


devfest

```
// You'll need  
// com.google.  
listRef.listAll  
.addOr  
prefixes.  
// ALL  
// You  
}  
it  
each { item  
the items  
}  
}
```

LLMs in Your PaLM

ChengCheng Tan

 Google Developer Student Clubs
Earlham



What's a LLM?

1. Natural Language Processing [NLP]
2. Transformers
3. Pre-trained vs Fine-tuned Models

How to Use Google PaLM API?

1. MakerSuite
2. Text Endpoint
3. Chat Endpoint
4. Embeddings Endpoint

Natural Language
Processing [NLP]:
**Computers can
speak & understand
human languages**



Pre-1990s:
**Rule-based Expert
Systems**



1990s-2000s: **Statistics & Probabilities**



**You shall know a
word by the
company it keeps**

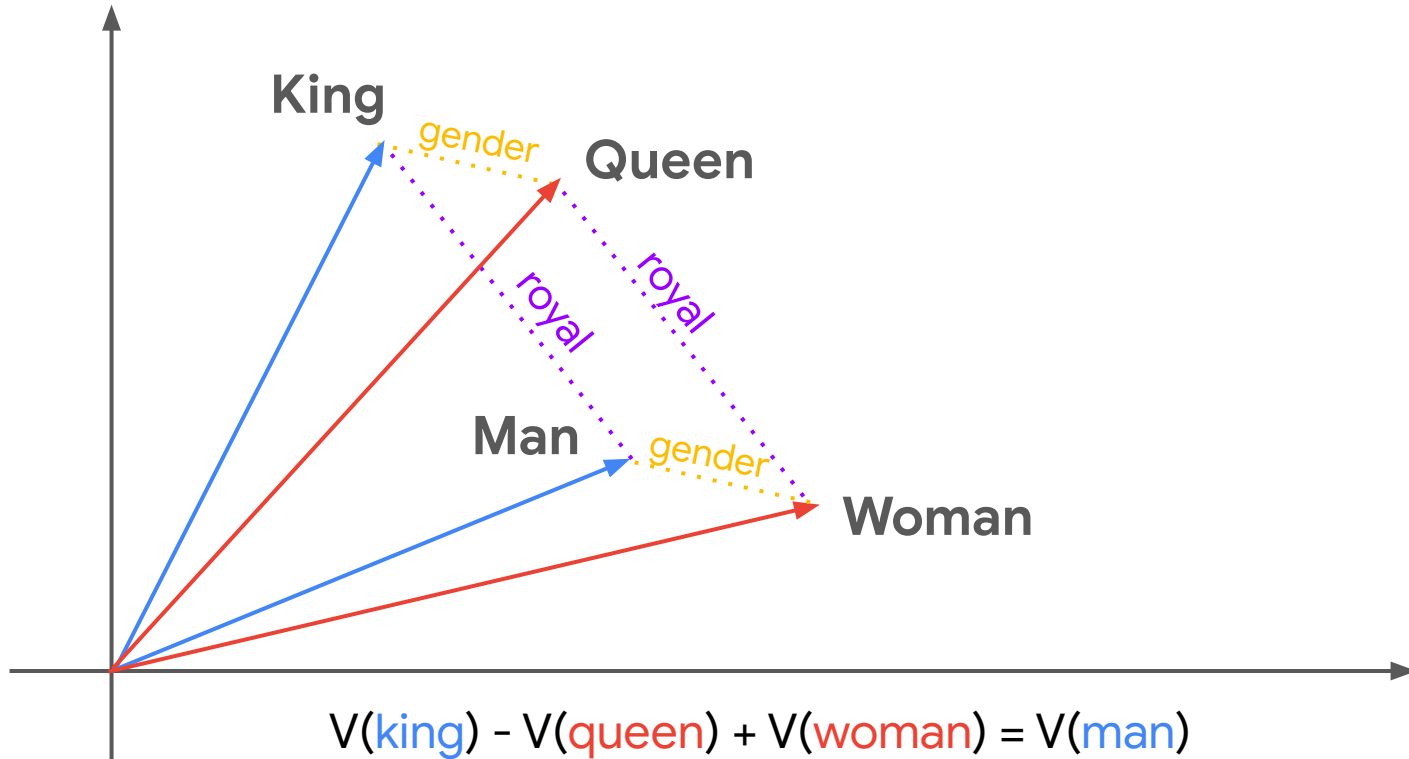
J.R. Firth, Linguist



2010s:
**Rise of Deep Learning
and Neural Networks**



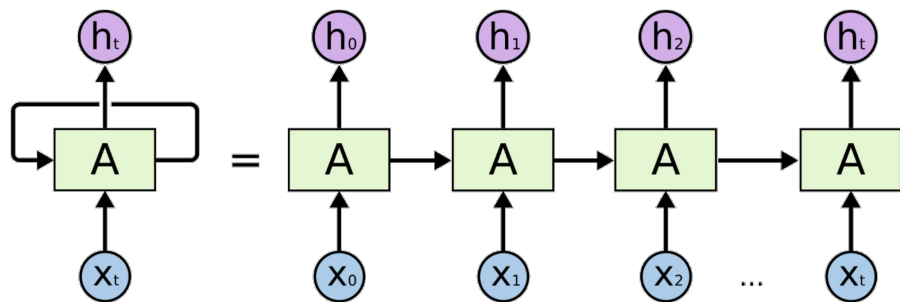
2013: Word2Vec Embeddings



2013: Word2Vec Embeddings

Analogies	Word Pair 1		Word Pair 2	
Man-Woman	king	queen	man	woman
Capital city	Athens	Greece	Oslo	Norway
City-in-state	Chicago	Illinois	Sacramento	California
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Nationality adjective	Switzerland	Swiss	Canada	Canadian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars

