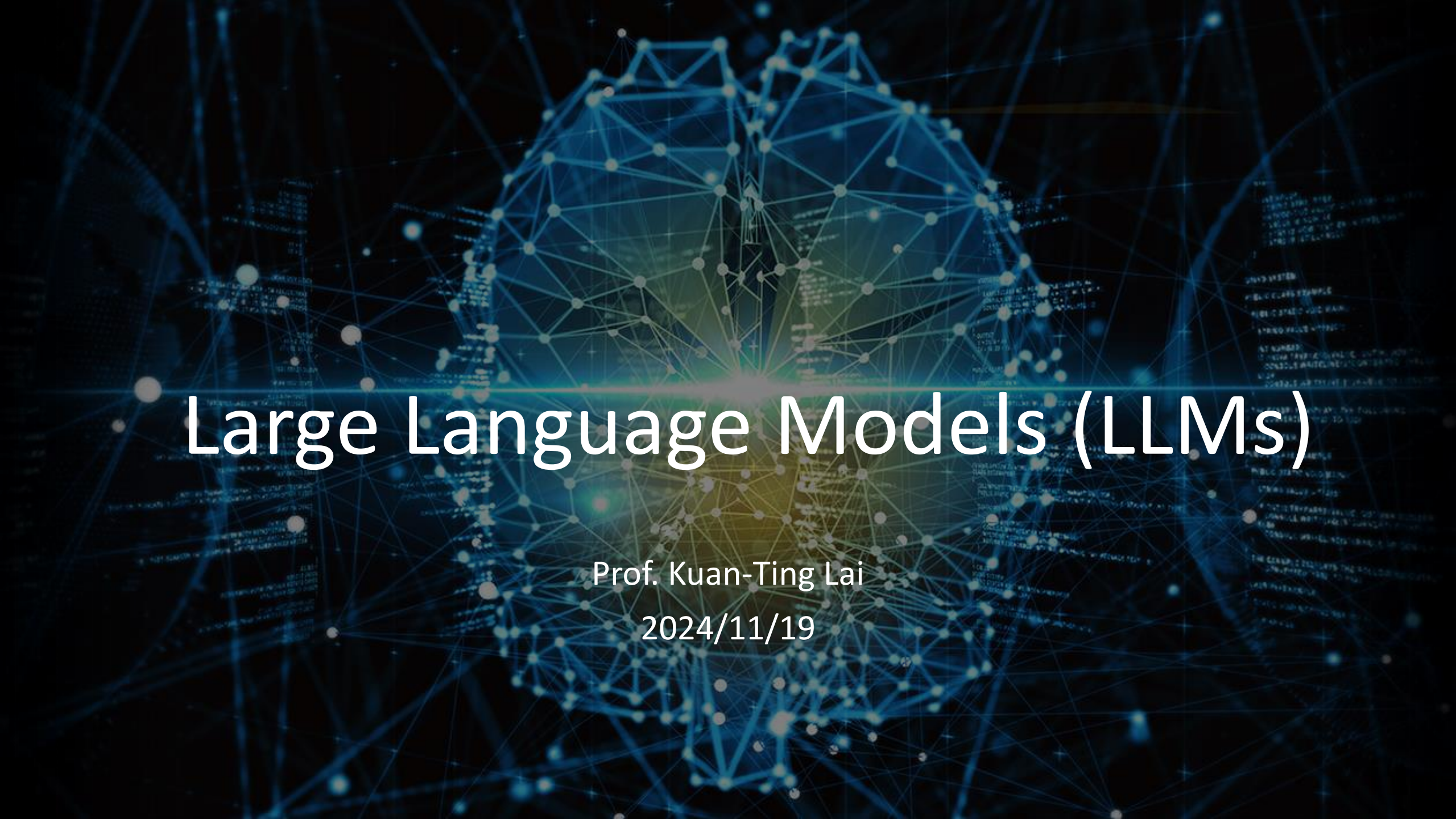


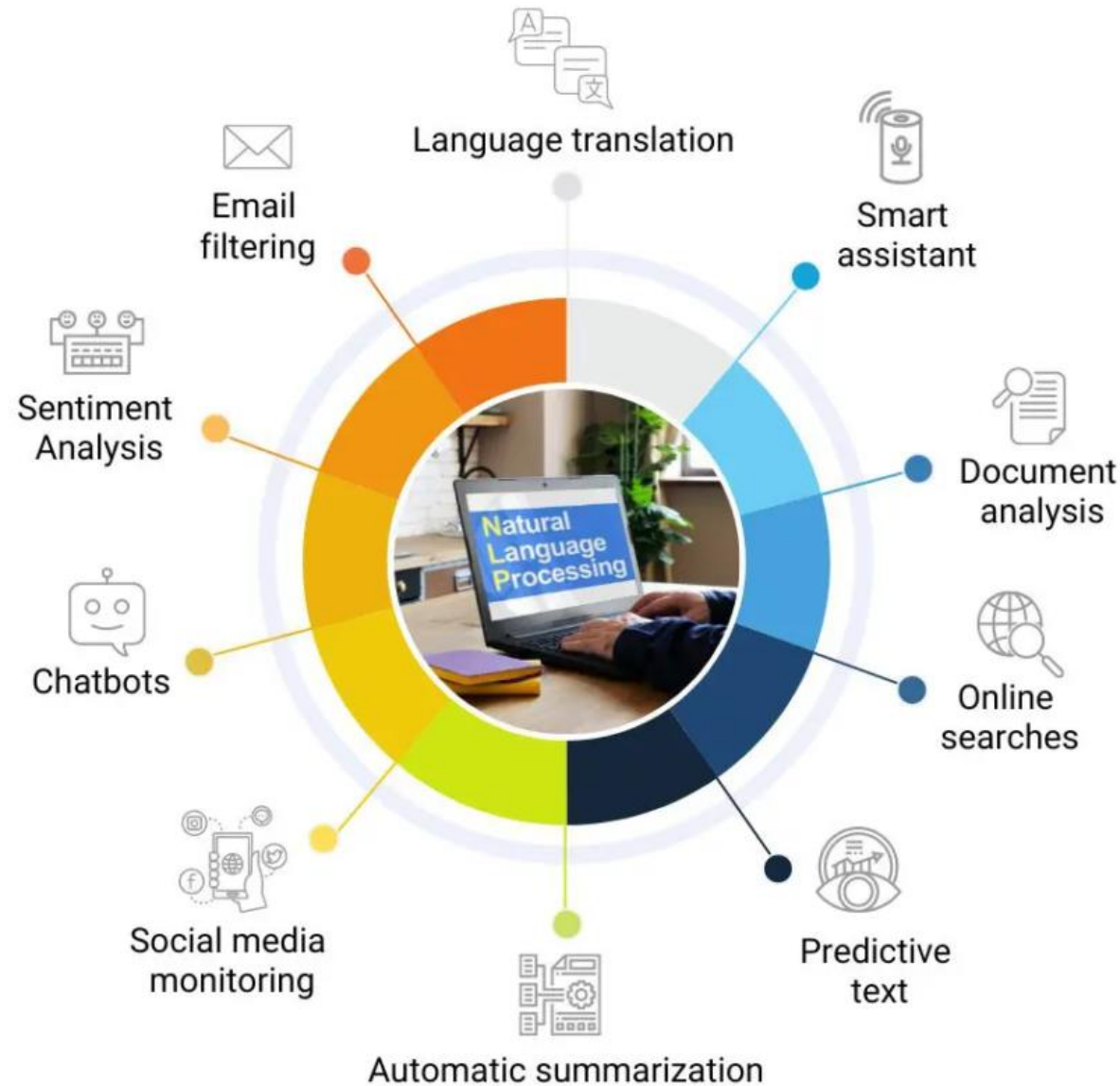


# Large Language Models (LLMs)

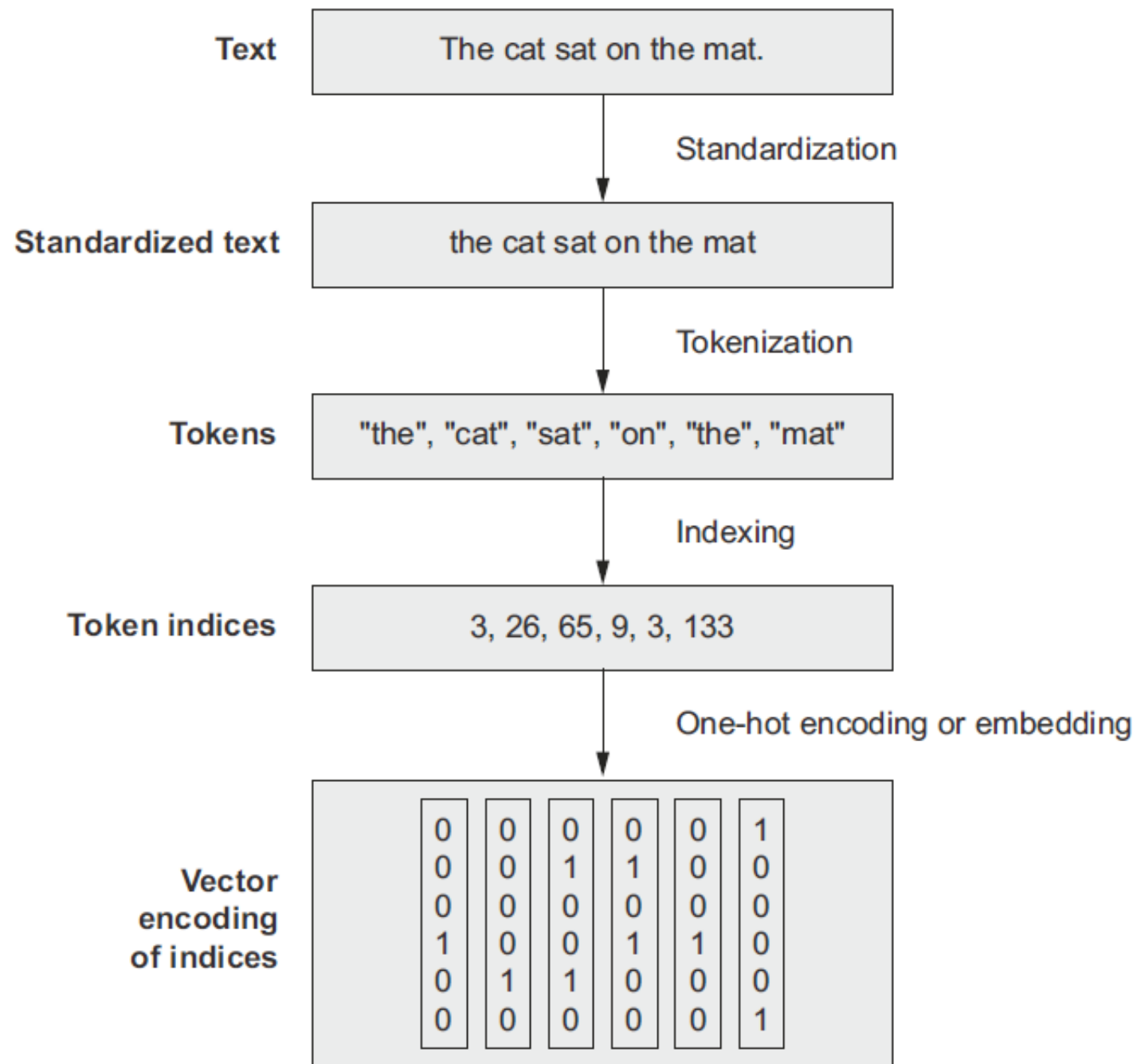
Prof. Kuan-Ting Lai

2024/11/19

# Natural Language Processing (NLP)

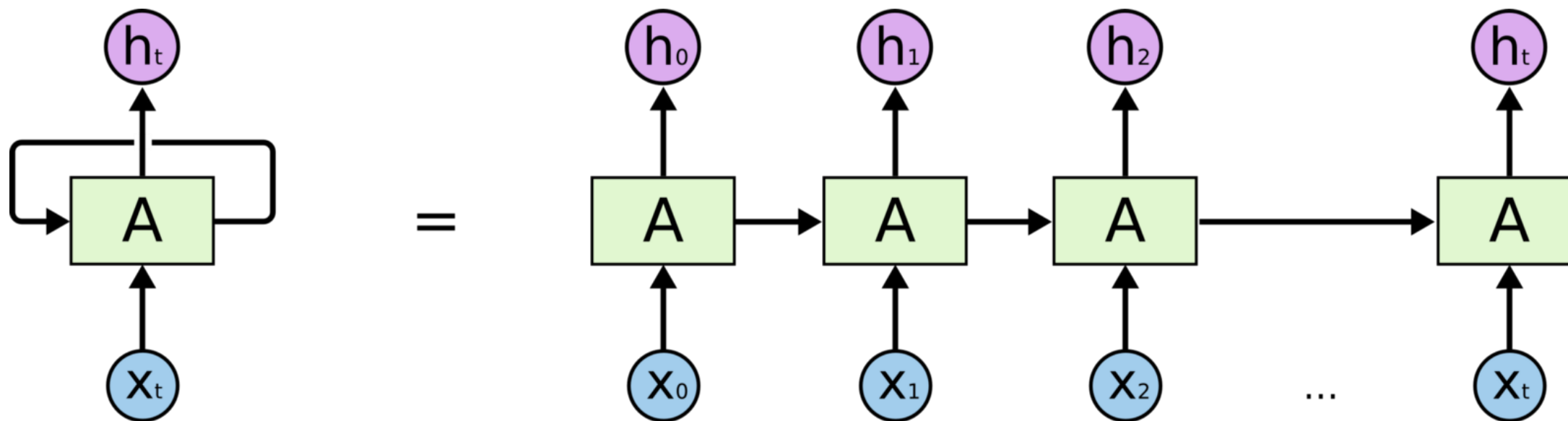


# NLP Preprocessing

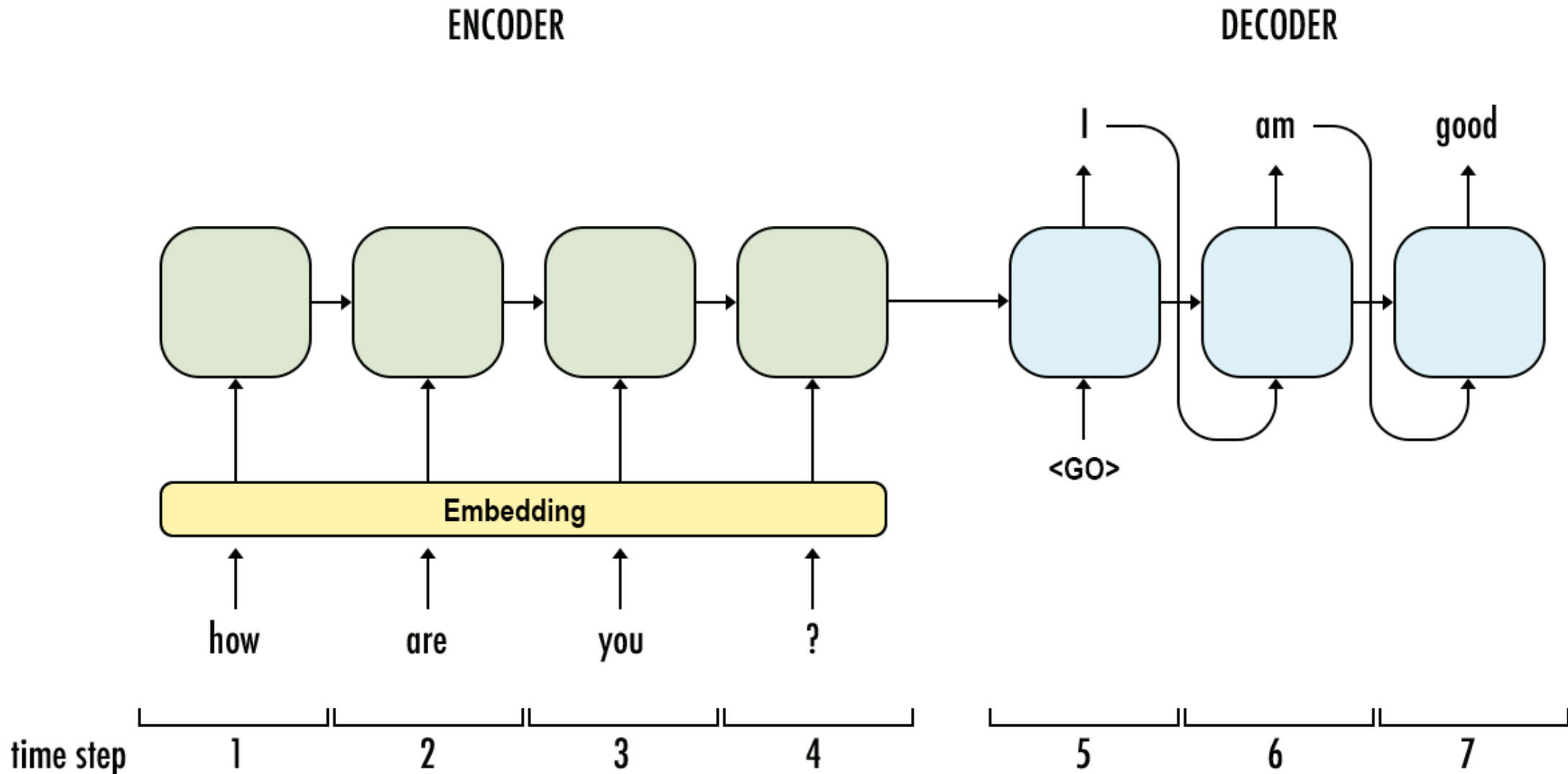


# Recurrent Neural Networks (RNNs)

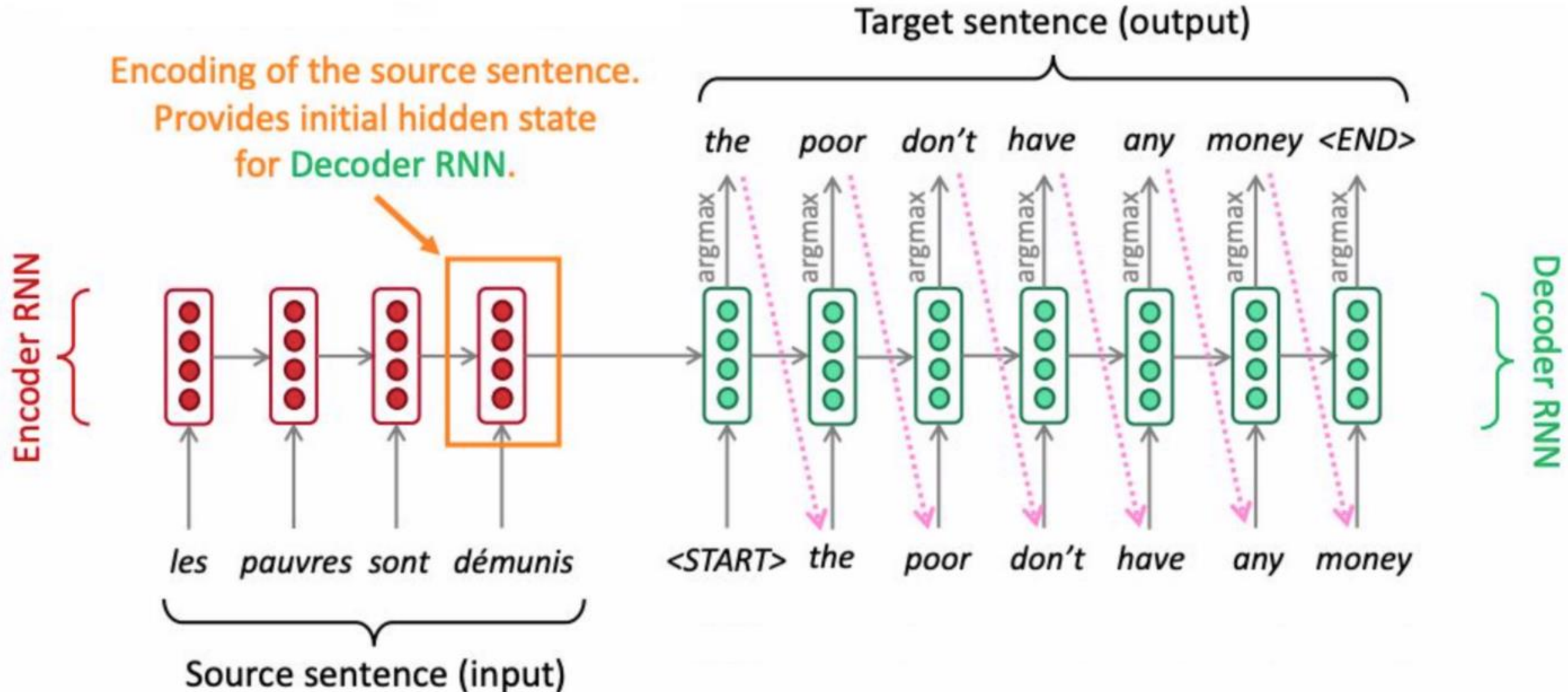
- Effective for speech recognition, language modeling, translation



