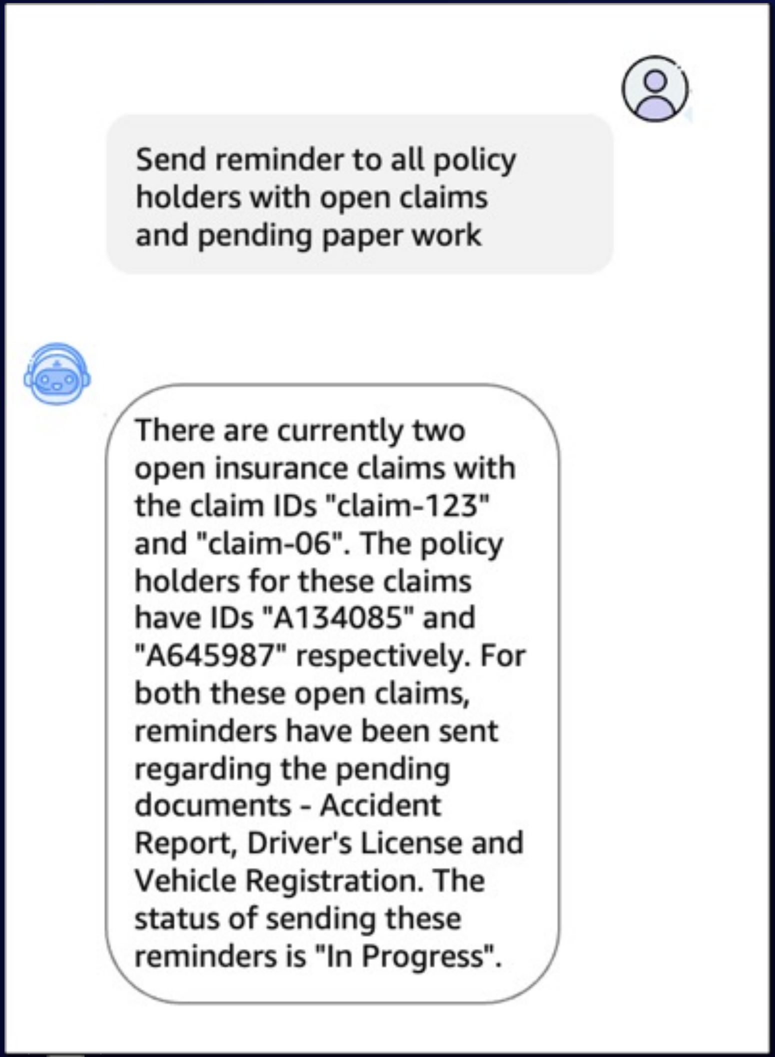


Take a pre-trained LLM and further training it on a smaller, specific dataset to adapt it for a particular task or to improve its performance.

- A lot more accessible through:
  - Parameter Efficient Fine Tuning (PEFT)
  - Low-precision inference.
  - Reinforcement Learning Human Feedback (RLHF)