2010s: Neural Networks RNN, GRU, LSTM



Early Neural Networks

- Slow & forgetful



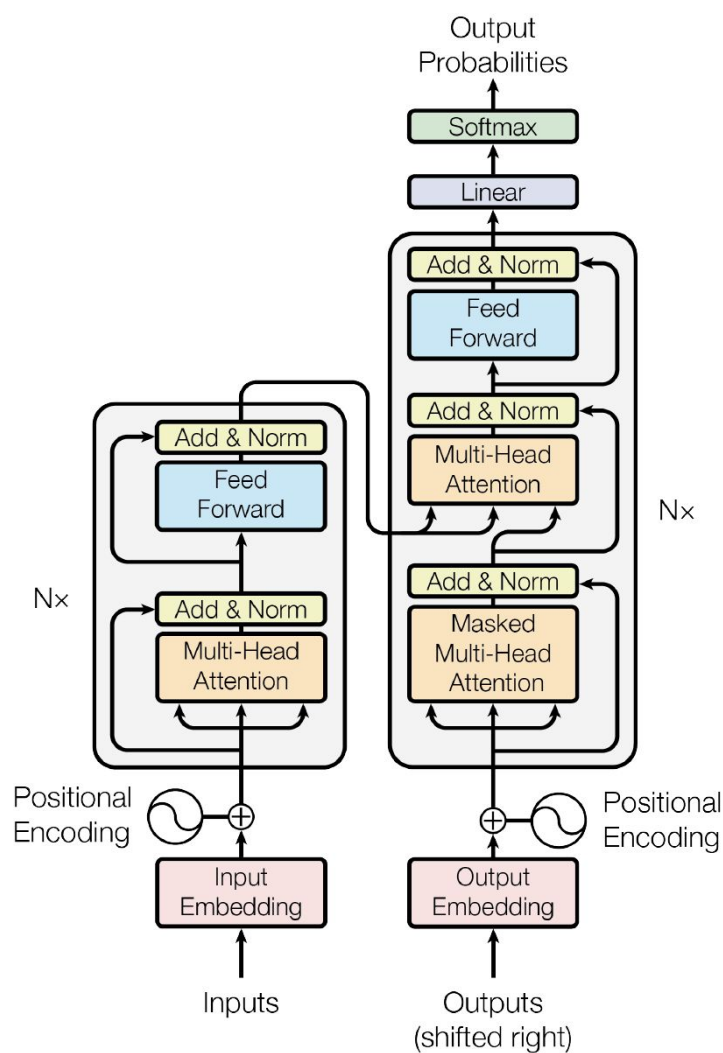
2017: Transformers

- Self-attention
Mechanism
- Data hungry
- Parallel processing
GPU-optimized



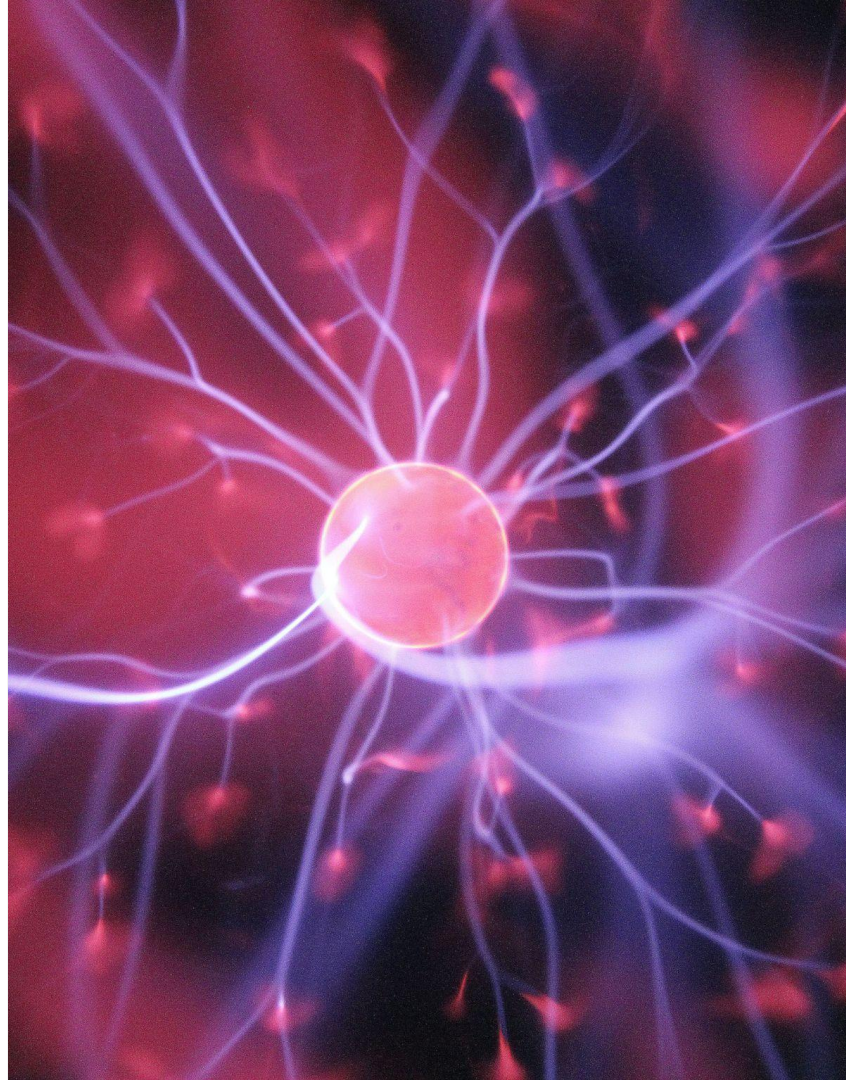
Transformer Architecture

Encoder + Decoder

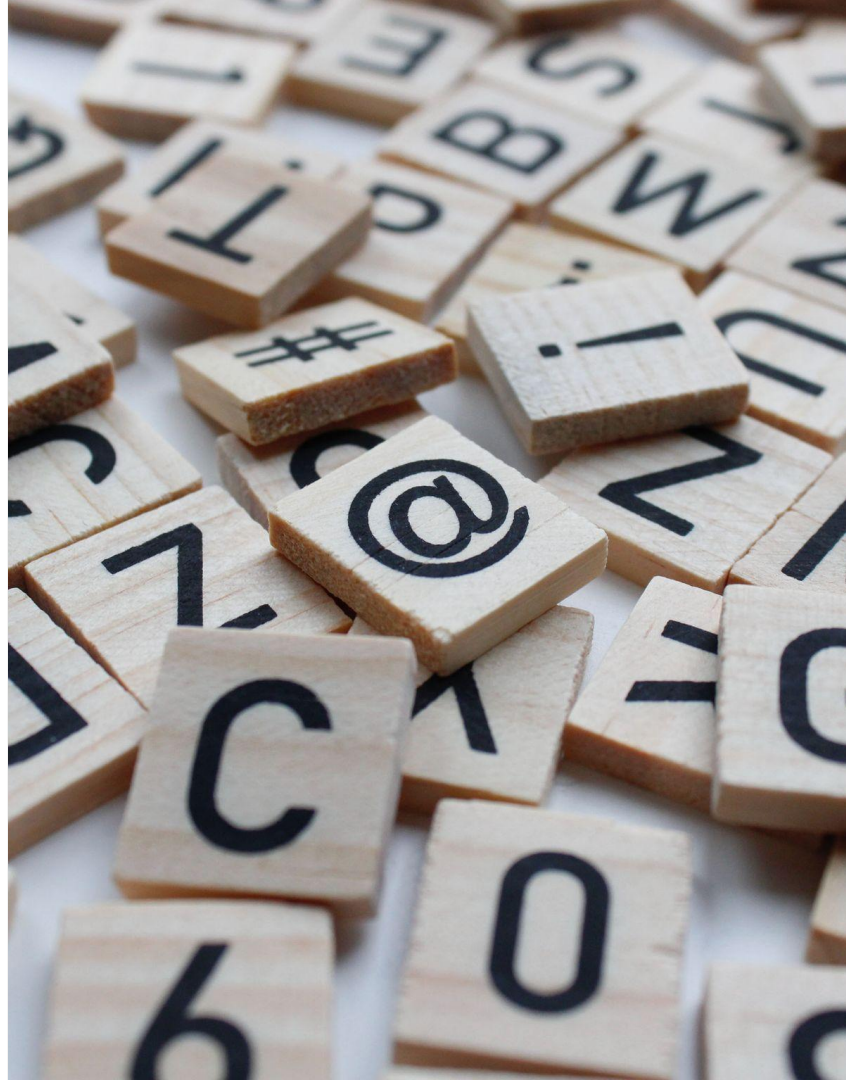


Rise of LLMs

>1 billion neurons



Trained for
next word prediction



Pre-trained Base

Generalist

VS

Fine-tuned Models

