

# Generative AI in practice: Concrete LLM use cases in Java with PaLM & Gemini

Guillaume Laforge  
Developer Advocate

Google Cloud

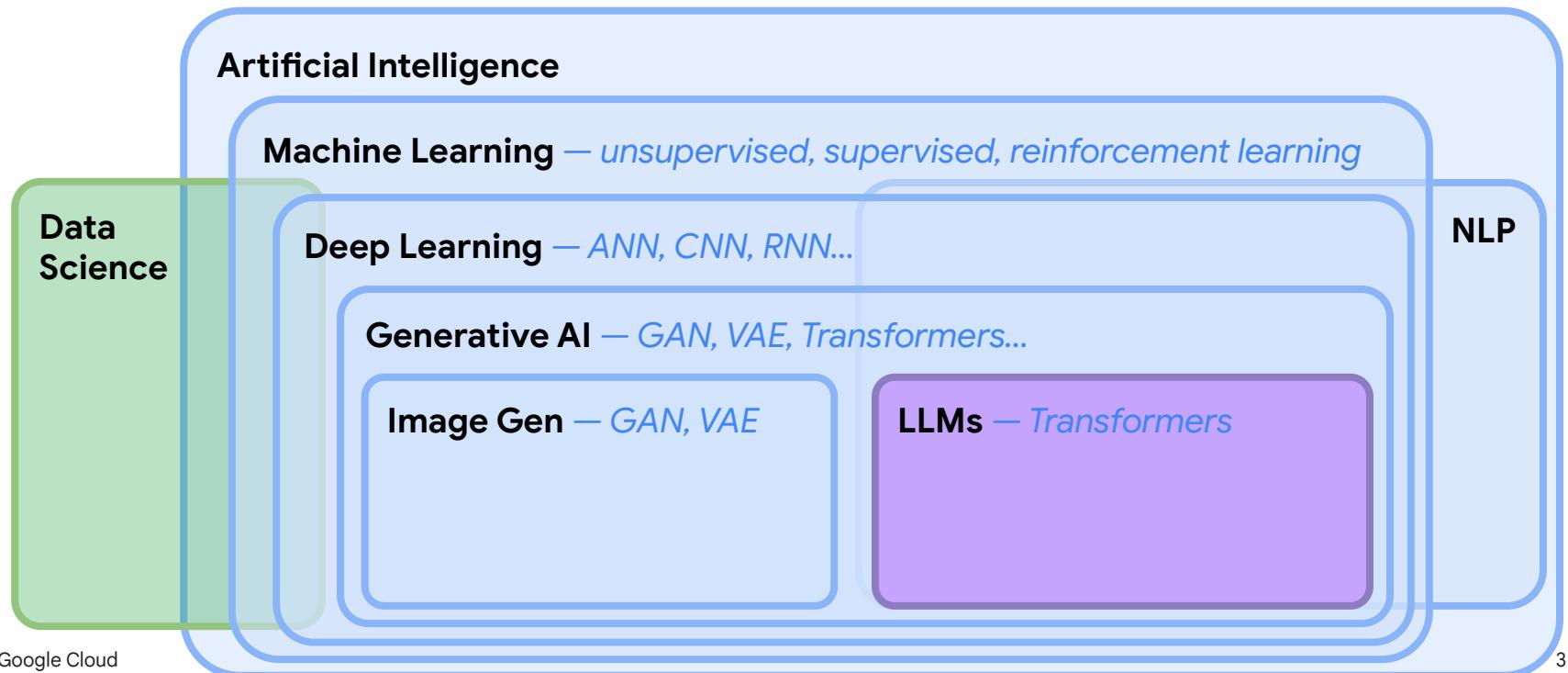
  
[@glaforge@uwyn.net](mailto:@glaforge@uwyn.net)



01

# A few notions and vocabulary

# Artificial Intelligence, Machine Learning, Data Science, Deep Learning...



# Google invented the Transformer architecture

						
<b>2017</b> Transformer	<b>2018</b> BERT	<b>2018</b> AlphaFold	<b>2019</b> T5	<b>2021</b> LaMDA	<b>2022</b> PaLM	<b>2023</b> PaLM 2
Google invents Transformer kickstarting LLM revolution	Google's groundbreaking large language model, BERT	AlphaFold predicts 3D models of protein structures	Text-to-Text Transfer Transformer LLM 10B P model open sourced	Google LaMDA model trained to converse	Google PaLM single model to generalize across domains	Google PaLM 2 model is the SOTA LLM