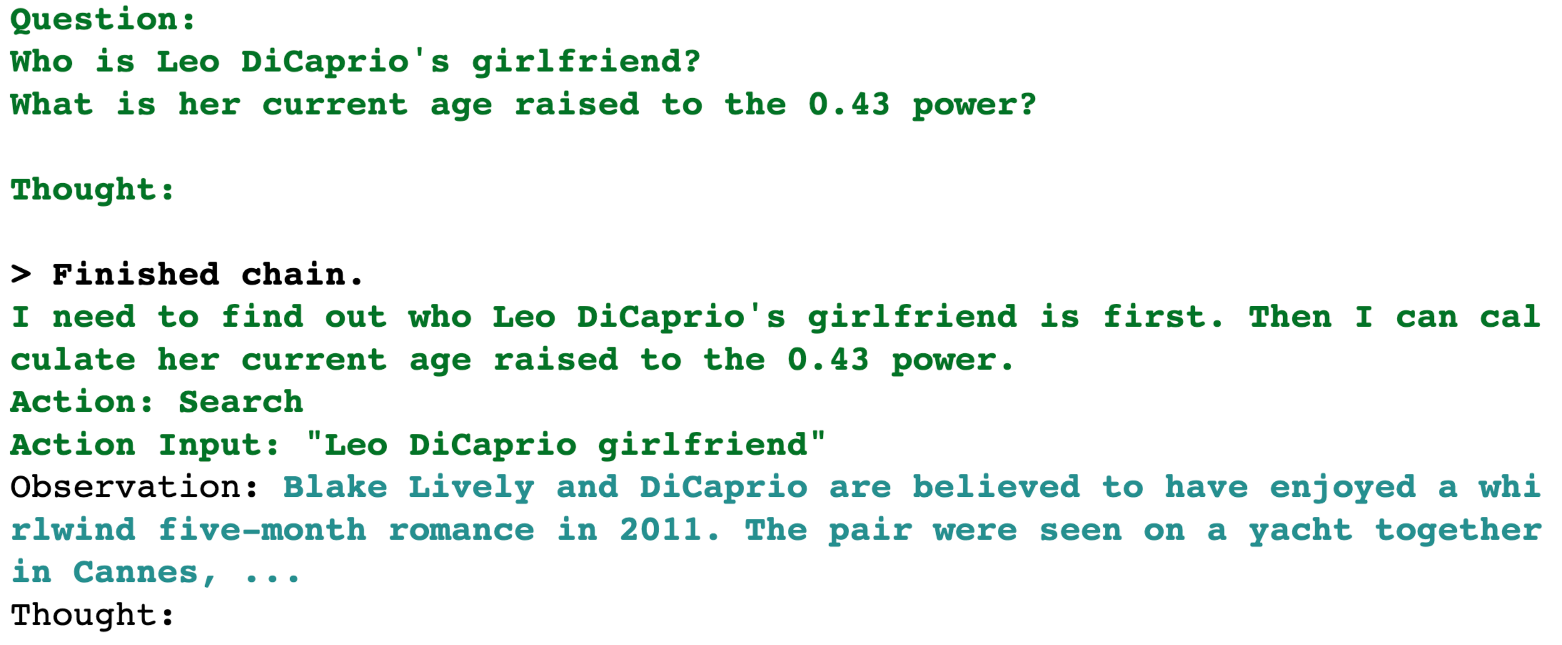
- Keep in mind:
  - More technical expertise
  - Data pipelines and more interactions
  - catastrophic forgetting