# Sequence-to-Sequence (Seq2Seq)



# Seq2seq Language Translation



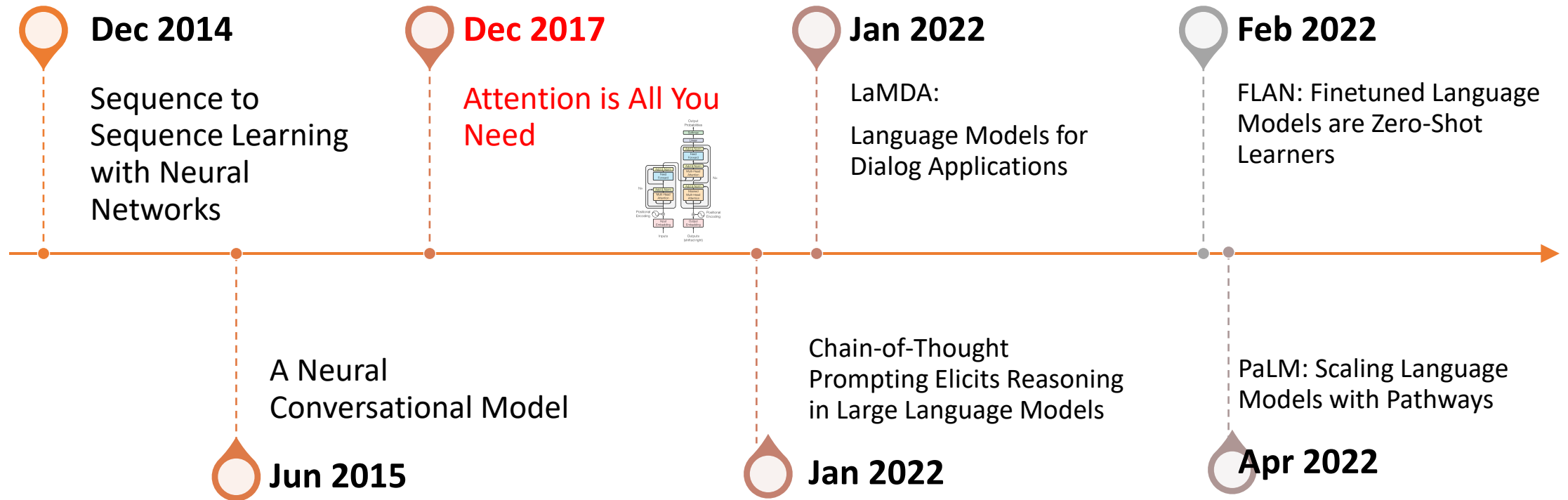
# Ed H. Chi (紀懷新)

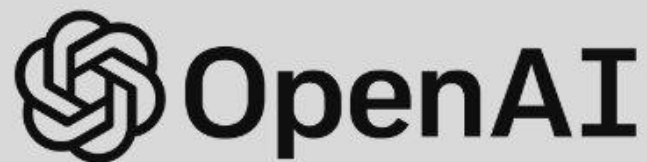
- Google Bard靈魂推手來自台灣！  
紀懷新解密Bard：AI有哪些限制？怎麼訓練(2023/8/17)

**學歷：**在六年半內取得明尼蘇達大學（University of Minnesota）學士、碩士和博士等3個學位。他獲選為電腦協會院士（ACM Fellow）和人機互動學會院士（CHI Academy），也因在資訊視覺化研究貢獻獲頒20年的Test of Time Award。他的相關研究與言論曾被《經濟學人》、《時代雜誌》、《洛杉磯時報》和《美聯社》等媒體報導和引用。



