



MELBOURNE EDITION

2024 SPONSORS



vNEXT



WHO AM I?

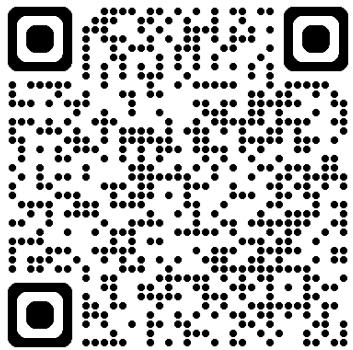
NAME

James Westall

WORK

Arinco

SOCIALS

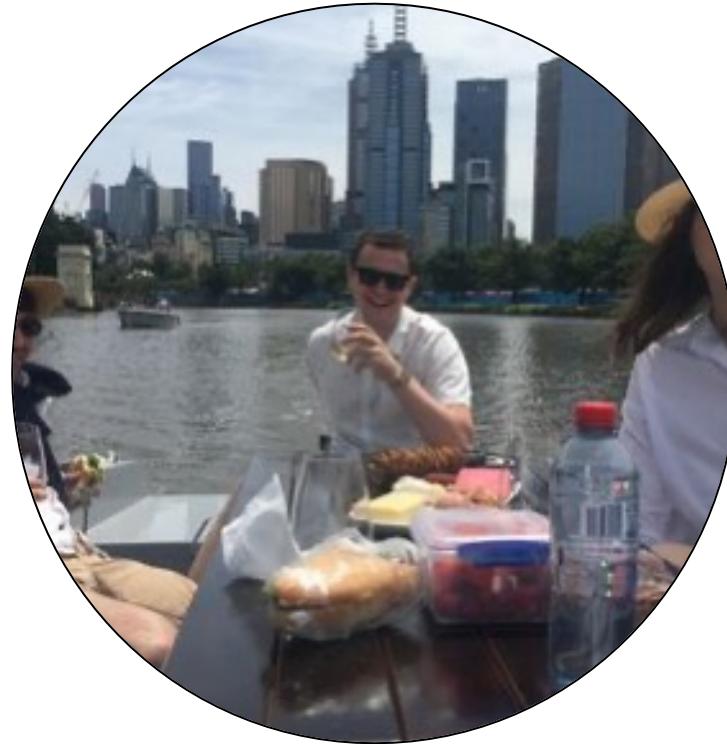


PASSIONS

Zooming down frosty hills
with fibreglass strapped to
my feet.