Responsible AI at the foundation

on

her

in

quantum

went

into

yesterday

morning

how

much

this

project

downtown

to

in

office

roof

at

has

a

He takes pride in his hard work and effort

accountant meticulously analyzed related financial documents

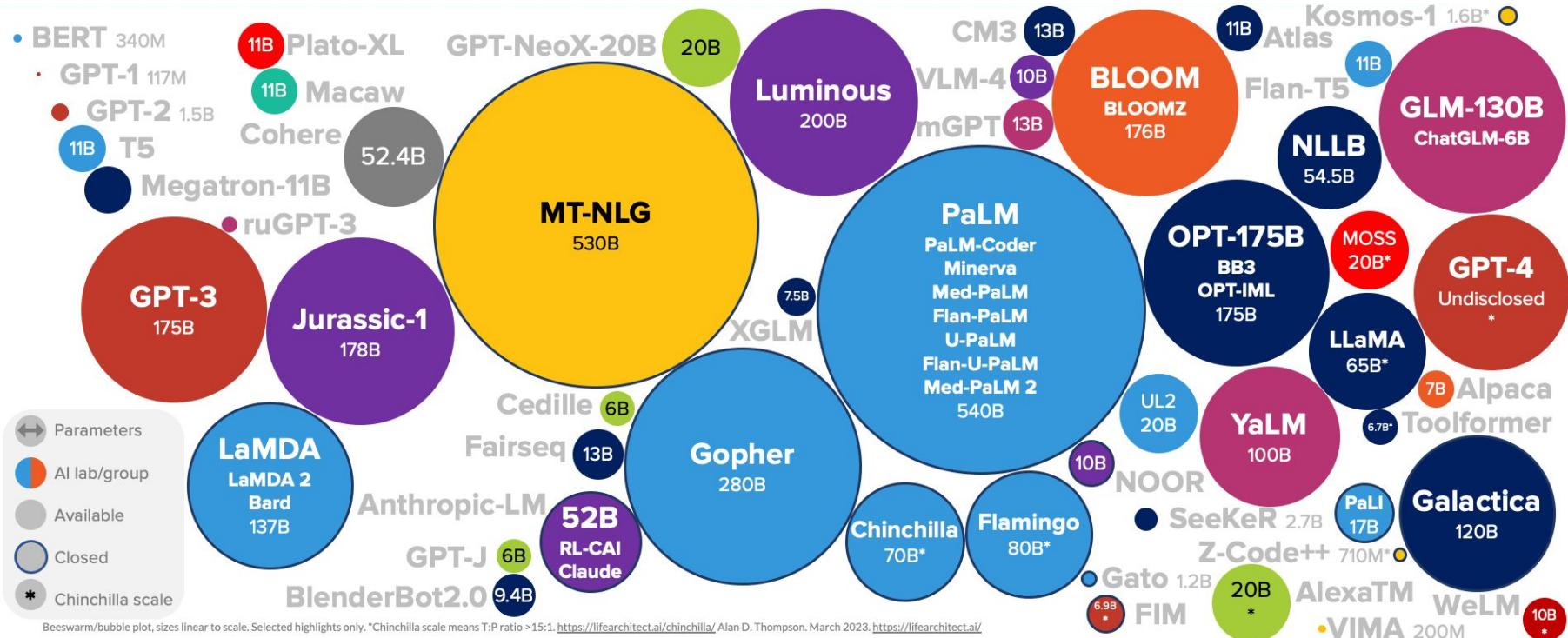
My commute is easy because I

in my neighbourhood

# So what are Large Language Models?

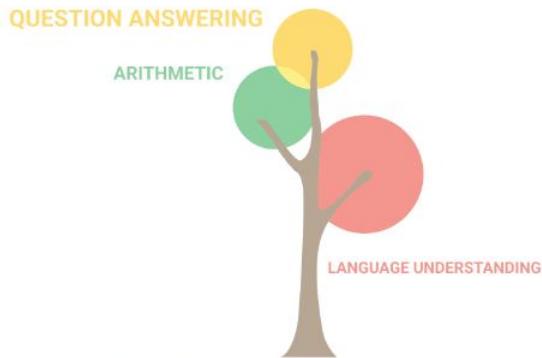
- Transformer-based neural network architecture that can **recognize**, **predict**, and **generate** human language
- Trained on huge corpuses of text, in various languages and domains
  - *Ex: PaLM 2 learned 340 billion **parameters**, and trained over 3.6 trillions of **tokens***
- Learn the **statistical relationships between words and phrases**, as well as the patterns of human language
- Can be **fine-tuned** for specific tasks or domain knowledge

# Language model sizes – March 2023



Beeswarm/bubble plot, sizes linear to scale. Selected highlights only. \*Chinchilla scale means T:P ratio >15:1. <https://lifearchitect.ai/chinchilla/>. Alan D. Thompson, March 2023. <https://lifearchitect.ai/>