# How Google Created LLMs





## Attention Is All You Need

**Ashish Vaswani\***  
Google Brain  
avaswani@google.com

**Noam Shazeer\***  
Google Brain  
noam@google.com

**Niki Parmar\***  
Google Research  
nikip@google.com

**Jakob Uszkoreit\***  
Google Research  
usz@google.com

**Llion Jones\***  
Google Research  
llion@google.com

**Aidan N. Gomez\* †**  
University of Toronto  
aidan@cs.toronto.edu

**Łukasz Kaiser\***  
Google Brain  
lukaszkaizer@google.com

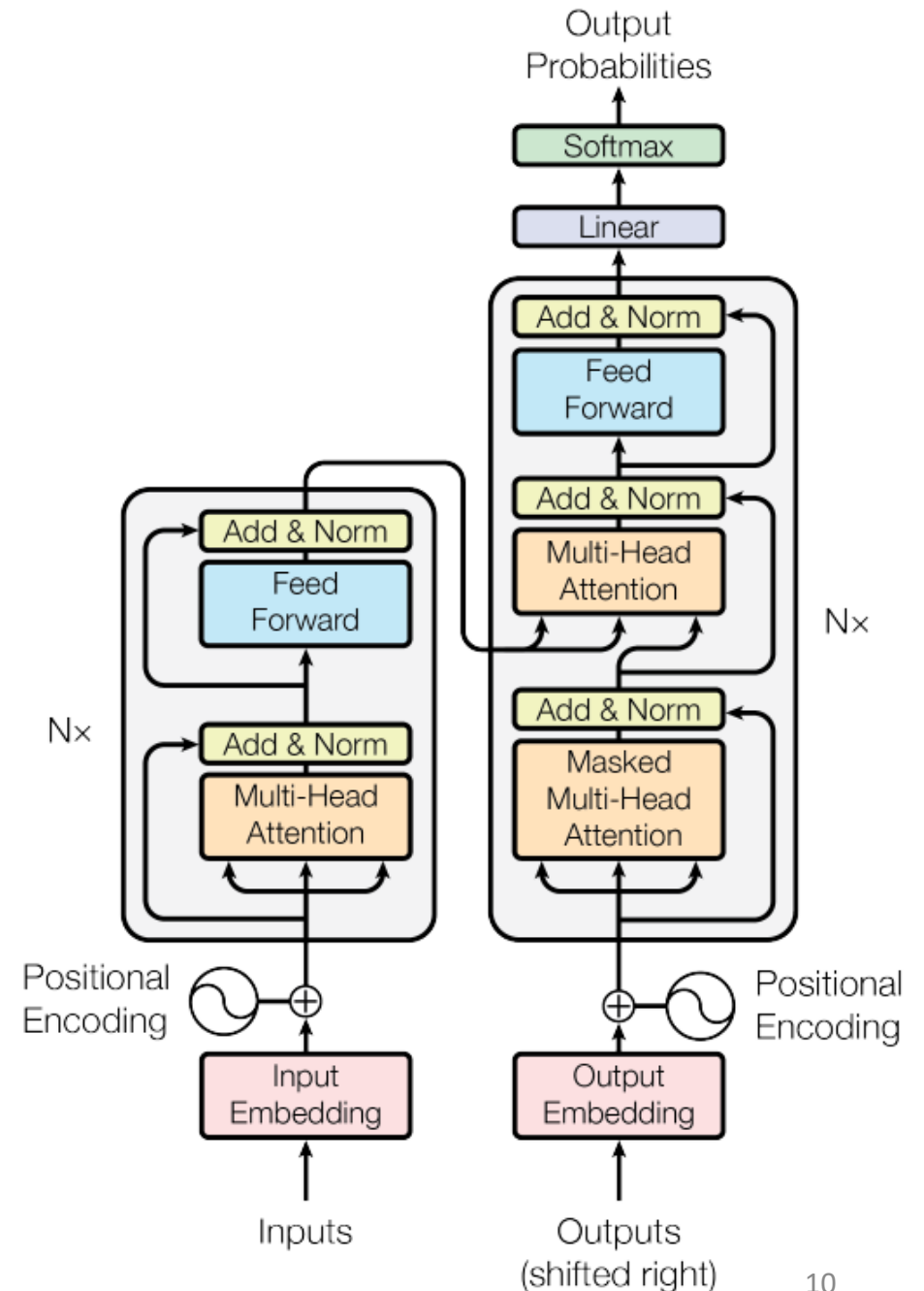
**Illia Polosukhin\* ‡**  
illia.polosukhin@gmail.com

<https://www.wired.com/story/eight-google-employees-invented-modern-ai-transformers-paper/>

Jakob  
Uszkoreit

# The Transformer Model

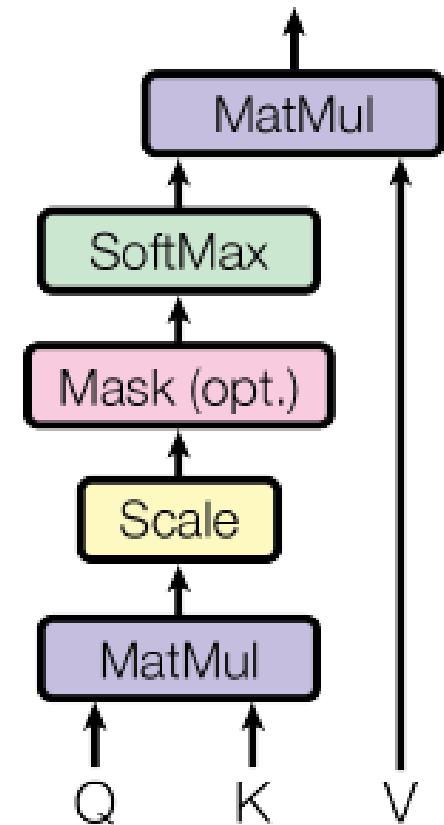
- Multi-head Attentions
  - Self-attention in encoders
  - Masked Self-attention in decoders
  - Encoder-decoder attention
- Encoder-decoder architecture
- Positional encoding



# Attention Module in Transformer

- Query (Q), Key (K), Value (V) attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



# Visualizing Attention

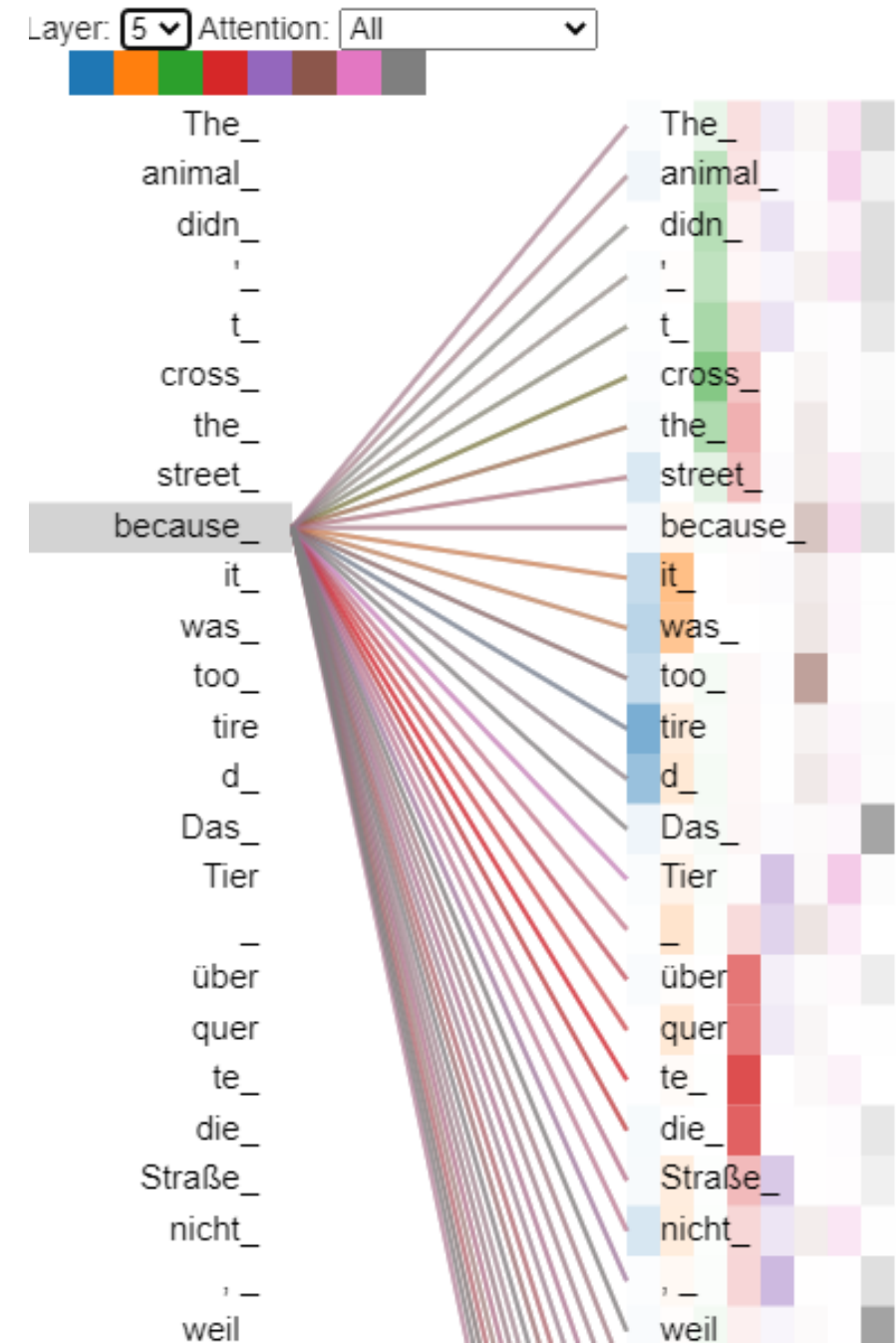
- Tensor2Tensor Notebook

[https://colab.research.google.com/github/tensorflow/tensor2tensor/blob/master/tensor2tensor/notebooks/hello\\_t2t.ipynb](https://colab.research.google.com/github/tensorflow/tensor2tensor/blob/master/tensor2tensor/notebooks/hello_t2t.ipynb)

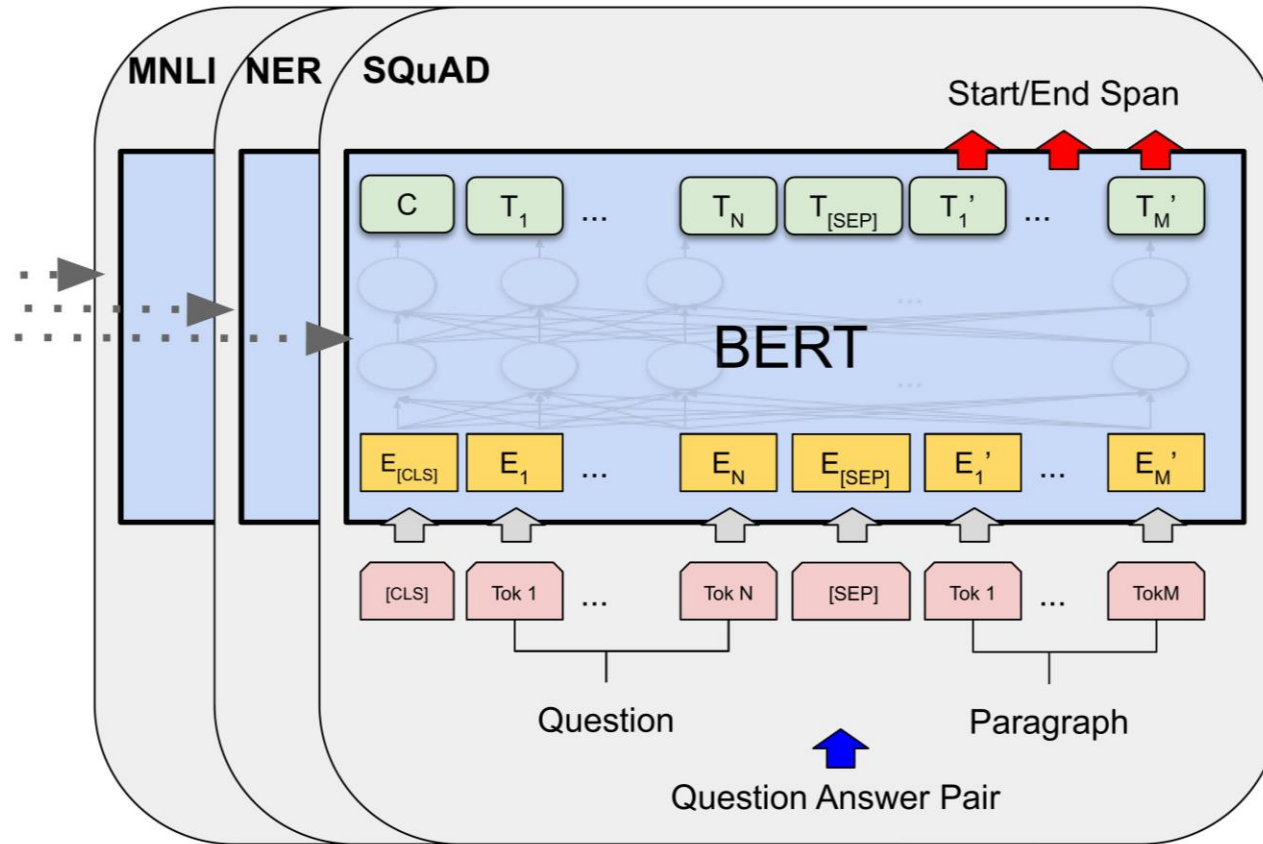
Inputs: The animal didn't cross the street because it was too tired



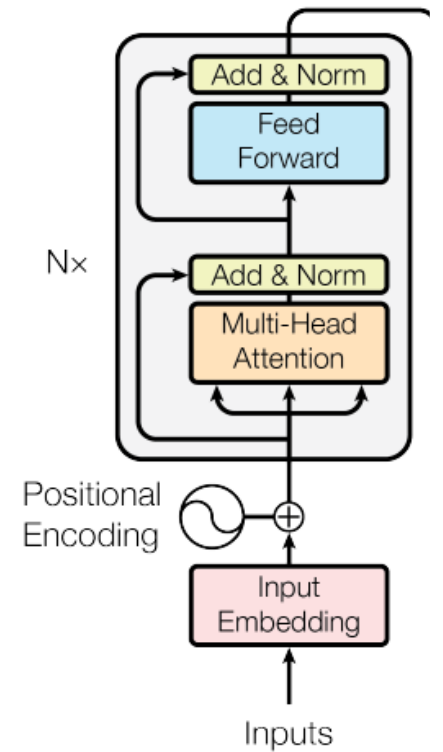
Outputs: Das Tier überquerte die Straße nicht, weil es zu müde war, weil es zu müde war.