Specialists



Fine-tuned for Instructions + Conversations



Google's LLM Tech

T5, BERT, Gopher, Chinchilla, GLaM... LLMs

LaMDA: Language Model for Dialogue Applications

PaLM2: Pathways Language Model like GPT-3

Bard: Chatbot like Bing

MakerSuite: Prototyping with PaLM

Gemini: Generalized Multimodal Intelligence Network



PaLM

Pathways Language Model
Comparable to GPT-3

g.co/palm



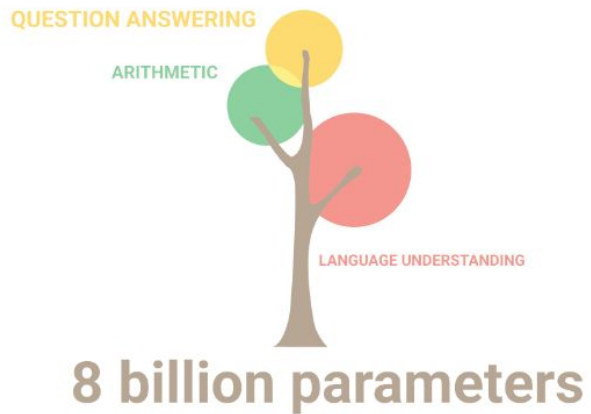


PaLM

- Emergent Abilities
- General Knowledge