# With larger models emerge new capabilities



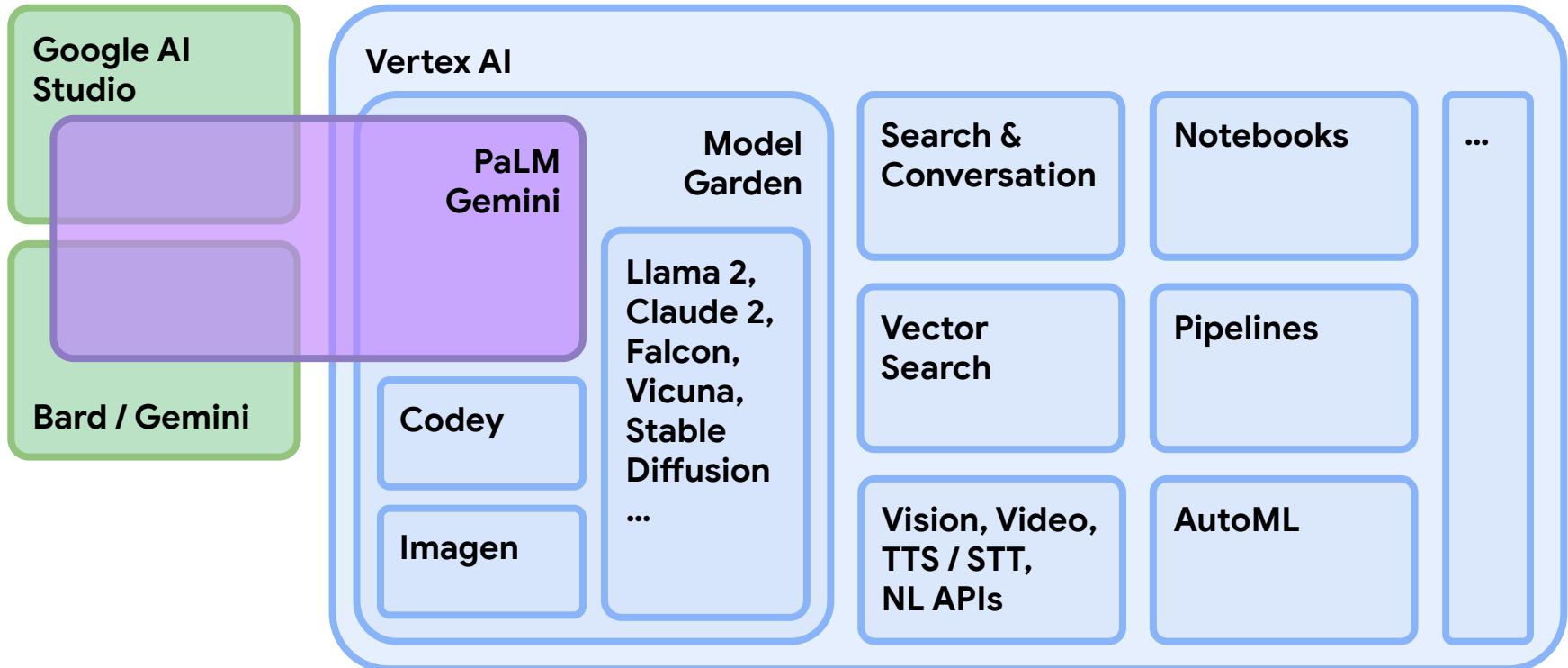
8 billion parameters

02

# What does Google (Cloud) offers?



# Gemini, PaLM, Codey, Imagen, Vertex AI...



# What is Gemini?

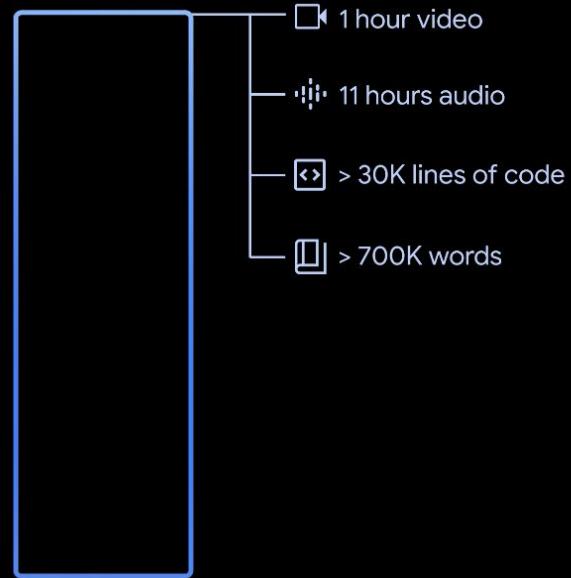


- **Gemini** is Google Deep Mind's most capable AI model
- It's a **multimodal** large language model: text, **images**, videos
- Comes in **3 sizes**: Nano, Pro, and Ultra
- Supports **function calling**
- Ranks at the top of the various LLM benchmarks  
*(general knowledge, translation, image understanding, reasoning, math, coding, and more...)*