# Bidirectional Encoder Representations from Transformers (BERT)

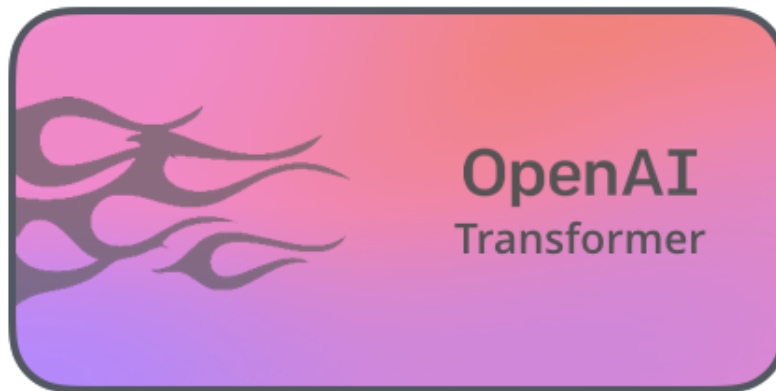


Fine-Tuning

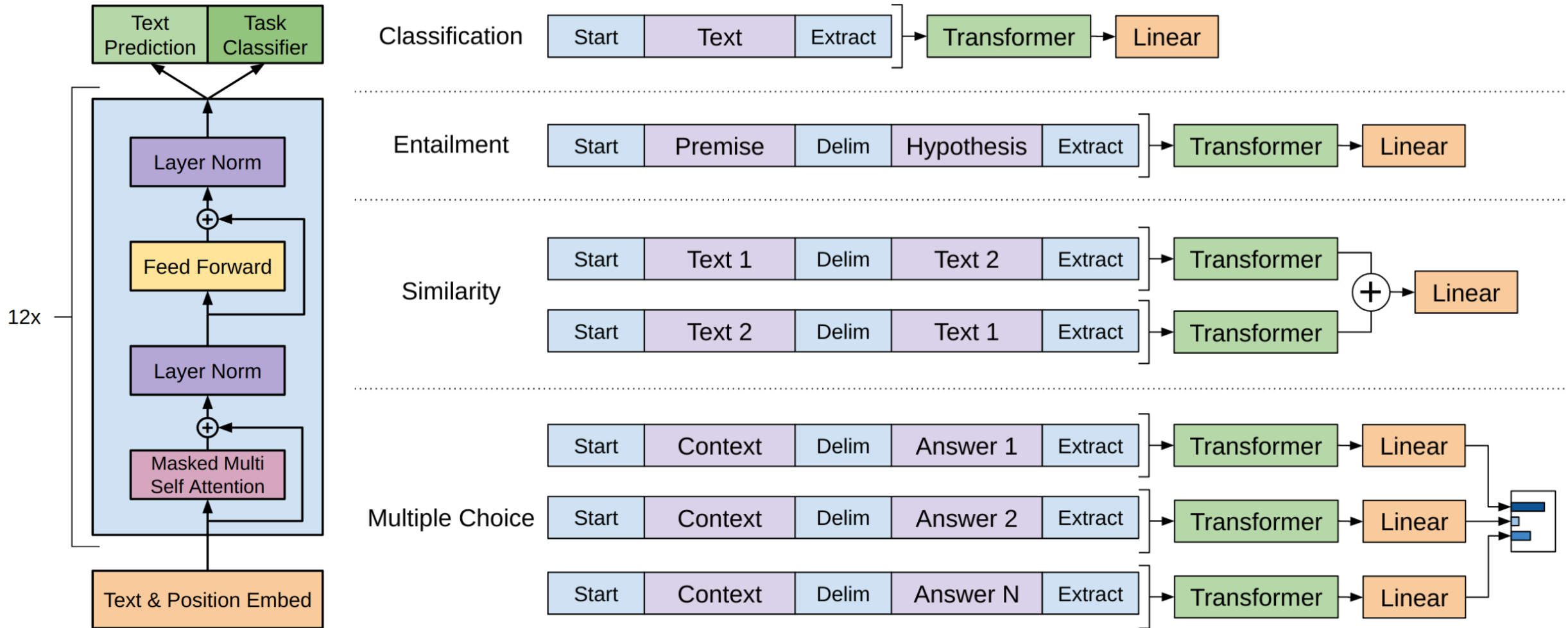


# OpenAI GPT: Pre-training Transformer Decoders

- Unsupervised pre-train transform decoders for predicting the next word (GPT: Generative Pre-Training)
- Use 12 Transformer decoders in GPT-1
  - GPT-1: [Improving Language Understanding with Unsupervised Learning \(2018\)](#)
  - GPT-2: [Better Language Models and Their Implications \(2019\)](#)
  - GPT-3: [Language Models are Few-Shot Learners \(2020\)](#)



# OpenAI GPT for Different Tasks



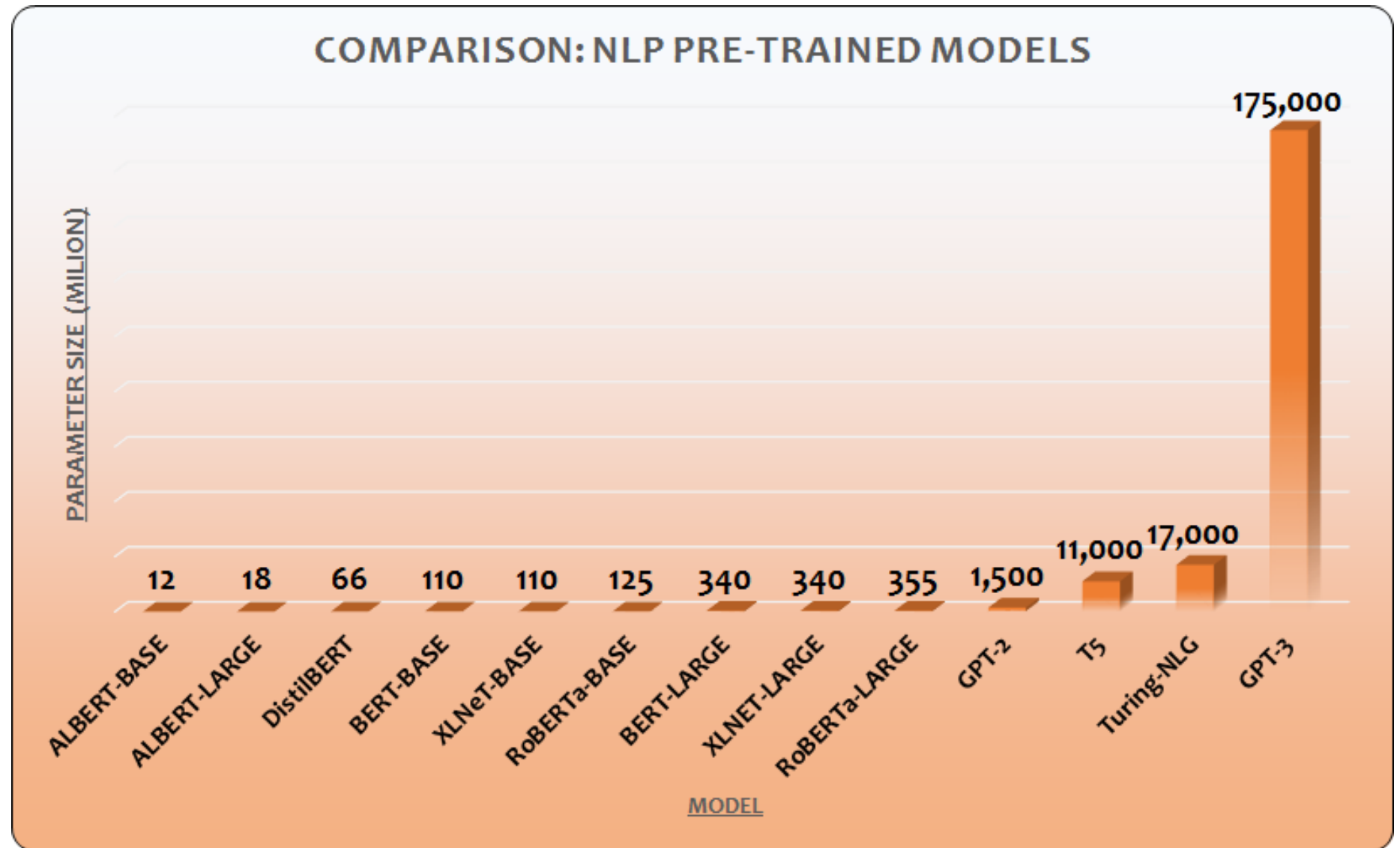
# OpenAI GPT-2

- Pre-trained using 40GB of Internet text
- Scale-up of GPT with 10X parameters trained with 10X data
- Other tricks
  - Layer normalization was moved to the input of each sub-block
  - An additional layer normalization was added after the final self-attention block

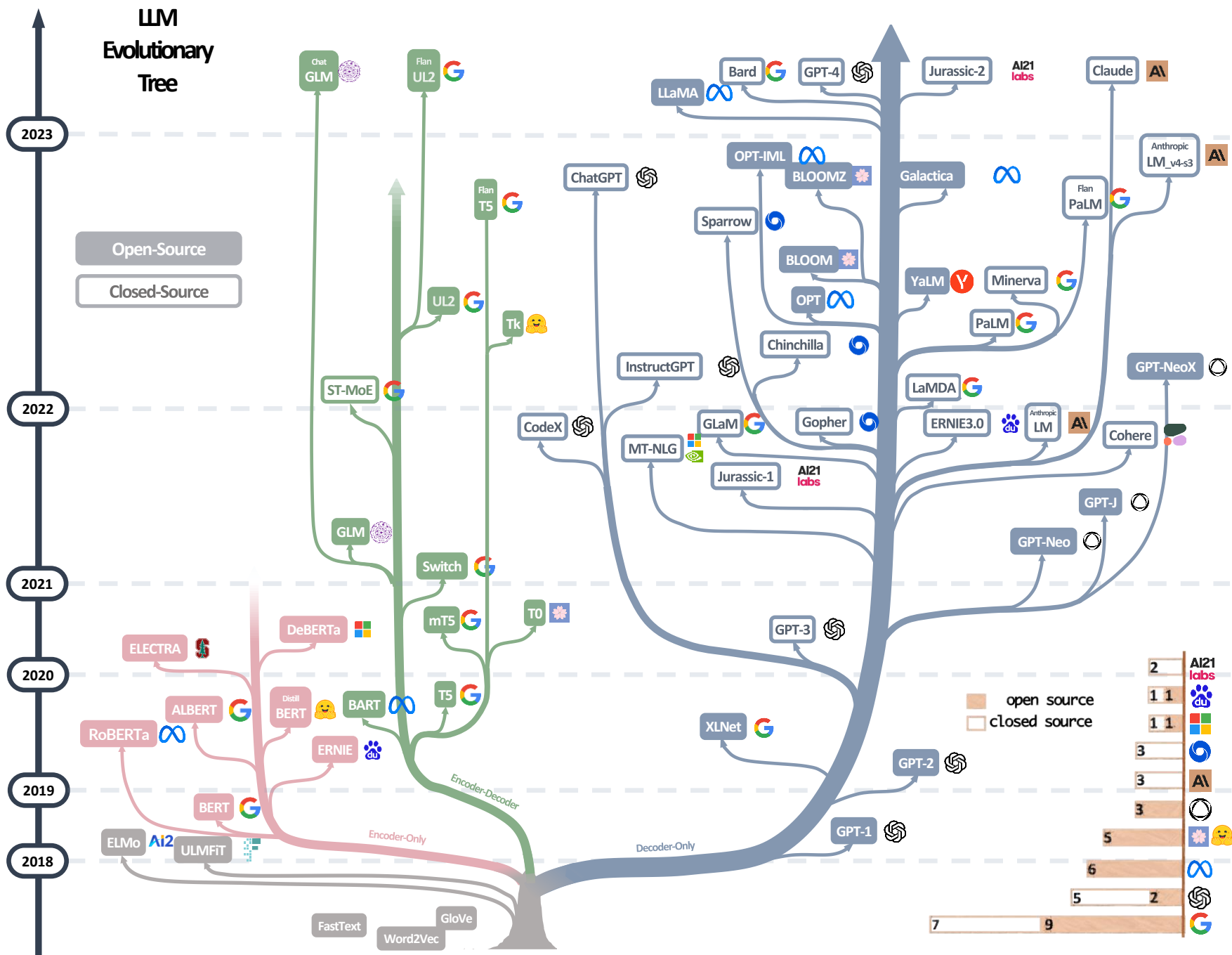
| Parameters | Layers | $d_{model}$ |
|------------|--------|-------------|
| 117M       | 12     | 768         |
| 345M       | 24     | 1024        |
| 762M       | 36     | 1280        |
| 1542M      | 48     | 1600        |