Delectable treats and eats

Azure



ZERO TO HERO

Prompt Engineering For Dummies



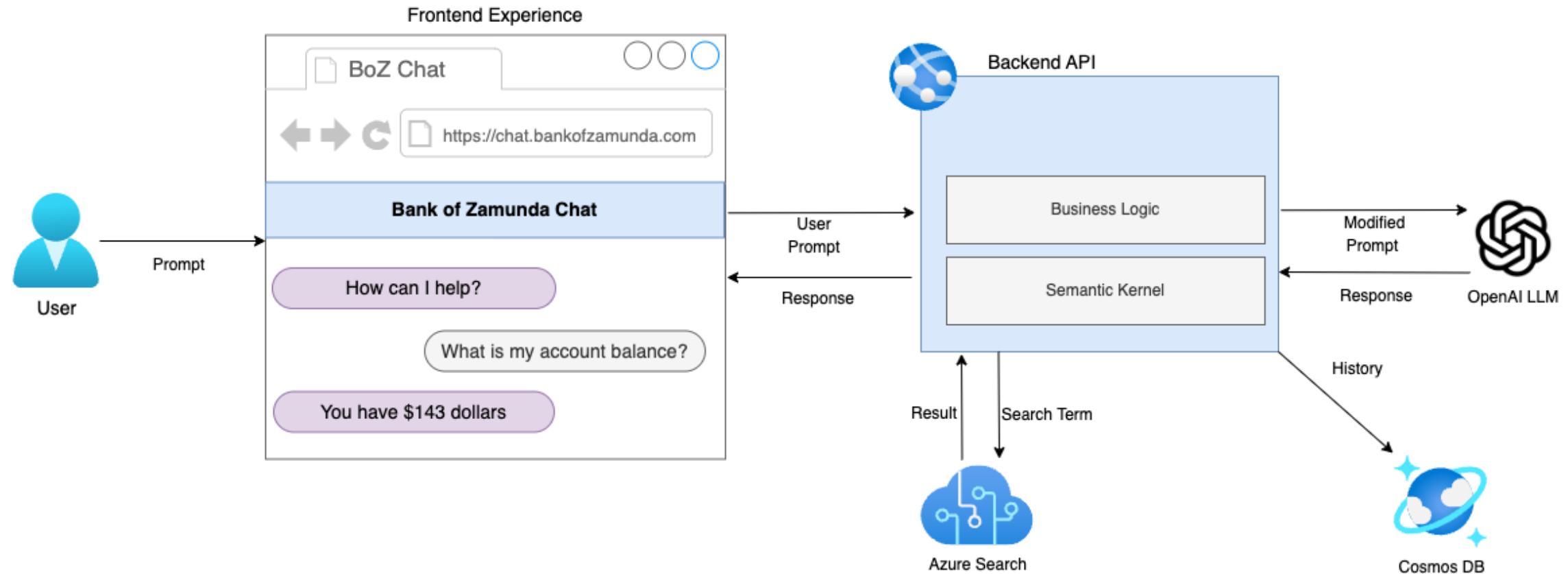
**LARGE LANGUAGE MODELS
JUST WANT TO ANSWER YOU.**

SO PROVIDE SMARTER INPUTS.



That's right! It goes into the square hole

RETRIEVAL AUGMENTED GENERATION (RAG)



UNDERSTANDING TOKENS

- Azure OpenAI is priced based on token usage, in 1000 token blocks.
- Your prompt and response must be less than the model limits.
 - GPT-4 = 8192
 - GPT-3.5 = 4096
 - GPT-3 = 4096
- As a general rule of thumb, 1 token ~ 4 chars in English
- Limits currently do exist for tokens per hour (very high)

Tokens Characters
18 74

How much wood would a wood chuck chuck? If a wood chuck could chuck wood?
?

Tokens Characters
44 242

Arinco brings a unique proposition to today's market. We have made it our specialty in leveraging emerging technologies, modern productivity tools and cloud native platforms to modernise and build the next generation of intelligent solutions.

TEXT TOKEN IDS

TEXT TOKEN IDS



BASIC PROMPT ENGINEERING

Simple & Complex Instructions:

Simple Prompt: Write me a marketing slogan, summarising who Arinco is.

Complex Prompt : Write five marketing slogans, summarising who Arinco is. The slogan should make some reference to Artificial Intelligence and modernisation. Synonyms of these themes are acceptable replacement.

Response: Arinco are cloud native innovators, leveraging AI technology to modernise your ecosystem

Usage Scenario:

Creative license, non restrictive scenarios with assumed GPT knowledge

Primary Content/Grounding.

Prompt: In computer science, the time complexity is the computational complexity that describes the amount of computer time it takes to run an algorithm. Time complexity is commonly estimated by [...]

Summarize the previous text in a few sentences:

Response: Time complexity is a measure of how long an algorithm takes to run. It is typically expressed as a function of the size of the input. The most common measure is worst-case time complexity, which is the maximum amount of time required for inputs of a given size.

Usage Scenario:

Providing organisational input for prompts, ensuring your data is presented to GPT



BASIC PROMPT ENGINEERING

Zero Shot:

Prompt: Headline: Coach confident injury won't derail Warriors

Topic:

Response: The coach is confident that the injury won't derail the Warriors' season. The team is still focused on their goals and that they will continue to work hard to achieve them.

Usage Scenario:

Creative license, non restrictive scenarios

Few-Shot:

Prompt: Headline: Twins' Correa to use opt-out, test free agency

Topic: Baseball

Headline: Qatar World Cup to have zones for sobering up

Topic: Soccer

Headline: Yates: Fantasy football intel for Week 6

Topic: Football

Headline: Coach confident injury won't derail Warriors

Topic:

Response: Basketball

Usage Scenario:

Ensuring expected response format



SYSTEM (META) PROMPTS

Prompt: ArchieBot helps the company employees understand information about the Contoso organisation, product offerings, and questions about the employee handbook. Be brief in your answers.

Instructions.

- If asking a clarifying question to the user would help, ask the question.
- You should decline to respond if a user asks about the following topics: Free Ducks at the park, Violence, Sexual Content, If birds are real
- If you decline to respond you should refer the employee to "hr@contoso.com" for an answer
- For tabular information return it as an html table. Do not return markdown format.

<User Prompt>

Usage Scenario:

Defining complex sets of rules which your LLM must follow.



BRINGING IT ALL TOGETHER

OpenAI GPT Tokens

System Prompt

Grounding Context

User Prompt

Prompt Start

<|im_start|>system

ArchieBot helps the company employees understand information about the Contoso organisation, product offerings, and questions about the employee handbook. Be brief in your answers.

Instructions.