# Gemini 1.5

up to  
10M

in research



Gemini 1.0 Pro

32K

GPT-4 Turbo

128K

Claude 2.1

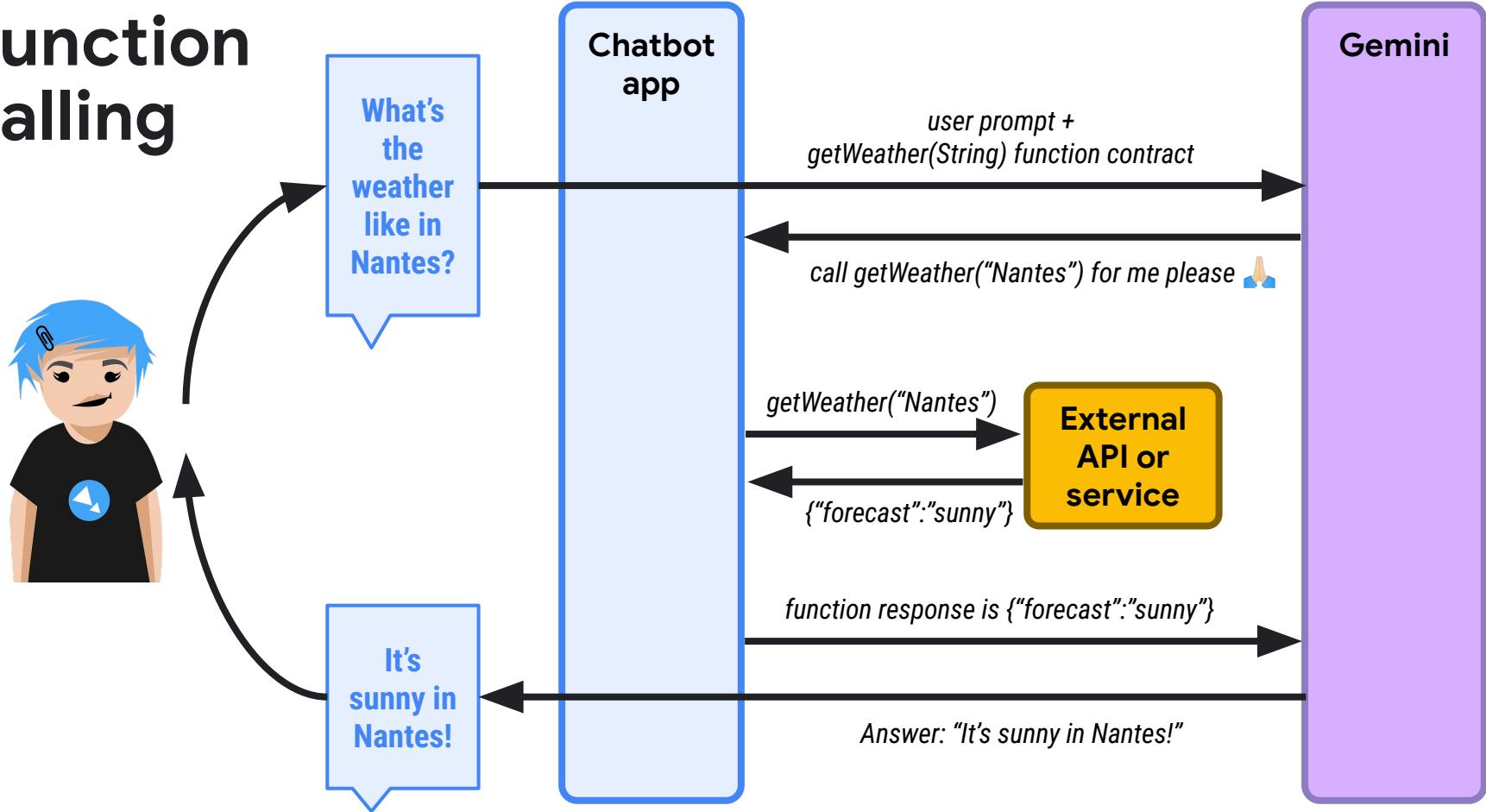
200K

Gemini 1.5 Pro

1M

Tokens

# Function calling



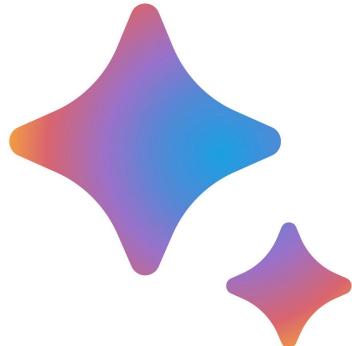
03

**Time to have a  
look at some  
demos!**



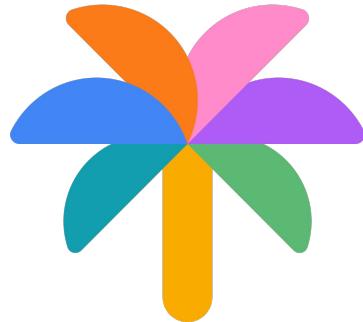
# Let's play with Bard and MakerSuite!

Bard  
Gemini



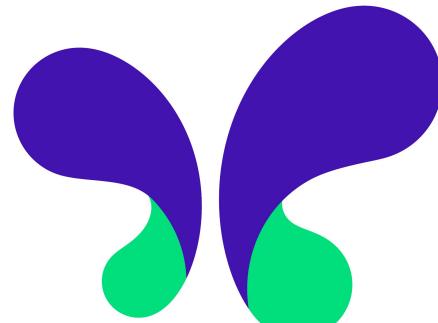
Google Cloud

Gemini



PaLM

MakerSuite  
Google AI  
Studio





# Untitled prompt



Sample prompts



Summarize a paragraph



Casual pon

Write your prompt



What's the name of the first cat who stepped on the moon?

Félicette



## Untitled prompt

FREE-FORM

STRUCTURED

SAVE

&lt; &gt; VIEW CODE

## Prompt

What is the name of the first cat who stepped on the moon?



## Response



Markdown

Your question contains a factual error. No cat has ever stepped on the moon.

We want your feedback.

Model

text-bison (latest)



Temperature



1

&gt; Advanced

SUBMIT

RESET PARAMETERS



What is the name of the first cat who stepped on the moon?



[View other drafts](#)



There has never been a cat on the moon.

On October 18, 1963, a French stray cat named Félicette became the first and only cat to be launched into space. She flew on a suborbital mission, reaching an altitude of 154 kilometers before returning safely to Earth. However, she was never landed on the moon.

It is possible that in the future, cats could be sent to the moon, but there are currently no plans to do so.

Check completed • [Understand the results](#)



Search related topics



First cat on moon



Enter a prompt here



▼ Context [?](#)

Give me in output a dessert recipe that contains the fruit or vegetable in input

▼ Examples [?](#)

[ADD INPUT COLUMN](#)

INPUT Add an input prefix

banana

OUTPUT Add an output prefix

banana split



apple

apple tart



[+ Write an input to add a new row](#)

[+ Write an output to add a new row](#)

▼ Test [?](#)

Batch testing

[\[+\] ADD TO EXAMPLES](#)

INPUT Add an input prefix

carrot

Add an output prefix

OUTPUT

Markdown



carrot cake

## Prompt

In the following biography, extract the name of the person, their job, and their company, following the following JSON format:

```
{"person": "John Doe", "job": "Computer Scientist", "company": "Meta"}
```

Here is the biography:

Guillaume Laforge is Developer Advocate for Google Cloud, where he spreads the word to developers about the serverless compute space and service orchestration solutions.

Guillaume is a Java Champion, a co-author of Groovy in Action, and a founding member of the French tech podcast Les CastCodeurs.