# Size does Matter! GPT-3

- 175 Billion Parameters!
- $175 \times 4 = 700\text{GB}$
- 55 years and \$4,600,000 to train - even with the lowest priced GPU cloud on the market.



# 2023 LLM Evolutionary Tree



# LLM Architecture Comparison

| Model          | Architecture    | Parameters | Context Window | Key Features                                                    |
|----------------|-----------------|------------|----------------|-----------------------------------------------------------------|
| GPT-4o         | Decoder-only    | 128B       | 128k tokens    | High quality, moderate price, good performance, and low latency |
| Llama 3.2 1B   | Decoder-only    | 1B         | 128k tokens    | Fastest output speed, low price                                 |
| Gemini 1.5 Pro | Decoder-only    | 2B         | 2M tokens      | Large context window, moderate price, good performance          |
| Claude 3.5     | Decoder-only    | 200B       | 200k tokens    | High quality, moderate price, good performance                  |
| Mistral 3B     | Decoder-only    | 3B         | 128k tokens    | Low price, good performance                                     |
| Flan-UL2       | Encoder-Decoder | 20B        | 512 tokens     | Good for fine-tuning, versatile applications                    |
| Bard           | Encoder-Decoder | 13B        | 512 tokens     | High quality, good for conversational AI                        |
| Bloom          | Decoder-only    | 176B       | 2048 tokens    | Open-source, multilingual capabilities                          |

# LLM Leaderboard

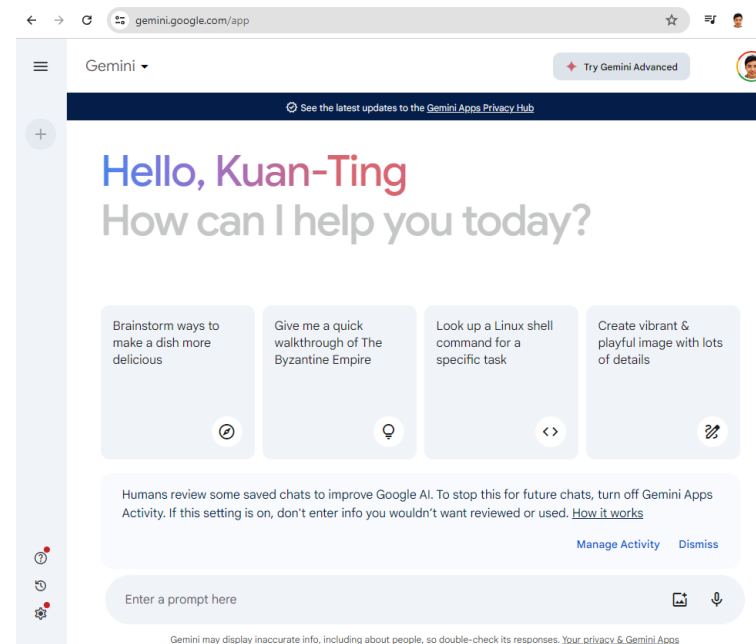
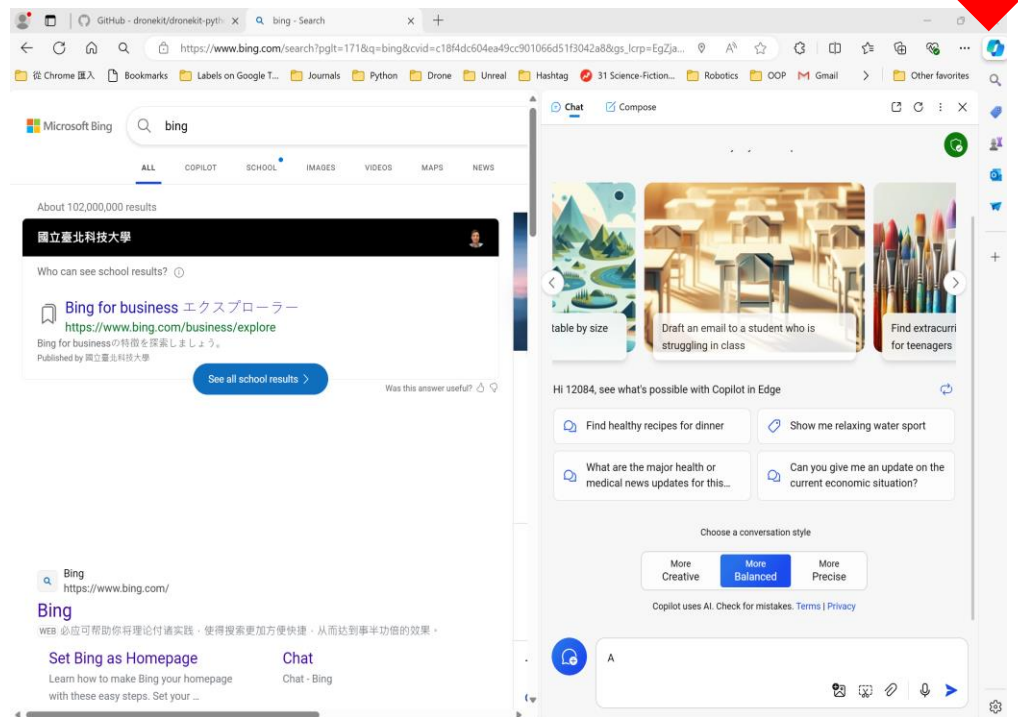
- <https://artificialanalysis.ai/leaderboards/models>

|                          | FEATURES ↗                                                                                  |                  | QUALITY ↗                                | PRICE ↗                           | OUTPUT TOKENS/S ↗           | LATENCY ↗                          |                                                     |
|--------------------------|---------------------------------------------------------------------------------------------|------------------|------------------------------------------|-----------------------------------|-----------------------------|------------------------------------|-----------------------------------------------------|
| MODEL ↕                  | CREATOR ↕                                                                                   | CONTEXT WINDOW ↕ | QUALITY INDEX<br><i>Normalized avg</i> ↕ | BLENDED<br><i>USD/1M Tokens</i> ↕ | MEDIAN<br><i>Tokens/s</i> ↕ | MEDIAN<br><i>First Chunk (s)</i> ↕ | FURTHER ANALYSIS                                    |
| o1-preview               |  OpenAI    | 128k             | 85                                       | \$26.25                           | 35.6                        | 28.82                              | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| o1-mini                  |  OpenAI    | 128k             | 82                                       | \$5.25                            | 74.0                        | 13.64                              | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| Claude 3.5 Sonnet (Oct)  | ANTHROPIC                                                                                   | 200k             | 80                                       | \$6.00                            | 55.3                        | 0.87                               | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| Gemini 1.5 Pro (Sep)     |  Google    | 2m               | 80                                       | \$2.19                            | 59.2                        | 0.77                               | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| GPT-4o                   |  OpenAI    | 128k             | 77                                       | \$4.38                            | 83.7                        | 0.45                               | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| Claude 3.5 Sonnet (June) | ANTHROPIC                                                                                   | 200k             | 77                                       | \$6.00                            | 55.4                        | 0.93                               | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| GPT-4o (May '24)         |  OpenAI  | 128k             | 77                                       | \$7.50                            | 91.6                        | 0.45                               | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| Qwen2.5 72B              |  Alibaba | 131k             | 75                                       | \$0.39                            | 47.2                        | 0.55                               | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |
| GPT-4 Turbo              |  OpenAI  | 128k             | 74                                       | \$15.00                           | 36.5                        | 0.64                               | <a href="#">↗ Model</a> <a href="#">↗ Providers</a> |

# Free Lunch

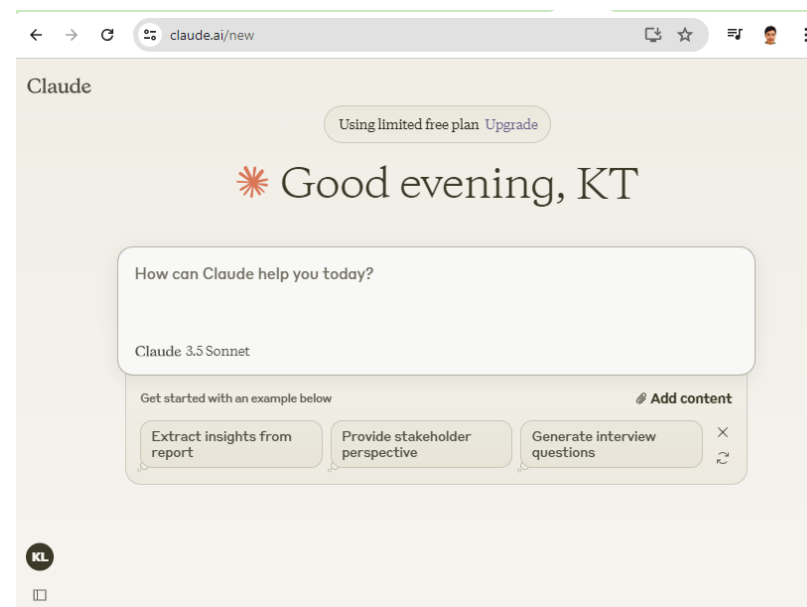
Click

# Microsoft Copilot



# Google

# Gemini



# ANTHROPIC

# Claude

# ChatGPT Plugin!

Click

A screenshot of the ChatGPT website's 'GPTs' page. The browser address bar shows 'chatgpt.com/gpts'. The sidebar on the left contains a list of GPTs categorized by 'Today', 'Previous 7 Days', and 'Previous 30 Days'. The 'Explore GPTs' button is highlighted. The main content area features the title 'GPTs', a description, a search bar, and a 'Featured' section with four GPTs: 'Kraftful product coach', 'Grimoire', 'Scholar AI', and 'Books'.

chatgpt.com/gpts

My GPTs + Create

## GPTs

Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.

Top Picks Writing Productivity Research & Analysis Education Lifestyle Programming

### Featured

Curated top picks from this week



**Kraftful product coach**  
Ask about best practices for product managers. Get top gurus' advice on product management....  
By kraftful.com



**Grimoire**  
Code Wizard Copilot 🧙 Build a website with a sentence. 20+ Hotkeys for quick coding flows....  
By gptavern.mindgoblinstudios.com



**Scholar AI**  
AI Innovator — search and review 200M+ scientific papers, patents, and books. Research literature,...  
By scholarai.io



**Books**  
Your AI guide in the world of literature and reading, utilizing Web Browsing to provide accurate,...  
By Josh Villocido

# Custom ChatGPT

The screenshot shows the ChatGPT web interface. On the left is a sidebar with a list of recent chats and a section for 'Add Team workspace'. The main area displays the 'ChatGPT 4o' header and a dropdown menu. A red arrow labeled (1) points to the user profile icon in the top right corner. Another red arrow labeled (2) points to the 'My GPTs' option in the dropdown menu, which is also highlighted with a red rectangular box. Below the menu, the OpenAI logo is centered, and four suggested prompts are shown in rounded boxes: 'Study vocabulary', 'Experience Seoul like a local', 'Email for plumber quote', and 'Python script for daily email reports'. At the bottom, there is a text input field with the placeholder 'Message ChatGPT' and a send button. A disclaimer at the very bottom states 'ChatGPT can make mistakes. Check important info.'

chatgpt.com

ChatGPT 4o

(1)

(2)

My plan

My GPTs

Customize ChatGPT

Settings

Log out

Study vocabulary

Experience Seoul like a local

Email for plumber quote

Python script for daily email reports

Message ChatGPT

ChatGPT can make mistakes. Check important info.