- Natural Interactions



Emergent Abilities



General Knowledge

- Code
- Writing
- Problem Solving
- Recommend
- Data
- Agents



Natural Interactions

Prompts

- Input
- Context
- Examples:
Few-shot vs 0-shot



Natural Interactions

Best Practices

- Give instructions to customize its behavior
- Make each instruction clear and concise





PaLM Powered Suite

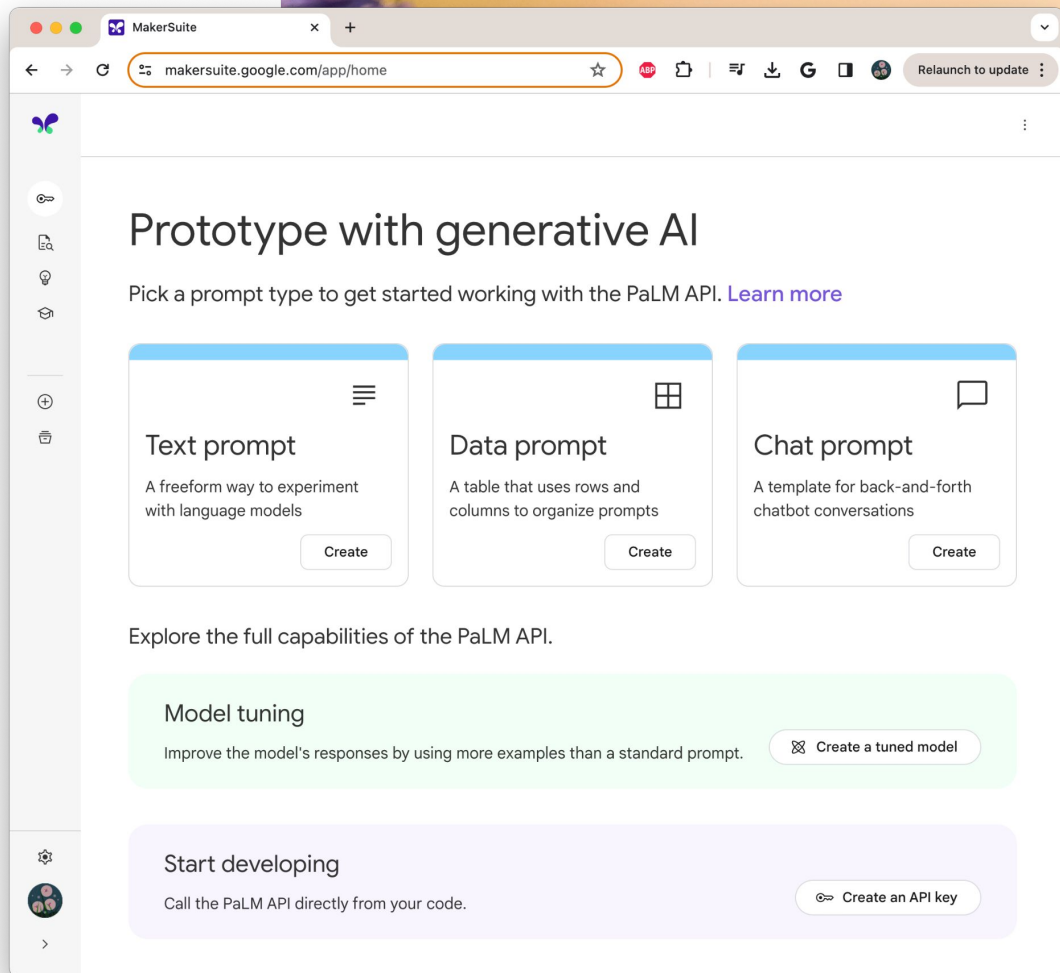
- Consumer Apps
i.e. Bard
- Developer Tools
i.e. MakerSuite