Guillaume is the co-founder of the Apache Groovy programming language project, which he led under the umbrella of G2One (a Groovy/Grails startup where he was the VP Technology), then SpringSource, VMware, and the Pivotal spin-off. When the project joined to the Apache Software Foundation, he was the Chair of the Apache Groovy Project Management Committee (PMC). He is a member of the Micronaut Technology Advisory Board. He initiated the creation of the Grails web application framework, and founded the Gaelyk project, a lightweight toolkit for developing applications in Groovy for Google App Engine.

## Response

- Markdown

```
{"person": "Guillaume Laforge", "job": "Developer Advocate", "company": "Google Cloud"}
```

# Python is all the rage in AI...



What's in it for us,  
**Java** developers?

# Generating bedtime stories for kids

<https://bed-time-stories.web.app/>

## Bedtime Stories

Let Generative AI create cool stories for you and your kids, with the PaLM API!

This application is a [Micronaut](#) application, developed with [Apache Groovy](#), deployed on Google [Cloud Run](#), and calling the [Vertex AI PaLM API](#).

Pick an example character, or invent your own:

▼

Pick a setting defining when and where the action takes places, or create your own:

▼

# Bedtime Stories

Let Generative AI create cool stories for you and your kids, with the PaLM API!

This application is a [Micronaut](#) application, developed with [Apache Groovy](#), deployed on Google [Cloud Run](#), and calling the [Vertex AI PaLM API](#).

Pick an example character, or invent your own:

a brave knight in a shiny silver armor

a brave knight in a shiny silver armor

Pick a setting defining when and where the action takes places, or create your own:

in a small village on a quiet island

in a small village on a quiet island

Pick a possible plot, or roll your own:

a mission to break the spell that has been cast on her kingdom

a mission to break the spell that has been cast on her kingdom

 Generate story

Once upon a time, in a small village on a quiet island, there lived a brave knight in a shiny silver armor. His name was Sir Lancelot, and he was the most courageous knight in the kingdom.

One day, a wicked wizard cast a spell on the kingdom, turning everyone in the village into stone. Sir Lancelot was the only one who was not affected by the spell. He knew that he had to break the spell and save the kingdom.

Sir Lancelot set out on his quest to break the spell. He traveled through the forest, over the mountains, and across the sea. He faced many dangers along the way, but he never gave up.

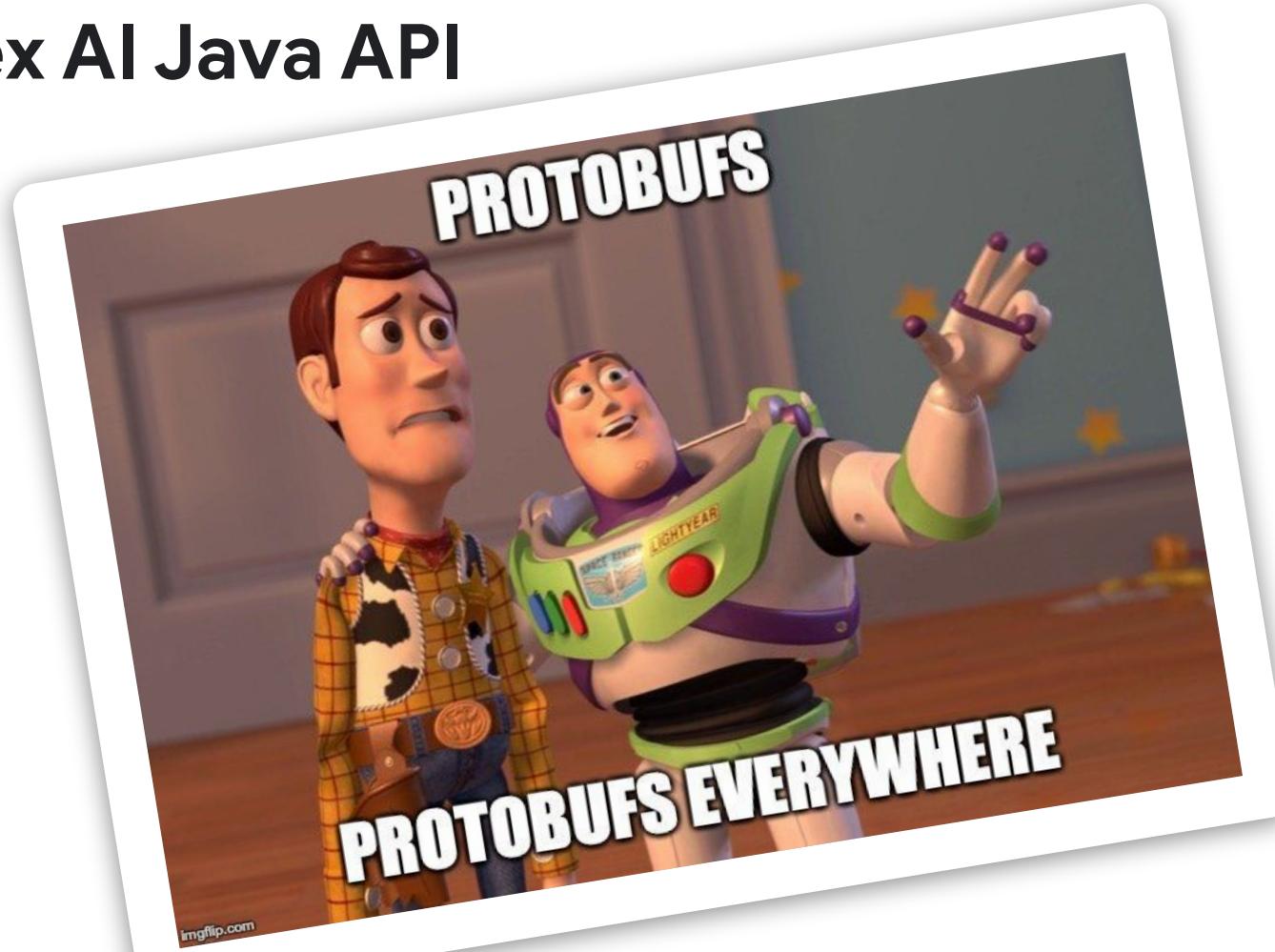
One day, Sir Lancelot came to a castle. He knew that the wizard must be inside, so he went inside to confront him.