Create

Configure



Name

AloT Lab Assistant

Description

Knowledgeable AloT Lab assistant answering student queries

Instructions

cooperated with Qualcomm and Amazon AWS to offer industry-university co-taught courses. In terms of international academic activities, he has assisted many journals and conferences in the review of papers, such as IEEE Transactions on Multimedia, IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), IEEE IoT Journal, etc. Since 2019, he has served as the co-host of the sub-project of the "Key Technology Research and Development Project for Self-Driving Subsystems" of the Ministry of Science and Technology.

Conversation starters

Could you introduce the research of the lab?

×

How can I apply for joining the lab?S

×

請問目前實驗室還有名額嗎?

×

實驗室學長姐畢業後就業狀況如何?

×

想加入實驗室需要什麼條件?

×

×

Knowledge

If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled



2022\_AI\_Wings\_early\_acc...  
PDF



2018-ACMMM-VIVID.pdf  
PDF

Preview



## AloT Lab Assistant

Knowledgeable AloT Lab assistant answering student queries

Could you introduce the research of the lab?

How can I apply for joining the lab?S

請問目前實驗室還有名額嗎?

實驗室學長姐畢業後就業狀況如何?



Message AloT Lab Assistant



?

# 實驗室AI小助理

← → ↺

chatgpt.com/g/g-9SS1JyEAQ-aiot-lab-assistant/c/e560d75c-f428-4975-a6eb-26342d45dab3

📄

AIoT Lab Assistant ▾

ChatGPT

AIoT Lab Assistant

Canva

Explore GPTs

Today

實驗室研究方向

實驗室名額查詢

OpenAI Products Diagram

ChatGPT Papers Timeline

Cartoon Girl Smiling Actions

Previous 7 Days

ZNS Simulator Optimization Tool

Python Code UML Diagram

UML Diagram Python Code

Content Summary Request

Previous 30 Days

UML Diagram of Python

Python Code UML Diagram

Zone SSD Garbage Collection

Bluetooth wBMS Prototype Deve

生成式AI應用技巧

Drone System Integration Challe

Add Team workspace

📤 👤

實驗室的研究方向為何?

🧠

實驗室的研究方向主要集中在以下幾個領域：

1. 電腦視覺：這是賴教授的主要研究方向，涵蓋了許多電腦視覺算法和應用，例如動作辨識、物體檢測和追蹤等。

2. 無人機和物聯網 (IoT)：實驗室致力於開發智能無人機系統，例如AI Wings無人機系統，這是一個集成了物聯網技術和人工智慧的無人機平台。

3. 虛擬現實到現實學習 (Virtual-to-Real Learning)：這項研究結合虛擬現實和深度強化學習，目的是在虛擬環境中訓練模型，然後將其應用於現實世界。例如，VIVID是一個用於視覺深度學習的虛擬環境，能夠模擬和訓練無人機和自駕車的深度模型。

4. 深度強化學習和虛擬現實：這個領域的研究目的是利用先進的3D引擎生成真實感圖像，用於訓練深度神經網絡，以進行各種任務，例如自動駕駛、碰撞避免和圖像分割等。

這些研究方向不僅體現在理論研究上，還通過與業界的合作，例如與Qualcomm和Amazon AWS的產學合作課程，推動技術應用的實現。

Could you introduce the research of the Lab?

🧠

Certainly! The AIoT Lab at National Taipei University of Technology, founded by Dr. Kuan-Ting Lai, focuses on several cutting-edge research areas that integrate artificial intelligence, computer vision, Internet of Things (IoT), and drone technology. Here's an overview of the lab's key research directions:

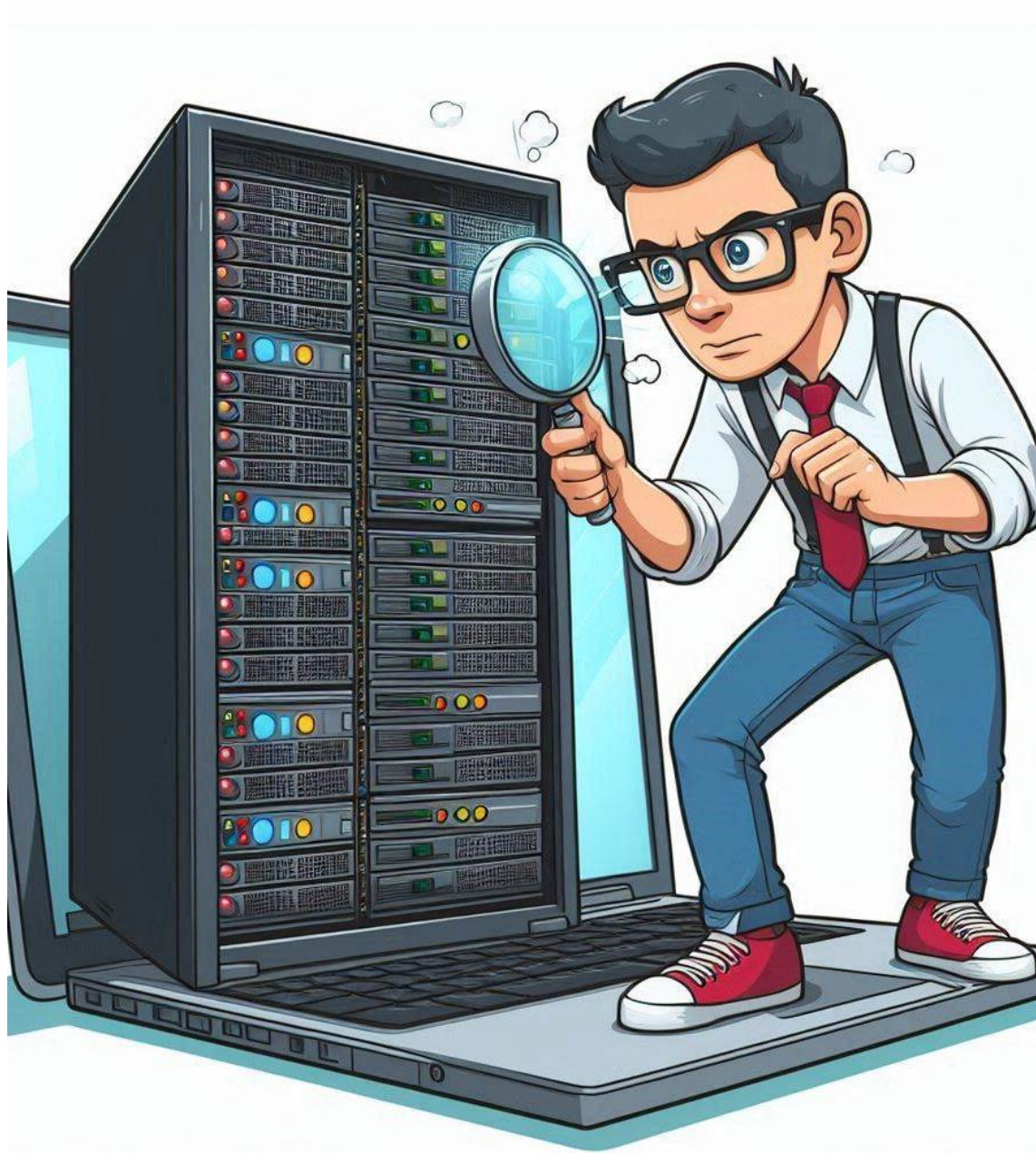
1. Computer Vision

↓

📎

Message AIoT Lab Assistant

📤

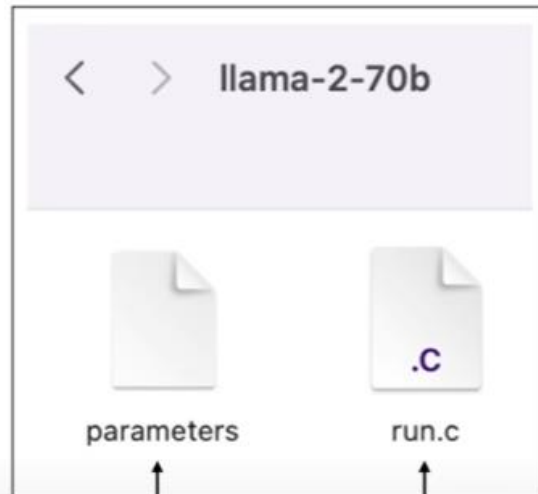


Install LLMs on  
Your PC

# Training is like a Mountain, Running is like a Feather (訓練重如泰山, 執行輕如鴻毛)

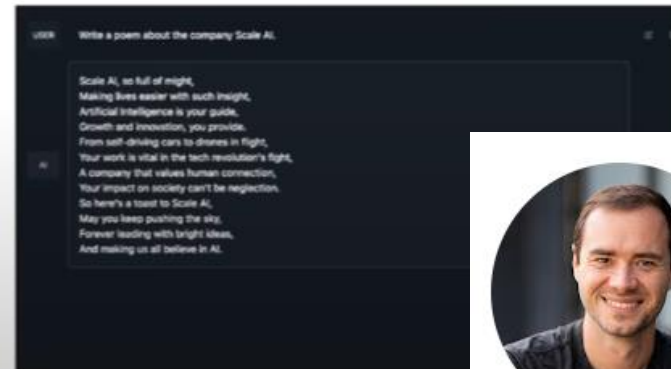
## Large Language Model (LLM)

MacBook



140GB

~500 lines  
of C code



**Andrej Karpathy**

@AndrejKarpathy · 512K subscribers · 15 videos

SuperThanks: very optional, goes to Eureka Labs. ...more

[eurekalabs.ai](https://eurekalabs.ai) and 4 more links

Subscribed

<https://www.youtube.com/@AndrejKarpathy>

# Tools for Installing LLM

- GPT4All

- Hugging Face



- LangChain



- Llama.cpp

- Ollama



# TAIDE: Taiwan LLM ([taide.tw](https://taide.tw))

[認識TAIDE](#)[團隊陣容](#)[學習資源](#)[最新動態](#)[訓練資料](#)[TAIDE模型](#)[中文](#) [EN](#)

Hi 你想要問什麼？

有台語諧音的YT旅標

當然，台語諧音的YT旅標：

全台灣好

## TAIDE計畫目的

除協助 AI 法制化規劃外，亦將成立 AI 產品與系統評測中心，並評估生成式 AI 對話引擎模型發展之語言模型先期應用，逐步建立 AI 應用之規範制度與驗測環境。

整體計畫目的



# Using GPT4All to Install TAIDE

GPT4All v3.0.0

Home  
Chats  
Models  
LocalDocs  
Settings

← Existing Models

## Explore Models

taide

Enter "taide"

ZoneTwelve/TAIDE-LX-7B-Chat-GGUF

Select "ZoneTwelve"

Created by ZoneTwelve.

- Published on Mon April 15, 2024.
- This model has 17 likes.
- This model has 845 downloads.
- More info can be found [here](#).

| File size | RAM required | Parameters | Quant | Type |
|-----------|--------------|------------|-------|------|
| 3.69 GB   | ?            | ?          | q4_0  |      |

Download

QuantFactory/TAIDE-LX-7B-GGUF

Created by QuantFactory.