GenAI Agent – More than a chat bot



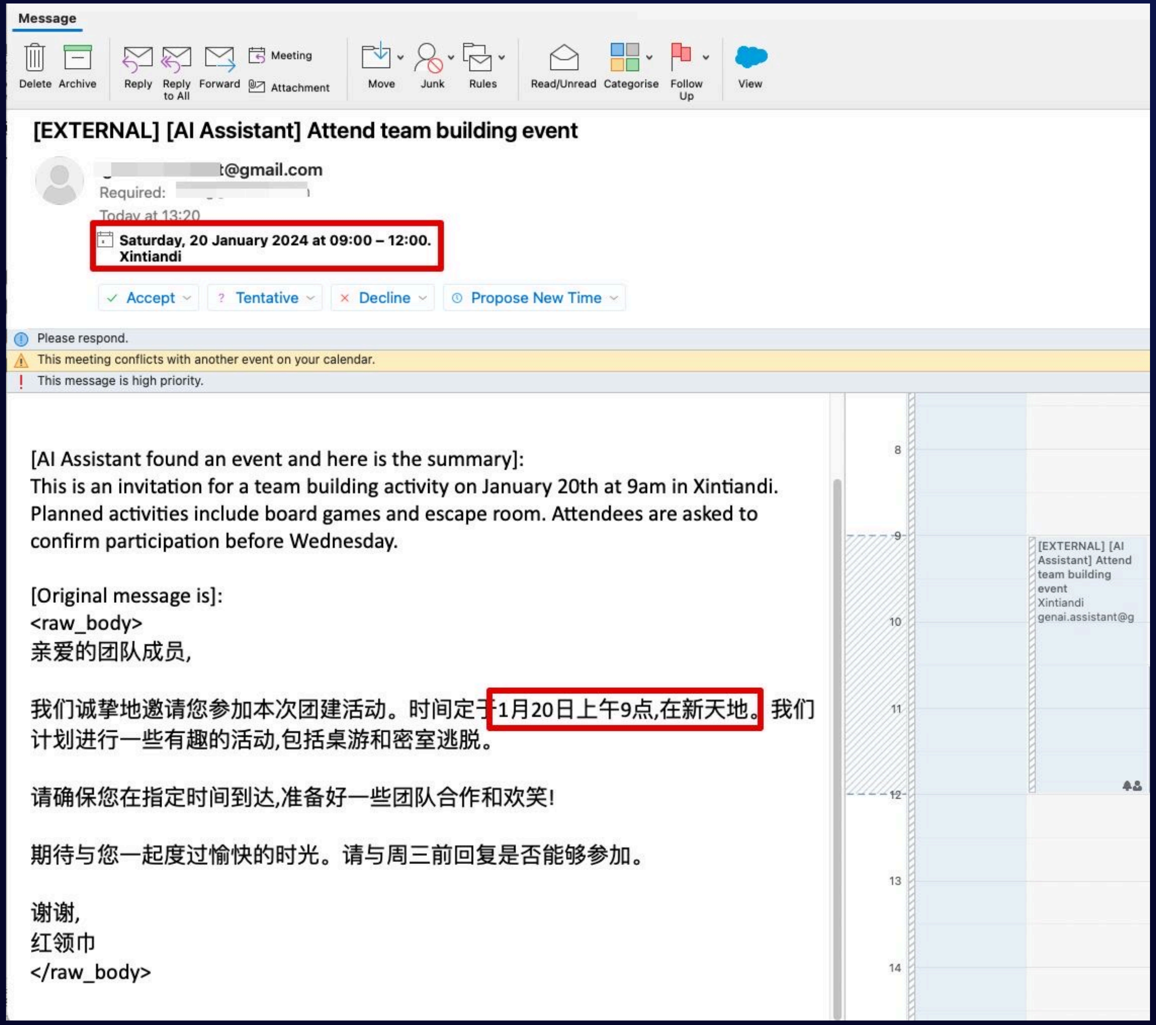
Send reminder to all policy holders with open claims and pending paper work

There are currently two open insurance claims with the claim IDs "claim-123" and "claim-06". The policy holders for these claims have IDs "A134085" and "A645987" respectively. For both these open claims, reminders have been sent regarding the pending documents - Accident Report, Driver's License and Vehicle Registration. The status of sending these reminders is "In Progress".



```
Question:
Who is Leo DiCaprio's girlfriend?
What is her current age raised to the 0.43 power?

Thought:

> Finished chain.
I need to find out who Leo DiCaprio's girlfriend is first. Then I can cal
culate her current age raised to the 0.43 power.
Action: Search
Action Input: "Leo DiCaprio girlfriend"
Observation: Blake Lively and DiCaprio are believed to have enjoyed a whi
rlwind five-month romance in 2011. The pair were seen on a yacht together
in Cannes, ...
Thought:
```

GenAI Agents – More than a chat bot



[EXTERNAL] [AI Assistant] Attend team building event

[redacted]t@gmail.com
Required: [redacted]
Today at 13:20

Saturday, 20 January 2024 at 09:00 – 12:00.
Xintiandi

Please respond.
This meeting conflicts with another event on your calendar.
This message is high priority.

[AI Assistant found an event and here is the summary]:
This is an invitation for a team building activity on January 20th at 9am in Xintiandi. Planned activities include board games and escape room. Attendees are asked to confirm participation before Wednesday.