The wizard was a powerful sorcerer, and he put Sir Lancelot through many difficult tests. But Sir Lancelot was determined to break the spell, and he never gave up.

Finally, Sir Lancelot defeated the wizard and broke the spell. The people of

Made with ❤ by [Guillaume Laforge](#) — Code available on [Github](#)

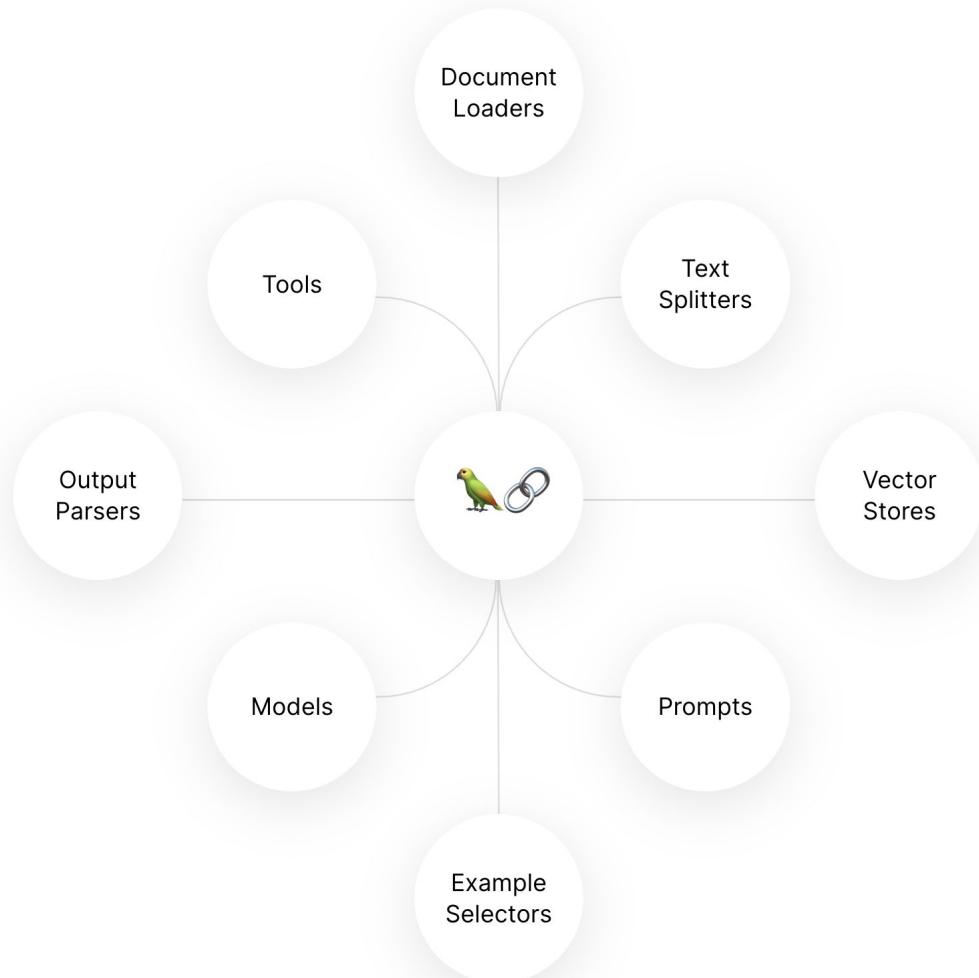
# Vertex AI Java API



# Gemini...



# LangChain4j



# Searching the Apache Groovy documentation

Apply the RAG pattern:  
Retrieval Augmented Generation



Google Cloud

★ Apache Groovy

Learn Documentation Download Support Contribute Ecosystem Blog posts Socialize

[Limitations](#)  
[Compatibility with AST transformations](#)  
[Prefix and postfix operations](#)  
**1.4.6. Record classes (incubating)**  
[Special record features](#)  
[Compact constructor](#)  
[Serializability](#)  
[Groovy enhancements](#)  
[Argument defaults](#)  
[Declarative `toString`](#)  
[Customization](#)  
[Obtaining a list of the record component values](#)  
[Obtaining a map of the record component values](#)  
[Obtaining the number of components in a record](#)  
[Obtaining the nth component from a record](#)  
[Optional Groovy features](#)  
[Groovy annotations](#)

A workaround is to use the `+=` operator instead.

### 1.4.6. Record classes (incubating)

Record classes, or *records* for short, are a special kind of class useful for modelling plain data aggregates. They provide a compact syntax with less ceremony than normal classes. Groovy already has AST transforms such as `@Immutable` and `@Canonical` which already dramatically reduce ceremony but records have been introduced in Java and record classes in Groovy are designed to align with Java record classes.

For example, suppose this example, let's simulate a message body. We can do:

```
record Message(String body)
```

We'd use the record class:

```
def msg = new Message("Hello, world!")
```

The reduced ceremony of using records is apparent here. In fact, it's a lot more concise than the original code.

How can I do string interpolation in Groovy?