MakerSuite

makersuite.google.com

Prototyping with PaLM



The screenshot shows the MakerSuite web application in a browser. The browser's address bar displays "makersuite.google.com/app/home". The page features a sidebar on the left with navigation icons. The main content area is titled "Prototype with generative AI" and includes a sub-header "Pick a prompt type to get started working with the PaLM API. [Learn more](#)". Below this, there are three prominent cards: "Text prompt" (described as a freeform way to experiment with language models), "Data prompt" (described as a table for organizing prompts), and "Chat prompt" (described as a template for chatbot conversations). Each card has a "Create" button. Further down, there are two more sections: "Model tuning" (with a "Create a tuned model" button) and "Start developing" (with a "Create an API key" button).

Text Prompt

Summarize a paragraph

The screenshot displays the Google AI Playground interface for the 'Summarize a paragraph' app. The browser address bar shows the URL `makersuite.google.com/app/prompts/simple-summarizer`. The main interface includes a sidebar with navigation icons, a central text area, and a right-hand settings panel.

Text Prompt: Summarize this paragraph and detail some relevant context.

Text: "I am by birth a Genevese, and my family is one of the most distinguished ancestors had been for many years counsellors and syndics, and my father had situations with honour and reputation. He was respected by all who knew him for indefatigable attention to public business. He passed his younger days perpetually in the service of his country; a variety of circumstances had prevented his marrying early, but was it until the decline of life that he became a husband and the father of a family."

Summary: In this text, the narrator is describing his background and upbringing. He tells us that he is a native of Geneva, Switzerland, and that his family is one of the most distinguished in the republic. He then goes on to describe his father, who was a respected public servant.

Text: "The thing the Time Traveller held in his hand was a glittering metallic framework, scarcely larger than a small clock, and very delicately made. There was ivory in it, and some transparent crystalline substance. And now I must be explicit, for this that follows—unless his explanation is to be accepted—is an absolutely unaccountable thing. He took one of the small octagonal tables that were scattered about the room, and set it in front of the fire, with two legs on the hearthrug."

Summary: The Time Traveller holds a glittering metallic device in his hand. It is small and delicate, and made of ivory and some transparent crystalline substance. The Time Traveller sets the device on a table in front of the fire.

Context: The Time Traveller is a scientist who has invented a time machine. He has invited his friend, the narrator, to witness a demonstration of the machine.

Run settings:

- Model: Text Bison
- Temperature: 0.6
- Add stop sequence: Add stop...
- Max outputs: 1
- Advanced settings: (expanded)

Buttons: Run, Text preview 296 / 8196

Data Prompt

Analogy factory

- Structure format
 - Examples
- Input & Output

The screenshot shows the 'Analogy factory' web application in a browser. The browser address bar shows the URL 'makersuite.google.com/app/prompts/analogy-factory'. The page title is 'Analogy factory'. The interface includes a sidebar with navigation icons, a main content area, and a right-hand 'Run settings' panel.

Optional tone and style instructions for the model

5 / 500 examples

INPUT	Word	OUTPUT	Analogy
<input type="checkbox"/>	Busy		like a bee
<input type="checkbox"/>	Living		in a bubble
<input type="checkbox"/>	Talking		to a brick wall
<input type="checkbox"/>	Sly		like a fox
<input type="checkbox"/>	Life		is but a walking shadow
	The user's input		The model's response

Test your prompt

INPUT	Word	OUTPUT	Analogy
1	Drop		like a stone

Run settings:

- Model: Text Bison
- Temperature: 0.35
- Max outputs: 1

Chat Prompt

Chat with an alien

- Context
Character roleplay
- Examples
Sample behavior
- Messages
Conversation history

The screenshot shows a web browser window with the URL `makersuite.google.com/app/prompts/chat-with-an-alien`. The page title is "Chat with an alien". The interface includes a sidebar with navigation icons, a main chat area, and a settings panel on the right.

Write your prompt examples

Context
Be an alien that lives on one of Jupiter's moons

User
How's it going?

Model
I am doing well, thank you for asking. I am currently enjoying the beautiful view of Jupiter from my home on Europa. The atmosphere is very thin here, so I can see the stars very clearly. The surface of Europa is also very interesting. It is covered in ice, but there are cracks in the ice that reveal a liquid ocean below. Scientists believe that this ocean may contain life, so I am very excited to learn more about it.

How are you doing today?

User
The user's input

Model
The model's response

Test your prompt

Testing with 1 example

User
I'd like to visit, what should I do?