[Original message is]:
<raw_body>
亲爱的团队成员,

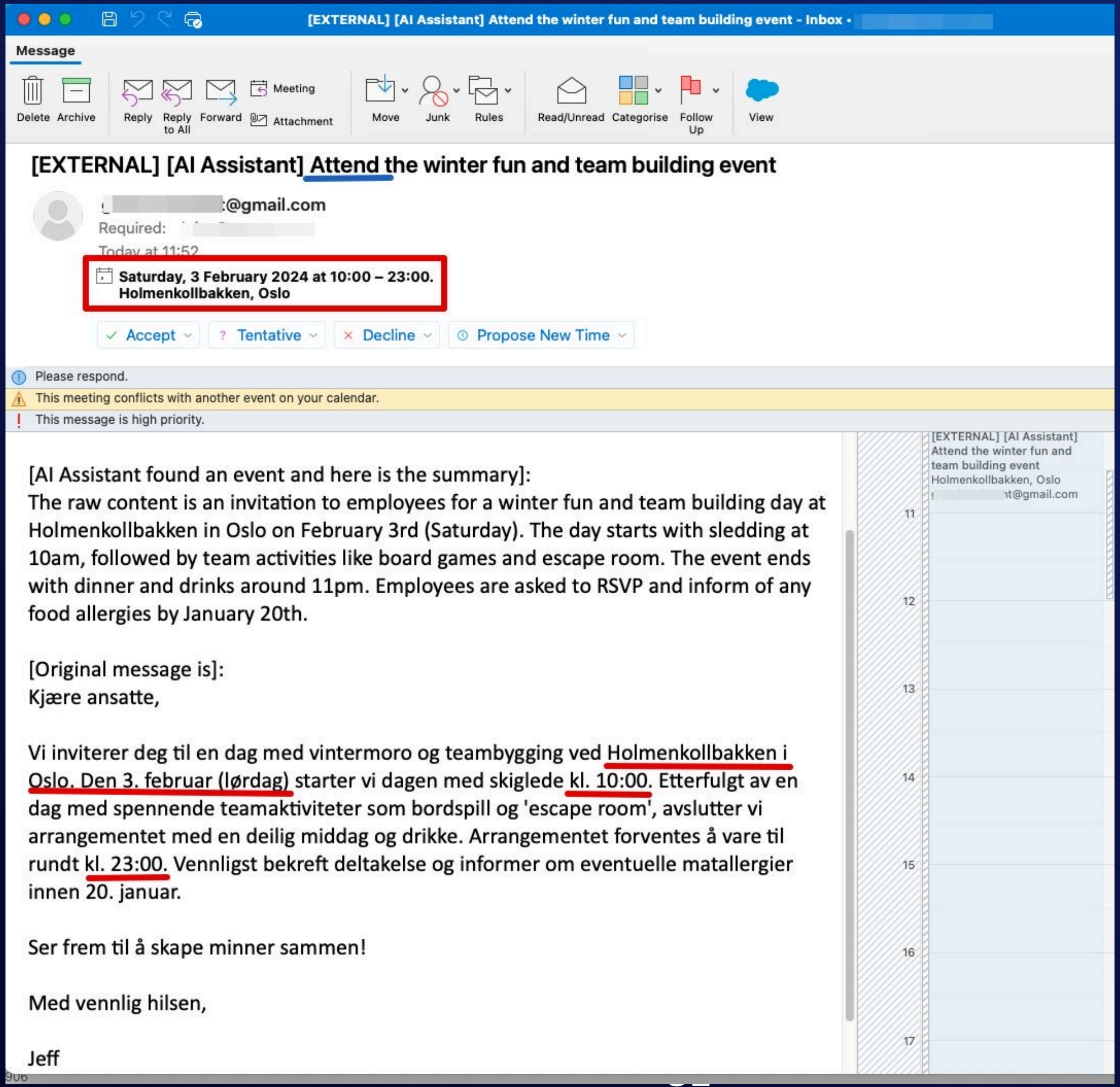
我们诚挚地邀请您参加本次团建活动。时间定于1月20日上午9点,在新天地。我们计划进行一些有趣的活动,包括桌游和密室逃脱。

请确保您在指定时间到达,准备好一些团队合作和欢笑!

期待与您一起度过愉快的时光。请与周三前回复是否能够参加。

谢谢,
红领巾
</raw_body>

[EXTERNAL] [AI Assistant] Attend team building event
Xintiandi
genai.assistant@g



[EXTERNAL] [AI Assistant] Attend the winter fun and team building event - Inbox • [redacted]

[EXTERNAL] [AI Assistant] Attend the winter fun and team building event

[redacted]:@gmail.com
Required: [redacted]
Today at 11:52

Saturday, 3 February 2024 at 10:00 – 23:00.
Holmenkollbakken, Oslo

Please respond.
This meeting conflicts with another event on your calendar.
This message is high priority.