String interpolation in Groovy is done by using the \${} syntax. For example, the following code will print the string "Hello, world!" to the console:

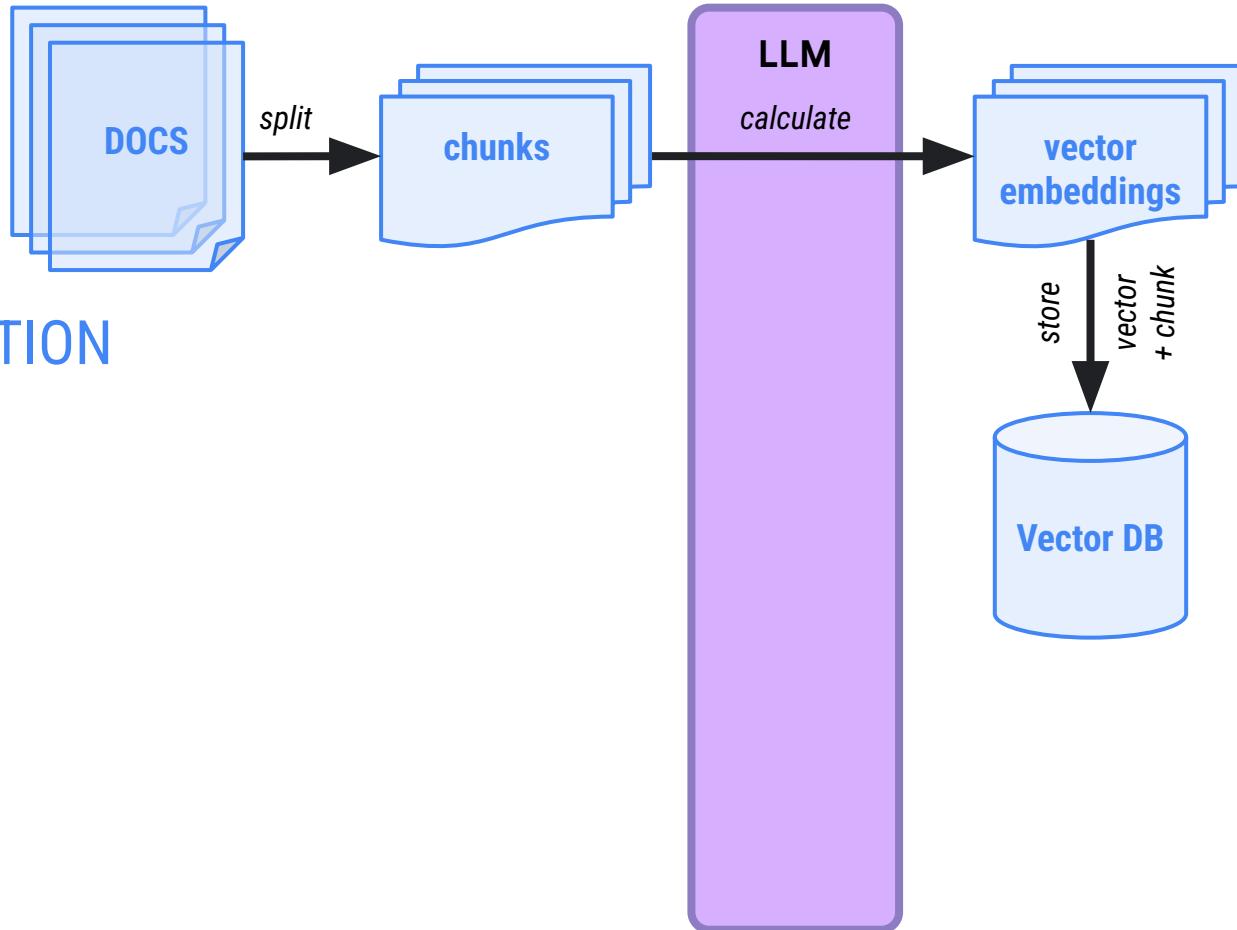
```
println "Hello, world!"
```

You can also use string interpolation to access variables and expressions. For example, the following code will print the string "The current date is \${date}" to the console:

```
println "The current date is ${date}"
```