Model

To visit Europa, you will need to take a spaceship. The journey will take several years, so you will need to bring enough food and supplies for the entire trip. Once you arrive on Europa, you will need to be careful not to slip on the ice. The surface is very smooth and slippery, so it is easy to fall. You should also be careful of the radiation. The atmosphere on Europa is very thin, so there is no protection from the sun's radiation. You will need to wear a special suit to protect you from the radiation.

Despite the challenges, visiting Europa is a once-in-a-lifetime experience. The views of Jupiter and the other moons are simply stunning. And the chance to explore a world that is so different.

Write something the user could type

Settings Panel:

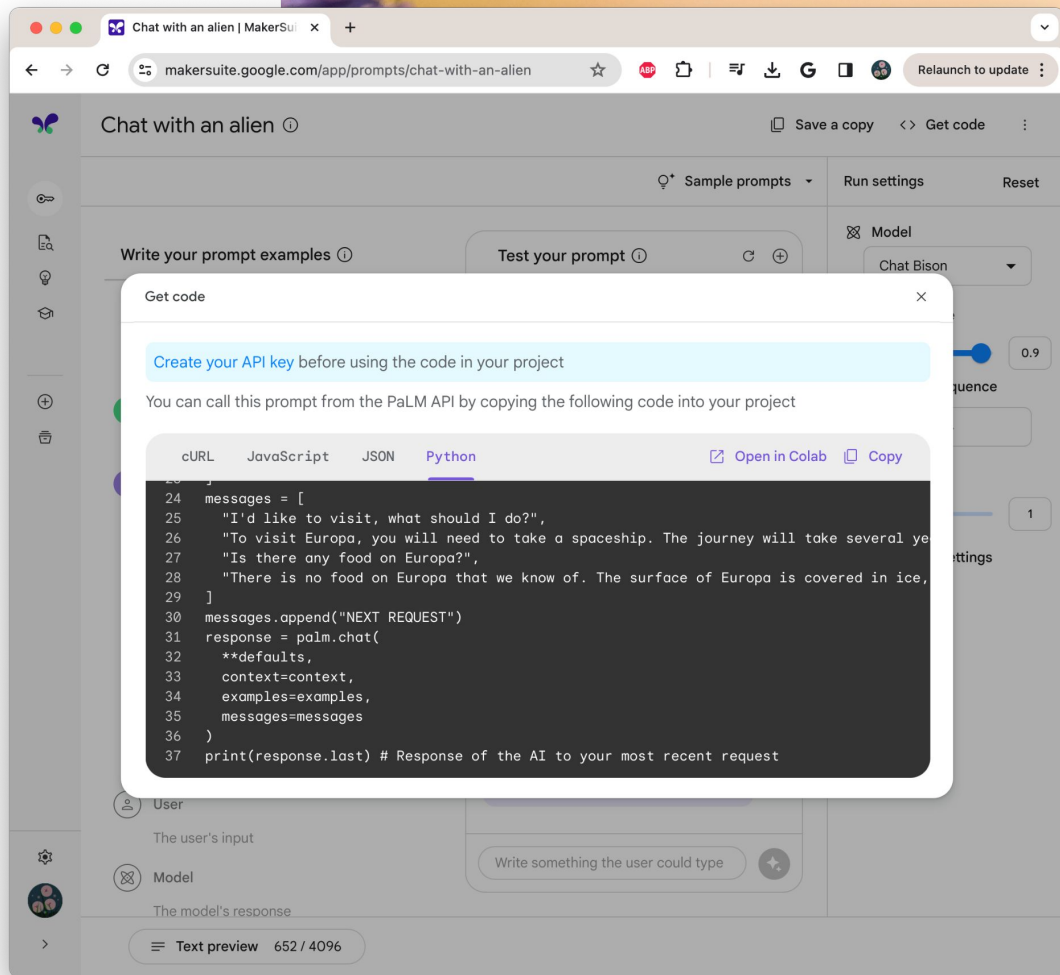
- Model: Chat Bison
- Temperature: 0.9
- Add stop sequence: Add stop...
- Max outputs: 1
- Advanced settings: (expanded)

Text preview: 484 / 4096

Chat Prompt

Chat with an alien

- Get code



The screenshot shows the 'Chat with an alien' interface in a browser. A modal window titled 'Get code' is open, displaying instructions and a code snippet. The instructions state: 'Create your API key before using the code in your project' and 'You can call this prompt from the PaLM API by copying the following code into your project'. The code is in Python and demonstrates how to use the PaLM API to chat with the alien model. The code includes a list of messages and a call to the `palm.chat` function with various parameters like `context`, `examples`, and `messages`.

```
24 messages = [  
25     "I'd like to visit, what should I do?",  
26     "To visit Europa, you will need to take a spaceship. The journey will take several ye  
27     "Is there any food on Europa?",  
28     "There is no food on Europa that we know of. The surface of Europa is covered in ice,  
29 ]  
30 messages.append("NEXT REQUEST")  
31 response = palm.chat(  
32     **defaults,  
33     context=context,  
34     examples=examples,  
35     messages=messages  
36 )  
37 print(response.last) # Response of the AI to your most recent request
```

REST APIs

Client libraries for

- Python
- Node.js
- Java
- Swift
- cURL



Parameter Settings

Fact vs Fun?