- Answer ONLY with the facts listed in the list of sources below.
- If there isn't enough information below, say you don't know.
- Do not generate answers that don't use the sources below.
- If asking a clarifying question to the user would help, ask the question.
- For tabular information return it as an html table. Do not return markdown format.
- Each source has a name followed by colon and the actual information, always include the source name for each fact you use in the response.
- Use square brackets to reference the source, e.g. [info1.txt].
- Don't combine sources, list each source separately, e.g. [info1.txt][info2.pdf].

Sources:

Contoso-Leave-Policy.pdf: <**REMOVED FOR BREVITY**>

<|im_end|>

<|im_start|>user

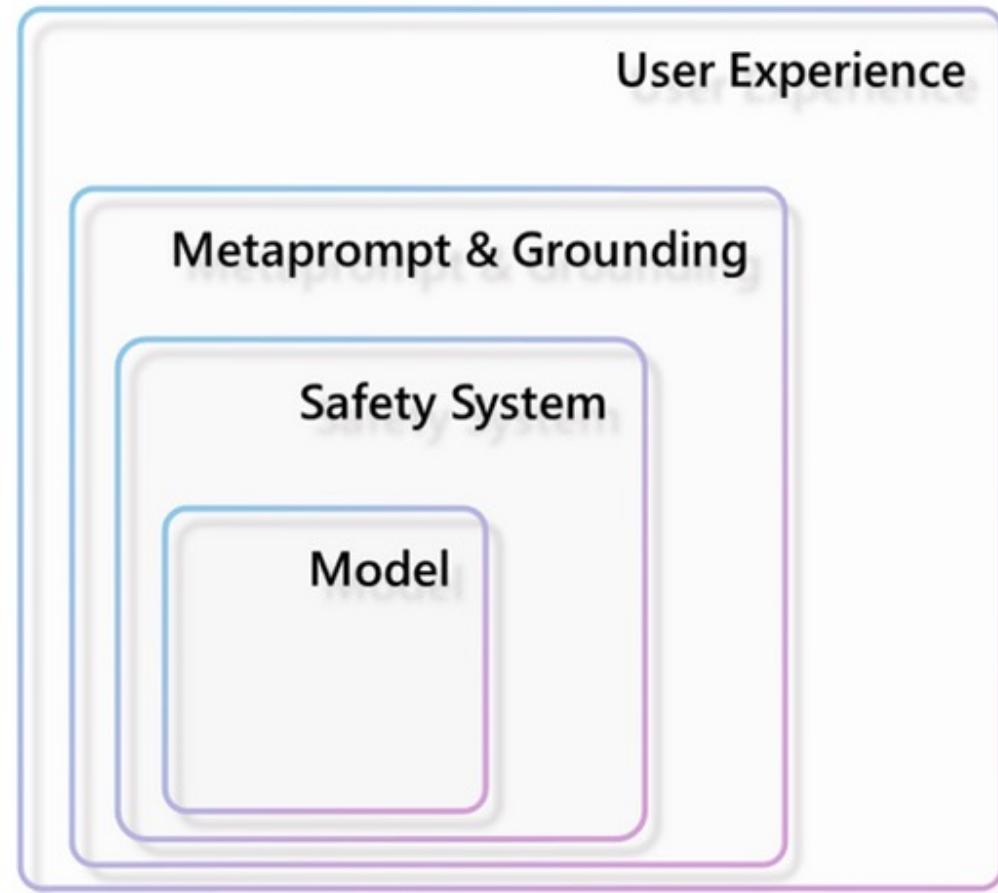
Summarise the Contoso policy on leave requests.

<|im_end|>

<|im_start|>ArchieBot



SECONDARY AREAS OF IMPACT



SHOULD I FINE TUNE MY LLM?

No.*

*For about 999/1000 scenarios.



VECTORS ARE BETTER

Similarity Example - What is an amphibian?

1.000000 What is an amphibian?

0.937651 A frog is an amphibian.

0.902491 Amphibians are four-limbed and ectothermic vertebrates of the class Amphibia.

0.873569 Cos'è un anfibio?

0.866632 Frogs, toads, and salamanders are all examples.

0.857454 A frog is green.

0.842596 They are four-limbed and ectothermic vertebrates.

0.802171 A dog is a mammal.

0.784479 It's not easy bein' green.

0.778341 A tree is green.

0.756669 A dog is a man's best friend.

0.734219 You ain't never had a friend like me.

0.721176 Rachel, Monica, Phoebe, Joey, Chandler, Ross



EVALUATING PROMPTS

1. Measuring prompt performance is hard.
2. In public facing scenarios you should be doing this.