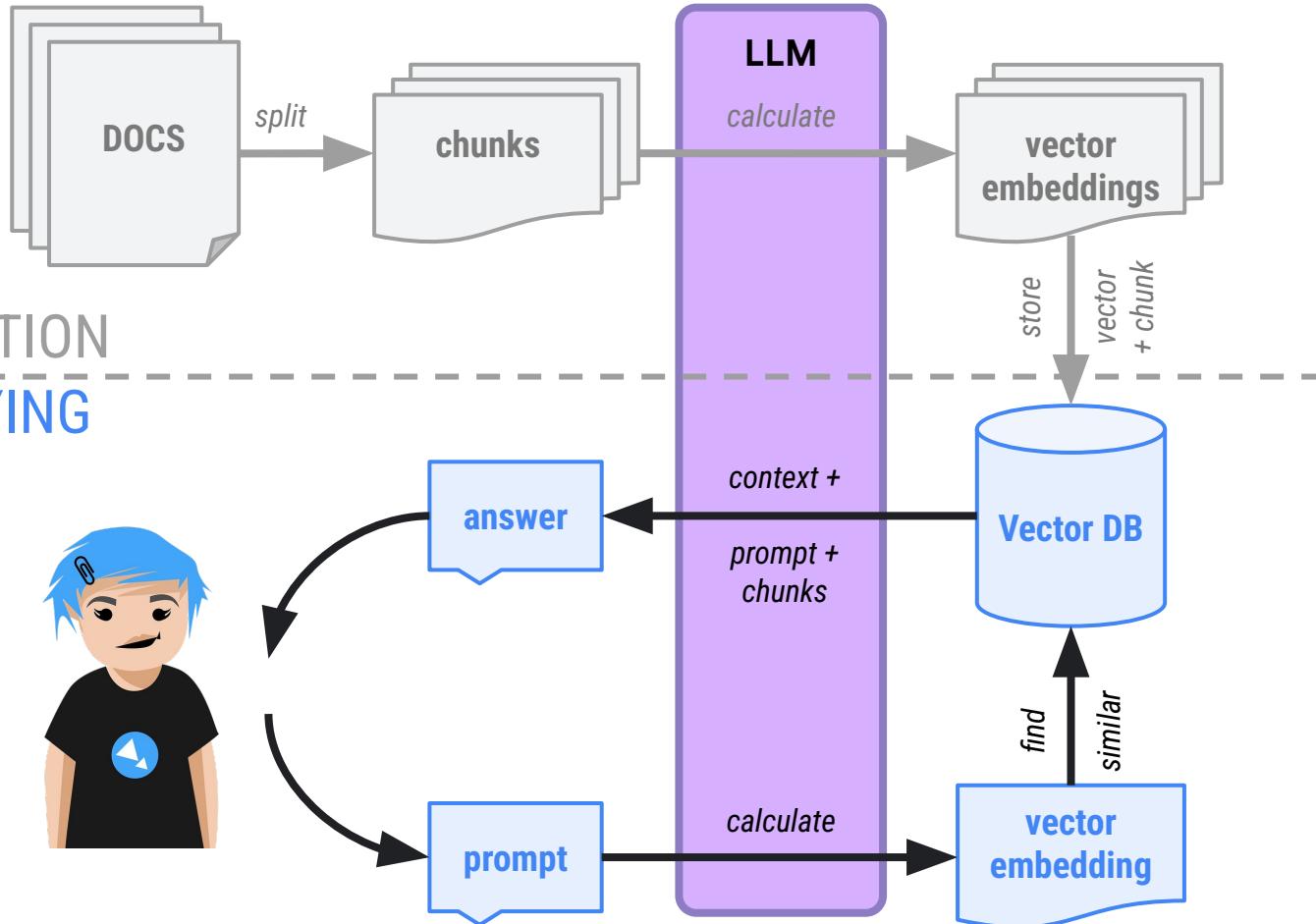
# RAG

## ① INGESTION



# RAG

- ① INGESTION
- ② QUERIES



# ★ Apache Groovy – LLM Doc Chatbot

Welcome to the [Apache Groovy](#) documentation chatbot. I'm a Large-Language-Model trained on the online Groovy [documentation](#). Feel free to ask me any question about the Groovy programming language.

Please ask your question

Send

Does Groovy support interfaces?

Yes, Groovy supports interfaces. The Groovy documentation provides an example of how to declare an interface and implement it in a class. The documentation also discusses how to extend an interface and how to make an instance of an object implement an interface at runtime.

Show me an example of an interface, and a class that implements that interface

Sure, here is an example of an interface and a class that implements that interface:

```
interface Greeter {  
    void greet(String name);  
}  
  
class SystemGreeter implements Greeter {  
    public void greet(String name) {  
        System.out.println("Hello, " + name);  
    }  
}
```

The Greeter interface defines a single method, greet, which takes a String argument and prints a greeting to the console. The SystemGreeter class implements the Greeter interface and provides an implementation for the greet method.

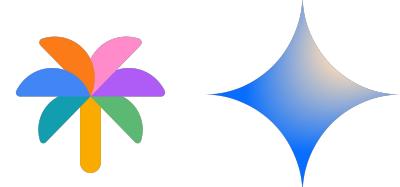
04

# Summary



# Summary

- **PaLM 2 & Gemini** are **powerful LLM** that allow you to build all sorts of generative AI ideas and projects
- Using the PaLM API from Java is not necessarily ideal, but with REST or LangChain4j, it's easy
- Fortunately, Gemini has a much better SDK! (and is also supported in LangChain4j)
- Easy to deploy & serve containerized apps on **Cloud Run**
- And **LangChain4J** is pretty sweet!



# More about LangChain4j

Current LangChain4j Integrations

Model Integrations	Embedding Stores	Document Loaders
Azure OpenAI	Chroma	txt
DashScope	Elasticsearch	html
Google Vertex AI	Milvus	doc
HuggingFace	Pinecone	pdf
LocalAI	Vespa	xls
OpenAI	Weaviate	ppt
	Redis	url
	Astra DB	
	Cassandra	

LangChain4j logo

DEVOXX BELGIUM

ING Google Cloud aws JET BRAINS ORACLE

[youtube.com/watch?v=BD1MSLbs9KE](https://youtube.com/watch?v=BD1MSLbs9KE)



# What to build next?

- **Chat Q&A** with the Apache Groovy documentation
  - improve the quality of responses
- **News curation** that collects & summarize my social feeds

What about you,  
**what would you build?**



# Going further

Learn more about **Generative AI**  
at [goo.gl/generativeai](https://goo.gl/generativeai)

**Guillaume Laforge**  
Developer Advocate  
**Google** Cloud

  
[@glaforge](https://twitter.com/glaforge)  
[@uwyn.net](mailto:glaforge@uwyn.net)

Google Cloud



# Thanks!

**Guillaume Laforge**  
Developer Advocate  
**Google** Cloud

  
[@glaforge](https://twitter.com/glaforge)@uwyn.net