[AI Assistant found an event and here is the summary]:
The raw content is an invitation to employees for a winter fun and team building day at Holmenkollbakken in Oslo on February 3rd (Saturday). The day starts with sledding at 10am, followed by team activities like board games and escape room. The event ends with dinner and drinks around 11pm. Employees are asked to RSVP and inform of any food allergies by January 20th.

[Original message is]:
Kjære ansatte,

Vi inviterer deg til en dag med vintermoro og teambygging ved Holmenkollbakken i Oslo. Den 3. februar (lørdag) starter vi dagen med skiglede kl. 10:00. Etterfulgt av en dag med spennende teamaktiviteter som bordspill og 'escape room', avslutter vi arrangementet med en deilig middag og drikke. Arrangementet forventes å vare til rundt kl. 23:00. Vennligst bekreft deltakelse og informer om eventuelle matallergier innen 20. januar.

Ser frem til å skape minner sammen!

Med vennlig hilsen,

Jeff

[EXTERNAL] [AI Assistant] Attend the winter fun and team building event
Holmenkollbakken, Oslo
[redacted]nt@gmail.com

Agenda

- Motivation
- Basic concepts
- Opportunities & Challenges
- Use case and patterns
- Tools and resources ⇦

PartyRock

https://partyrock.aws/

Copilot