- Published on Thu April 25, 2024.
- This model has 1 likes.
- This model has 84 downloads.
- More info can be found [here](#).

| File size | RAM required | Parameters | Quant | Type |
|-----------|--------------|------------|-------|------|
| 3.69 GB   | ?            | ?          | q4_0  |      |

Download

NOMIC

+ New Chat

TODAY

AI助手協助台語問答，展現語言能力。

Taiwanese Hokkien language basics

Please provide me with

I am an A

LAST YEAR

Naive Bayes Classifier

Prompts user, then

# Select Your Model

ZoneTwelve/TAIDE-LX-7B-Chat-GGUF

LocalDocs

You

你會說台語嗎? 列出幾個常用台語



GPT4All ZoneTwelve/TAIDE-LX-7B-Chat-GGUF

當然! 以下是幾種常見的台語詞彙和短語:

1. 你好 - 嗨 (Hōo-ē)
2. 再見 - 後，再會 (Lāi tsiānn khuan)
3. 謝謝 - 多謝 (Tō-si)
4. 不客氣 - 沒事 (Bú khiit-ling)
5. 對不起 - 抱歉 (Kò-ài hō)
6. 是的/不是 - 是 (Sī) 或否 (É)
7. 你好嗎? - 安頓了沒? (Ān tūn lóng mǎi)

You

列出台灣有台語諧音的YT地標



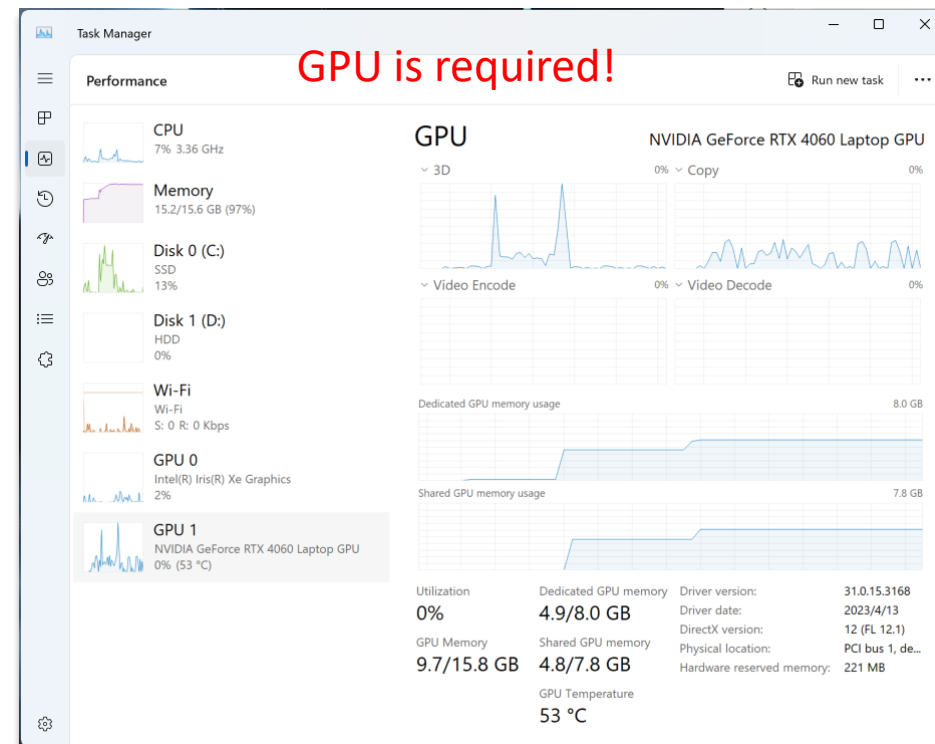
GPT4All ZoneTwelve/TAIDE-LX-7B-Chat-GGUF

以下是幾個具有台語諧音的台灣YouTube頻道或網紅:

1. 阿滴英文 - Derek (Dē) Cheng
2. 阿滴中文 - Derek Cheng (與阿滴英文同人)
3. 九妹 KIKI - 九妹的個人生活、旅遊和美食相關影片
4. 老皮 PAPA - 以台語講授各種主題，包括歷史、科學和文化等。
5. 陳沂軒 Yi-Hsuan Chen - 分享自己的藝術創作和手繪動畫作品。
6. 黃氏兄弟 HUANG BROTHERS - 以台語演唱流行歌曲和搞笑影片。
7. 大樹哥 TA-SHIH (Joe Tsai) - 分享自己種植樹木、園藝和生活經驗的頻道。
8. 林進榮 IIN JIN RONG - 以台語講授各種主題，包括投資理財、股票分析等。

Send a message...

## GPU is required!



# Meta's LLaMA 3



GPT4All v3.0.0

- Home
- Chats
- Models
- LocalDocs
- Settings

+ New Chat

TODAY

Taiwanese Hokkien langua...

Please provide me with

I am an A

LAST YEAR

Naive Bayes Classifier

Prompts user, then



Llama 3 Instruct

LocalDocs

You

你會說台語嗎? 列出幾個常用台語



GPT4All Llama 3 Instruct

I'd be happy to chat with you in Taiwanese Hokkien (Tâi-lû) ! 🍵 Here are some common phrases and words:

## Greetings

- Ah-bah (阿爸): Hello, older brother/sister
- Ah-ma (阿媽): Hello, older sister/mother
- Kau-im (講意): Good morning/afternoon/evening

## Basic Phrases

- Chī-phi (吃飯): Eat rice
- Má-chi (馬車): Go away/go home
- Bóo-kah (不卡): Don't worry
- Ah-hí (阿嬤): Thank you, auntie
- Kòo-sa (好事): Good thing

## Food

- Chēng-m̄ng (整命): Rice bowl with various toppings
- Mee-gōo (麵狗): Noodles
- Bāng-kah (棒加): Fried dough fritters
- O-á-chī (烏仔吃): Eat oyster omelet

## Directions

- Tē-m̄ng (底門): Downstairs/underneath
- Lóo-tiān (樓天): Upstairs/above
- Kòo-kah (好卡): Straight ahead
- Hái-lâi (海來): Come from the sea/coast

## Miscellaneous

- Ah-hí, ah-bah! (阿嬤, 阿爸!) Oh no, oh dear!
- Má-chi kóo-sa! (馬車好事!) Go away and have a good time!

Please note that Taiwanese Hokkien is a dialect with various regional variations. The phrases I listed are in the Min-nan dialect, which is commonly spoken in Taiwan.

Feel free to ask me any questions or practice your Tai-lû skills with me! 🍵

Send a message...



NOMIC

# Building Your Own LINE ChatBot



# OpenAI Platform (for Developers)

The screenshot shows the OpenAI Platform developer documentation page in a web browser. The browser's address bar displays 'platform.openai.com/docs/overview'. The page features a top navigation bar with links to 'Dashboard', 'Docs', 'API reference', 'Log in', and a 'Sign up' button. A left sidebar contains a search bar and a list of navigation items under 'GET STARTED' (Overview, Quickstart, Concepts, Models, Libraries, Changelog) and 'CAPABILITIES' (Text generation, Vision, Function calling, JSON mode, Advanced usage). Below these are 'ENDPOINTS' (Chat Completions, Fine-tuning, Batch, Image generation), 'Forum', and 'Help'. The main content area is titled 'OpenAI developer platform' and includes a 'Developer quickstart' section with a 5-minute timer and a code snippet for using the OpenAI API. Below this is a 'Meet the models' section with cards for 'GPT-4o' and 'GPT-4o mini' (marked as 'New'). The GPT-4o card describes it as a high-intelligence flagship model for complex tasks, while the GPT-4o mini card describes it as an affordable and intelligent small model for fast, lightweight tasks. Both cards list input/output types, context length, and pricing.

Overview - OpenAI API

platform.openai.com/docs/overview

OpenAI Platform

Dashboard Docs API reference Log in Sign up

Search **CTRL K**

**GET STARTED**

- Overview
- Quickstart
- Concepts
- Models
- Libraries
- Changelog

**CAPABILITIES**

- Text generation
- Vision
- Function calling
- JSON mode
- Advanced usage

**ENDPOINTS**

- Chat Completions
- Fine-tuning
- Batch
- Image generation

Forum Help

## OpenAI developer platform

### Developer quickstart

Set up environment and make your first API request in minutes

🕒 5 min

```
1 from openai import OpenAI
2 client = OpenAI()
3 completion = client.chat.completions.create(
4     model="gpt-4o",
5     messages=[
6         {"role": "user", "content": "write a haiku about ai"}
7     ]
8 )
```

## Meet the models

### GPT-4o

Our high-intelligence flagship model for complex, multi-step tasks

Text and image input, text output

128k context length

Input: \$5 | Output: \$15\*

### GPT-4o mini

**New**

Our affordable and intelligent small model for fast, lightweight tasks

Text and image input, text output

128k context length

Input: \$0.15 | Output: \$0.60\*

\* prices per 1M tokens

# OpenAI Assistant API

The screenshot shows a web browser window with the OpenAI Platform documentation. The page title is "Assistants quickstart - OpenAI". The URL is "platform.openai.com/docs/assistants/quickstart". The page has a navigation bar with "Dashboard", "Docs", "API reference", "Log in", and "Sign up". A left sidebar contains a list of topics: Chat Completions, Fine-tuning, Batch, Image generation, Text to speech, Speech to text, Embeddings, Moderation, Completions (Legacy), ASSISTANTS, Overview, Quickstart (highlighted), Deep dive, Tools, What's new?, Migration guide, CHATGPT, Actions, and Release notes. The main content area is titled "Step 1: Create an Assistant". It explains that an Assistant represents an entity that can be configured to respond to a user's messages using parameters like model, instructions, and tools. Below this is a code block titled "Create an Assistant" with Python code to create an assistant named "Math Tutor" using the gpt-4o model and a code interpreter tool. The next section is "Step 2: Create a Thread", explaining that a Thread represents a conversation between a user and one or many Assistants. It includes a code block titled "Create a Thread" with Python code to create a thread. The final section is "Step 3: Add a Message to the Thread", explaining that messages are added as Message objects to the Thread.

Assistants quickstart - OpenAI

platform.openai.com/docs/assistants/quickstart

OpenAI Platform

Dashboard Docs API reference Log in Sign up

Chat Completions  
Fine-tuning  
Batch  
Image generation  
Text to speech  
Speech to text  
Embeddings  
Moderation  
Completions (Legacy)  
**ASSISTANTS**  
Overview  
Quickstart  
Deep dive  
Tools  
What's new?  
Migration guide  
CHATGPT  
Actions  
Release notes

## Step 1: Create an Assistant

An **Assistant** represents an entity that can be configured to respond to a user's messages using several parameters like `model`, `instructions`, and `tools`.

```
Create an Assistant python
```

```
1 from openai import OpenAI
2 client = OpenAI()
3
4 assistant = client.beta.assistants.create(
5     name="Math Tutor",
6     instructions="You are a personal math tutor. Write and run code to answer math questions.",
7     tools=[{"type": "code_interpreter"}],
8     model="gpt-4o",
9 )
```

## Step 2: Create a Thread

A **Thread** represents a conversation between a user and one or many Assistants. You can create a Thread when a user (or your AI application) starts a conversation with your Assistant.

```
Create a Thread python
```

```
thread = client.beta.threads.create()
```

## Step 3: Add a Message to the Thread

The contents of the messages your users or applications create are added as **Message** objects to the Thread. Messages

# OpenAI Prompt

- Converting data into CSV

在最近發現的行星 Gooocrux 上，發現了許多水果。那裡長有名為 neoskizzles 的水果，它們是紫色的，味道像糖果。還有 loheckles，這是一種灰藍色的水果，味道非常酸，有點像檸檬。Pounits 是鮮綠色的，味道偏鹹而非甜。那裡還有許多 loopnovas，這是一種霓虹粉色的水果，味道像棉花糖。最後，還有一種名為 glowls 的水果，味道非常酸和苦，具有酸性和腐蝕性，並帶有淡橙色的色調。



CSV

```
水果,顏色,味道
neoskizzles,紫色,糖果
loheckles,灰藍色,非常酸像檸檬
pounits,鮮綠色,偏鹹而非甜
loopnovas,霓虹粉色,棉花糖
glowls,淡橙色,非常酸和苦,酸性和腐蝕性
```



## Parse unstructured data

Extract

Natural Language

Create tables from unstructured text.

### Prompt

SYSTEM

You will be provided with unstructured data, and your task is to parse it into CSV format.

USER

There are many fruits that were found on the recently discovered planet Gooocrux. There are neoskizzles that grow there, which are purple and taste like candy. There are also loheckles, which are a grayish blue fruit and are very tart, a little bit like a lemon. Pounits are a bright green color and are more savory than sweet. There are also plenty of loopnovas which are a neon pink flavor and taste like cotton candy. Finally, there are fruits called glowls, which have a very sour and bitter taste which is acidic and caustic, and a pale orange tinge to them.