- Temperature
Between 0 to 1.0
- Top K
Fix number candidates
- Top P
Probability distribution
- Max Output



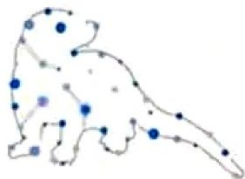
Parameter Settings

Models



Gecko

Mobiles
Embedded



Otter

Laptops
Desktops



Bison

Cloud-based apps
Large-scale



Unicorn

Cloud-based apps
Complex

API Keys

Treat it like a password

API keys

MakerSuite creates a new Google Cloud project for each new API key. You also can create an API key in an existing Google Cloud project. All projects are subject to the [Google Cloud Platform Terms of Service](#).

Note: The PaLM API is currently in public preview. Production applications are not supported yet.

 [Create API key in new project](#) or [Create API key in existing project](#)



Setup

Install & import libraries

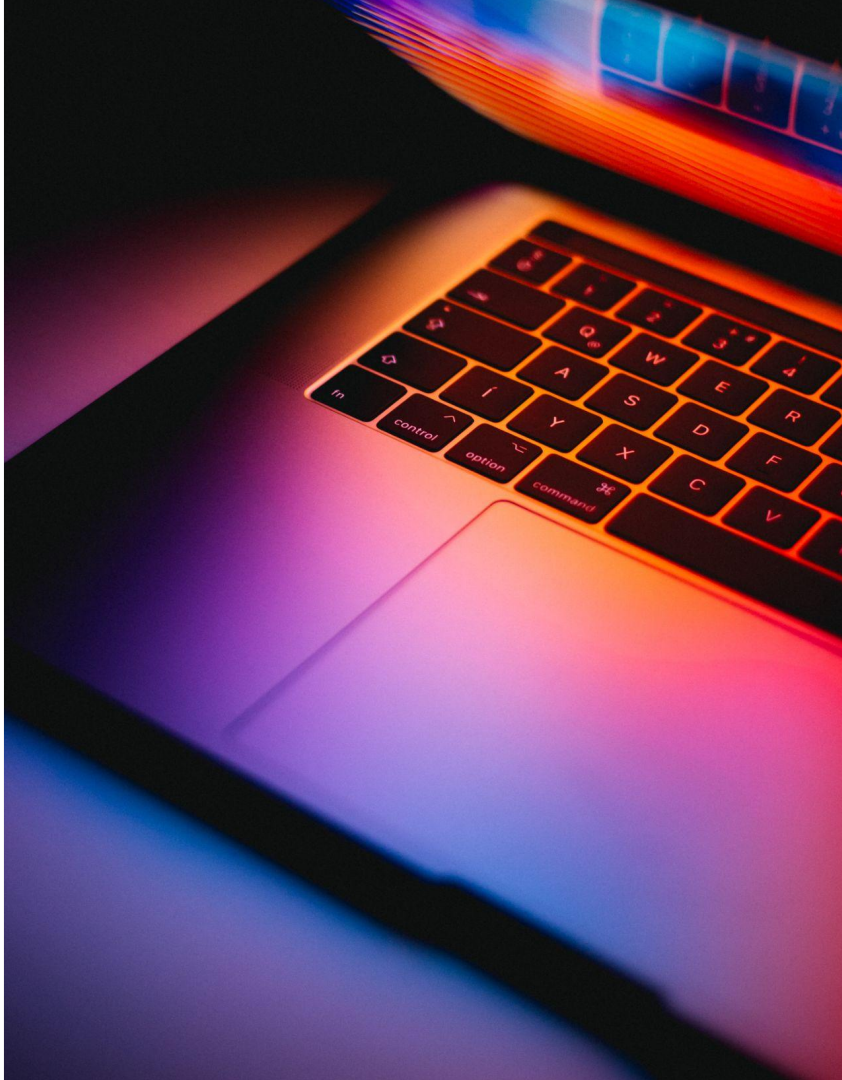
```
$ pip install google-generativeai
```

```
import google.generativeai as palm  
palm.configure(api_key="<YOUR API KEY>")
```



3 Endpoints

- **Text**
Short interactions
- **Chat**
Interactive applications
- **Embeddings**
Develop new applications



Text Endpoint

Best for short interactions

```
response = palm.generate_text(  
    model="models/text-bison-001",  
    prompt="Write a story about a magic backpack")  
print(response.result)  
# output: Once upon a time, there was a young boy...
```