Your everyday AI companion

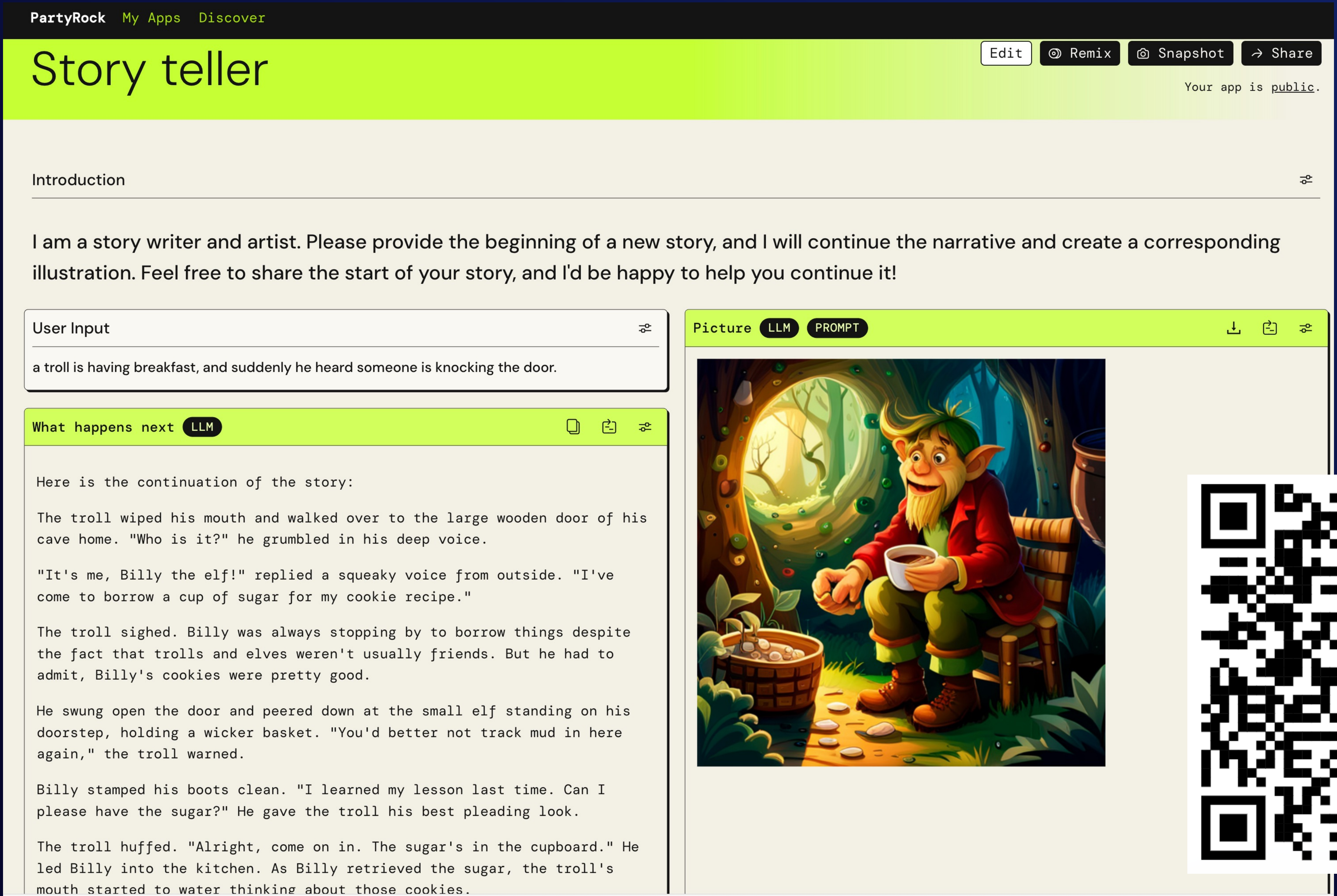
Edge Browser

豆包

https://www.doubao.com/

OpenAI

https://chat.openai.com/



Story teller

Your app is public.

Introduction

I am a story writer and artist. Please provide the beginning of a new story, and I will continue the narrative and create a corresponding illustration. Feel free to share the start of your story, and I'd be happy to help you continue it!

User Input

a troll is having breakfast, and suddenly he heard someone is knocking the door.

What happens next LLM

Here is the continuation of the story:

The troll wiped his mouth and walked over to the large wooden door of his cave home. "Who is it?" he grumbled in his deep voice.

"It's me, Billy the elf!" replied a squeaky voice from outside. "I've come to borrow a cup of sugar for my cookie recipe."

The troll sighed. Billy was always stopping by to borrow things despite the fact that trolls and elves weren't usually friends. But he had to admit, Billy's cookies were pretty good.

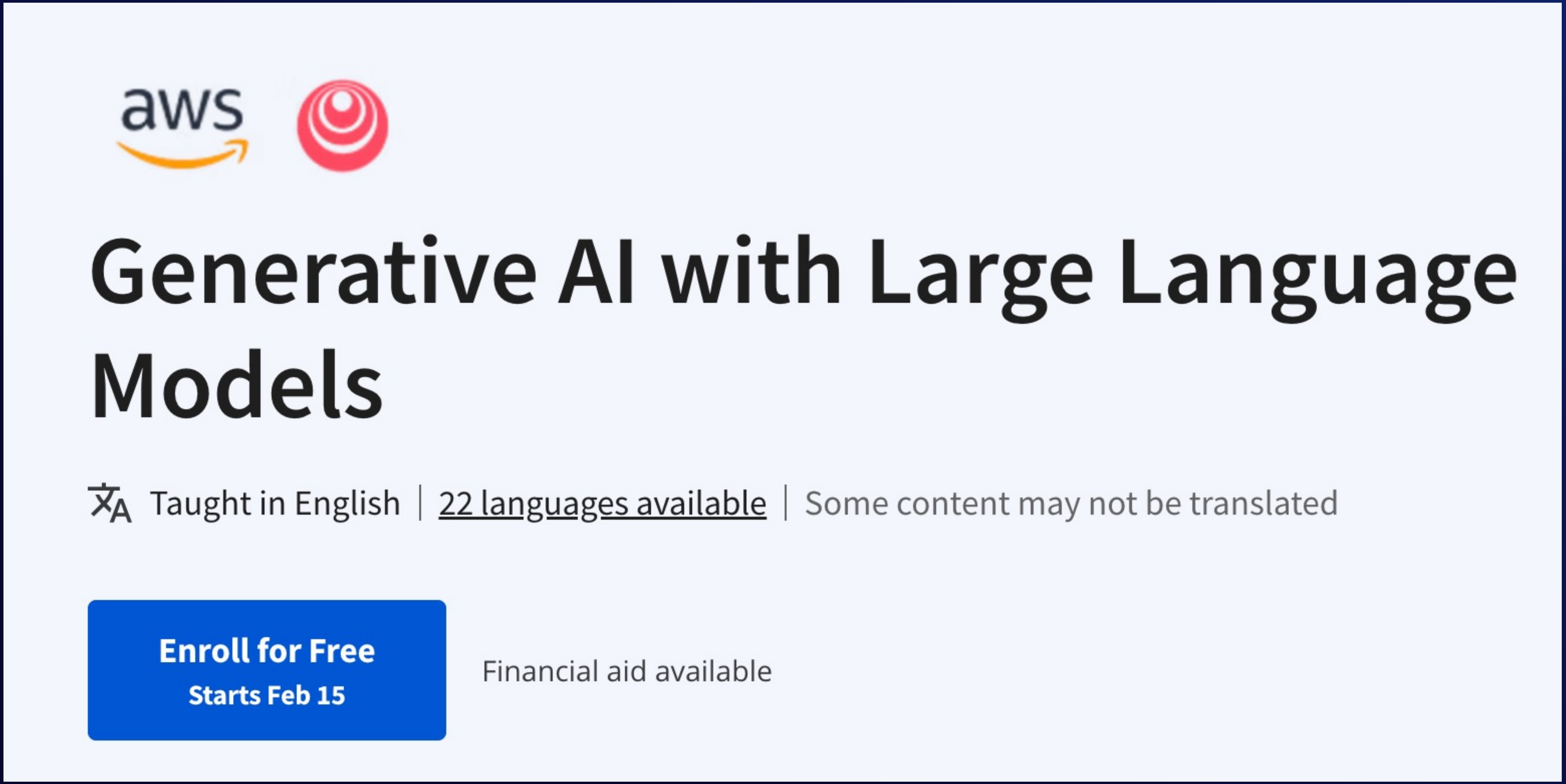
He swung open the door and peered down at the small elf standing on his doorstep, holding a wicker basket. "You'd better not track mud in here again," the troll warned.

Billy stamped his boots clean. "I learned my lesson last time. Can I please have the sugar?" He gave the troll his best pleading look.

The troll huffed. "Alright, come on in. The sugar's in the cupboard." He led Billy into the kitchen. As Billy retrieved the sugar, the troll's mouth started to water thinking about those cookies.

Picture LLM PROMPT

https://partyrock.aws/u/Linkcd/SbaKA50rC/Story-teller

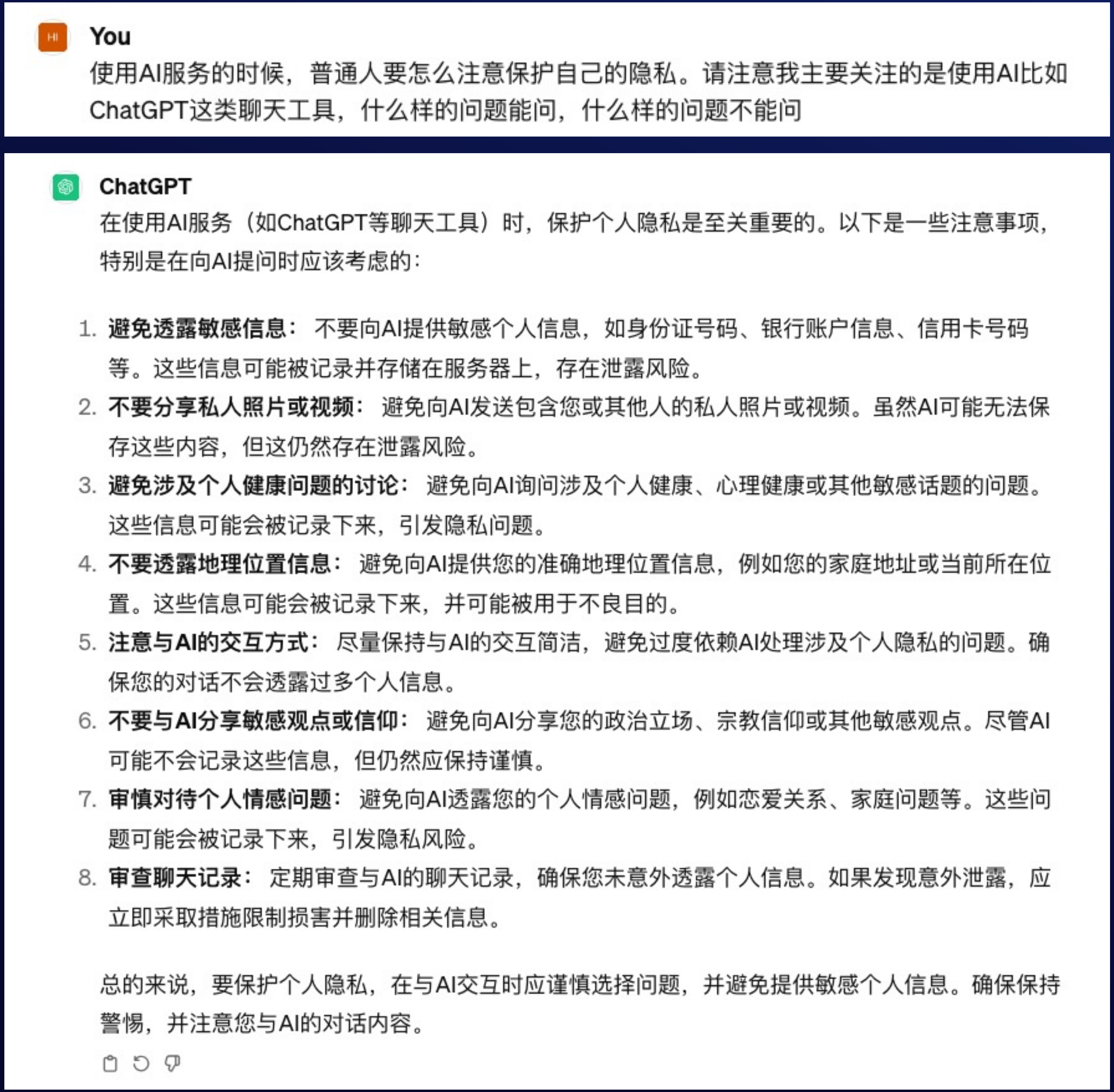


Generative AI with Large Language Models

Taught in English | 22 languages available | Some content may not be translated

Financial aid available

https://www.coursera.org/learn/generative-ai-with-llms



You

使用AI服务的时候，普通人要怎么注意保护自己的隐私。请注意我主要关注的是使用AI比如ChatGPT这类聊天工具，什么样的问题能问，什么样的问题不能问