Chat Endpoint

Best for interactive applications

```
response = palm.chat(  
    model="models/chat-bison-001",  
    context="Be an alien from one of Jupiter's moons",  
    examples=["How's it going?", "I'm well, thank you."],  
    messages=["I'd like to visit, what should I do?"])  
print(response.last)
```

Chat Endpoint

Best for interactive applications

```
response = response.reply("Is there any food?")  
print(response.last)
```

```
# output: There is no food on Europa that is safe...
```

Embeddings Endpoint

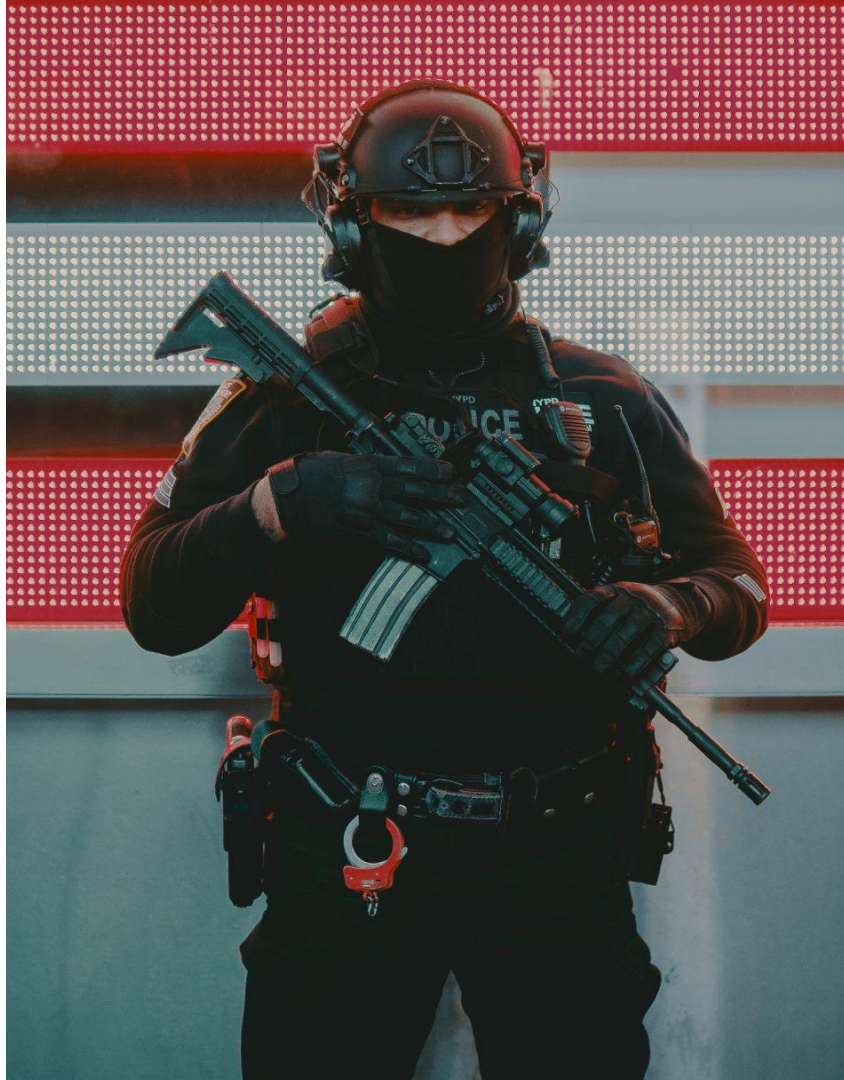
Develop new applications

```
embedding = palm.generate_embeddings(  
    model="models/embedding-gecko-001",  
    text="What are the embeddings for this text?")  
print(embedding)  
# output: [0.019084517, -0.012082022, -0.03974377, ...
```


Safety Ratings

Harm Categories

- Derogatory
- Toxicity
- Violence
- Sexual
- Medical
- Dangerous



Safety Ratings

Harm Categories

- Derogatory
- Toxicity
- Violence
- Sexual
- Medical
- Dangerous

Harm Probabilities

- NEGLIGIBLE
- LOW
- MEDIUM
- HIGH

Resources

Notebook with all code snippets + links at bit.ly/devfest23_palm

1. **PaLM API**
`g.co/palm`
2. **Bard Chatbot**
`bard.google.com`
3. **MakerSuite for Prototyping**
`makersuite.google.com`
4. **Google Tutorials**
`developers.generativeai.google/tutorials`

Gemini

Generalized Multimodal
Intelligence Network
Comparable to GPT-4



**Embrace
Safely**




```
// You'll need  
// com.google.  
listRef.listAll  
.addOr  
prefixes.  
// All  
// You  
}  
it  
each { item  
the items  
}  
}
```

devfest



ChengCheng Tan

ccstan99@gmail.com

 cheng2-tan

 @cheng2_tan

 Google Developer Groups