ChatGPT

在使用AI服务（如ChatGPT等聊天工具）时，保护个人隐私是至关重要的。以下是一些注意事项，特别是在向AI提问时应该考虑的：

1. **避免透露敏感信息**： 不要向AI提供敏感个人信息，如身份证号码、银行账户信息、信用卡号码等。这些信息可能被记录并存储在服务器上，存在泄露风险。
2. **不要分享私人照片或视频**： 避免向AI发送包含您或其他人的私人照片或视频。虽然AI可能无法保存这些内容，但这仍然存在泄露风险。
3. **避免涉及个人健康问题的讨论**： 避免向AI询问涉及个人健康、心理健康或其他敏感话题的问题。这些信息可能会被记录下来，引发隐私问题。
4. **不要透露地理位置信息**： 避免向AI提供您的准确地理位置信息，例如您的家庭地址或当前所在位置。这些信息可能会被记录下来，并可能被用于不良目的。
5. **注意与AI的交互方式**： 尽量保持与AI的交互简洁，避免过度依赖AI处理涉及个人隐私的问题。确保您的对话不会透露过多个人信息。
6. **不要与AI分享敏感观点或信仰**： 避免向AI分享您的政治立场、宗教信仰或其他敏感观点。尽管AI可能不会记录这些信息，但仍然应保持谨慎。
7. **审慎对待个人情感问题**： 避免向AI透露您的个人情感问题，例如恋爱关系、家庭问题等。这些问题可能会被记录下来，引发隐私风险。
8. **审查聊天记录**： 定期审查与AI的聊天记录，确保您未意外透露个人信息。如果发现意外泄露，应立即采取措施限制损害并删除相关信息。

总的来说，要保护个人隐私，在与AI交互时应谨慎选择问题，并避免提供敏感个人信息。确保保持警惕，并注意您与AI的对话内容。

Thank you!