### Sample response

```
Fruit,Color,Taste
neoskizzles,purple,candy
loheckles,grayish blue,tart
pounits,bright green,savory
loopnovas,neon pink,cotton candy
glowls,pale orange,sour and bitter
```

### API request

# No Free Lunch

The screenshot shows the 'Billing Overview' page in the OpenAI platform settings. The browser address bar shows 'platform.openai.com/settings/organization/billing/overview'. The page header includes 'AloT Lab / Default project' and navigation links for 'Dashboard', 'Docs', and 'API reference'. A left sidebar lists settings categories: 'Your profile', 'Organization' (with sub-items 'General', 'Members', 'Billing', 'Limits'), 'Project' (with sub-items 'General', 'Members', 'Limits'), and 'Create project'. The main content area is titled 'Billing' and has tabs for 'Overview', 'Payment methods', 'Billing history', and 'Preferences'. Under the 'Overview' tab, it displays 'Pay as you go' and a 'Credit balance' of '\$99.03'. A warning box states 'Auto recharge is off' and provides a link to 'Enable auto recharge'. Below this are buttons for 'Add to credit balance' and 'Cancel plan'. At the bottom, there are five quick links: 'Payment methods' (Add or change payment method), 'Billing history' (View past and current invoices), 'Preferences' (Manage billing information), 'Usage limits' (Set monthly spend limits), and 'Pricing' (View pricing and FAQs).

platform.openai.com/settings/organization/billing/overview

AloT Lab / Default project

Dashboard Docs API reference

SETTINGS

- Your profile
- Organization
  - General
  - Members
  - Billing**
  - Limits
- Project
  - General
  - Members
  - Limits
- + Create project

Forum

## Billing

Overview Payment methods Billing history Preferences

### Pay as you go

Credit balance ⓘ

# \$99.03

**Auto recharge is off**

When your credit balance reaches \$0, your API requests will stop working. Enable automatic recharge to automatically keep your credit balance topped up.

[Enable auto recharge](#)

Add to credit balance Cancel plan

**Payment methods**

Add or change payment method

**Billing history**

View past and current invoices

**Preferences**

Manage billing information

**Usage limits**

Set monthly spend limits

**Pricing**

View pricing and FAQs

# Model Pricing

- <https://openai.com/api/pricing/>

## GPT-4o

GPT-4o is our most advanced multimodal model that's faster and cheaper than GPT-4 Turbo with stronger vision capabilities. The model has 128K context and an October 2023 knowledge cutoff.

[Learn about GPT-4o ↗](#)

| Model             | Pricing                    | Pricing with Batch API*   |
|-------------------|----------------------------|---------------------------|
| gpt-4o            | \$5.00 / 1M input tokens   | \$2.50 / 1M input tokens  |
|                   | \$15.00 / 1M output tokens | \$7.50 / 1M output tokens |
| gpt-4o-2024-05-13 | \$5.00 / 1M input tokens   | \$2.50 / 1M input tokens  |
|                   | \$15.00 / 1M output tokens | \$7.50 / 1M output tokens |

## GPT-3.5 Turbo

The GPT-3.5 Turbo model is our fast, inexpensive model for simple tasks. It has a 16K context window and is optimized for dialog. The Instruct model supports a 4K context window.

[Learn about GPT-3.5 Turbo ↗](#)

| Model                  | Pricing                   | Pricing with Batch API*   |
|------------------------|---------------------------|---------------------------|
| gpt-3.5-turbo-0125     | \$0.50 / 1M input tokens  | \$0.25 / 1M input tokens  |
|                        | \$1.50 / 1M output tokens | \$0.75 / 1M output tokens |
| gpt-3.5-turbo-instruct | \$1.50 / 1M input tokens  | \$0.75 / 1M input tokens  |
|                        | \$2.00 / 1M output tokens | \$1.00 / 1M output tokens |

# Generate API Key

- Create API keys for different applications

## Your profile

Your details

Security

User API keys

Legacy



**User API keys have been replaced by project API keys.**

We recommend using project based API keys for more granular control over your resources. [Learn more](#)

[View project API keys](#)

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically disable any API key that has leaked publicly.

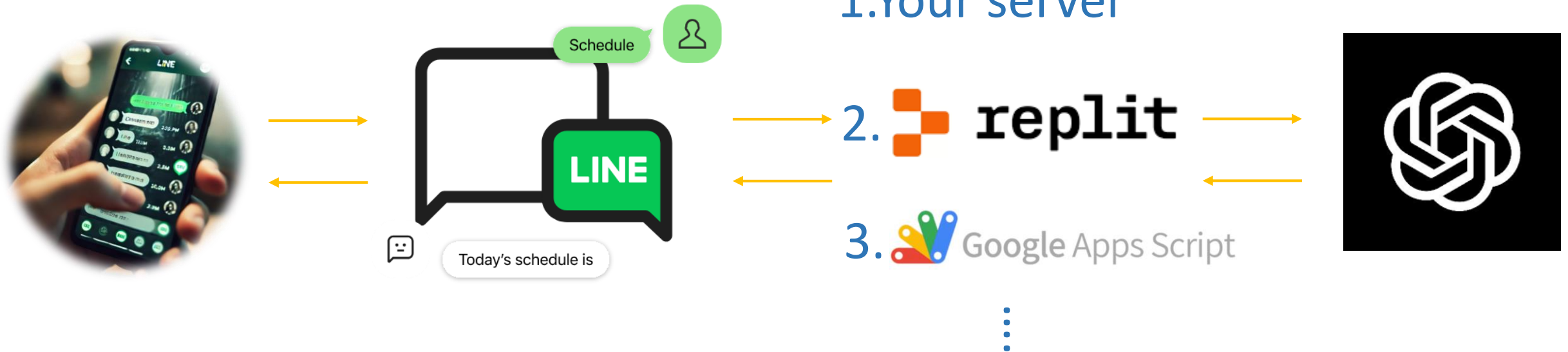
Enable tracking to see usage per API key on the [Usage page](#).

| NAME              | SECRET KEY | TRACKING ⓘ | LAST USED ⓘ  | PERMISSIONS |  |
|-------------------|------------|------------|--------------|-------------|--|
| <del>sk-...</del> | sk-...k8rJ | Enabled    | Mar 13, 2024 | All         |  |
| <del>sk-...</del> | sk-...ZE8S | Enabled    | Never        | All         |  |
| <del>sk-...</del> | sk-...A03V | Enabled    | Jul 20, 2024 | All         |  |

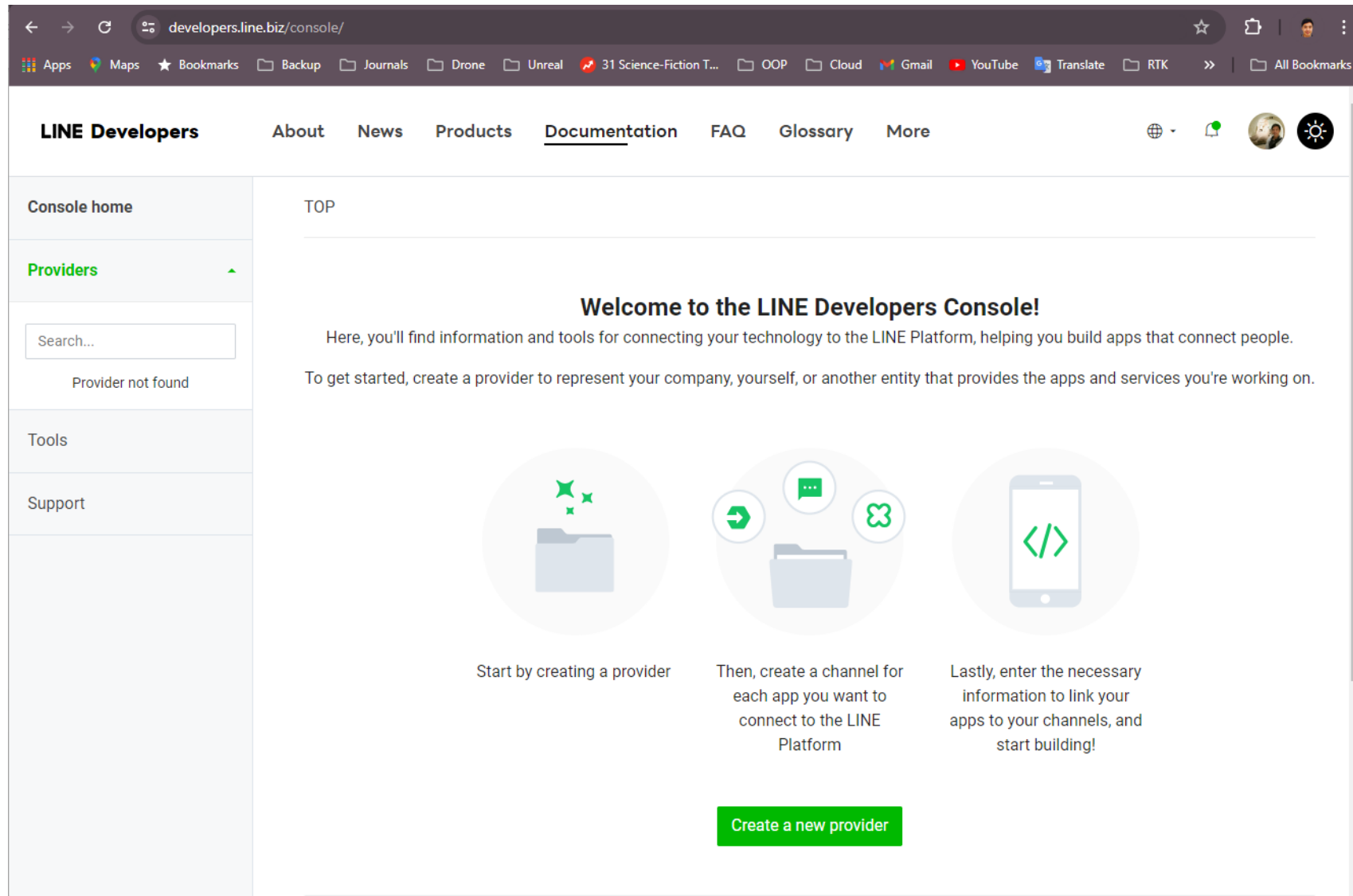
[+ Create new secret key](#)

# Connect to LineBot

- Connect LINE API via your server or cloud services



# Register LINE Business Account



# Setup LINE BOT

- [Messaging API | LINE Developers](#)



# Create a Messaging API channel

**LINE Developers**[About](#)[News](#)[Products](#)[Documentation](#)[FAQ](#)[Glossary](#)[More](#)



**Console home**

TOOLS

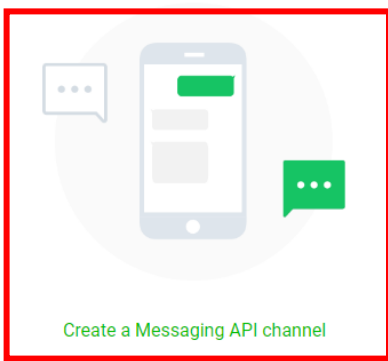
Support

TOP > AloTLab

To create one, choose a channel type below



Create a LINE Login channel



Create a Messaging API channel



Create a Blockchain Service channel



Create a LINE MINI App channel

Available features might defer based on the account with which you are currently logged in. For details, check the [document](#).

© LY Corporation

[About LINE Developers site](#) | [Terms and policies](#) | [About trademarks](#) | [Bug report form](#)

Family Sites  

Console home

Providers 

Search...

Admin

chatGPT

Tools

Support

TOP

## Create a new channel

Channel type

Messaging API

 Don't leave this empty

Provider

chatGPT

 Don't leave this emptyCompany or  
owner's country or  
region 

Taiwan

Corporations should select their company's country or region

 Don't leave this empty

Channel name

assistant

Note: The channel name can't be changed for seven days.

 Don't leave this empty Don't use special characters (4-byte Unicode) Enter no more than 20 characters

Channel description

chatGPT對話機器人

 Don't leave this empty Don't use special characters (4-byte Unicode) Enter no more than 500 characters

Category

音樂

 Don't leave this empty

Subcategory

專輯、歌曲、音樂活動

Linebot 名稱

# Add Our Class TA!

Console home

Providers

Search...

Admin

AloTlab

Tools

Support

TOP > AloTlab > AI小幫手 > Messaging API



AI小幫手

Admin

Messaging API

Basic settings

Messaging API

LIFF

Security

Statistics

Roles

Messaging API settings

Bot information

Bot basic ID @553avlxk

QR code



Scan this QR code with LINE to add your LINE Official Account as a friend. You can share the code with others.



AI小幫手

Today

Unread messages below



賴冠廷您好！  
我是AI小幫手。  
感謝您加入好友

此官方帳號將定期發放最新資訊給您  
敬請期待

5:52 PM

Enter a message

# Turn on Webhook

LINE Official Account Manager

assistant @493pyuzl 輕用量 0 聊天：關閉

主頁

分析

聊天

基本檔案

LINE VOOM

擴充功能

購物商城

設定

帳號設定

權限管理

回應設定

Messaging API

登錄資訊

帳務專區

總覽頁面

推廣方案

專屬ID

付款記錄

付款方式

電子發票資訊

優惠活動代碼

回應設定

可配合帳號的經營目的，設定聊天及自動訊息的回應方式。

回應功能

聊天

加入好友的歡迎訊息

Webhook

自動回應訊息

☐

☐

☒

☐

可透過聊天與好友互動。

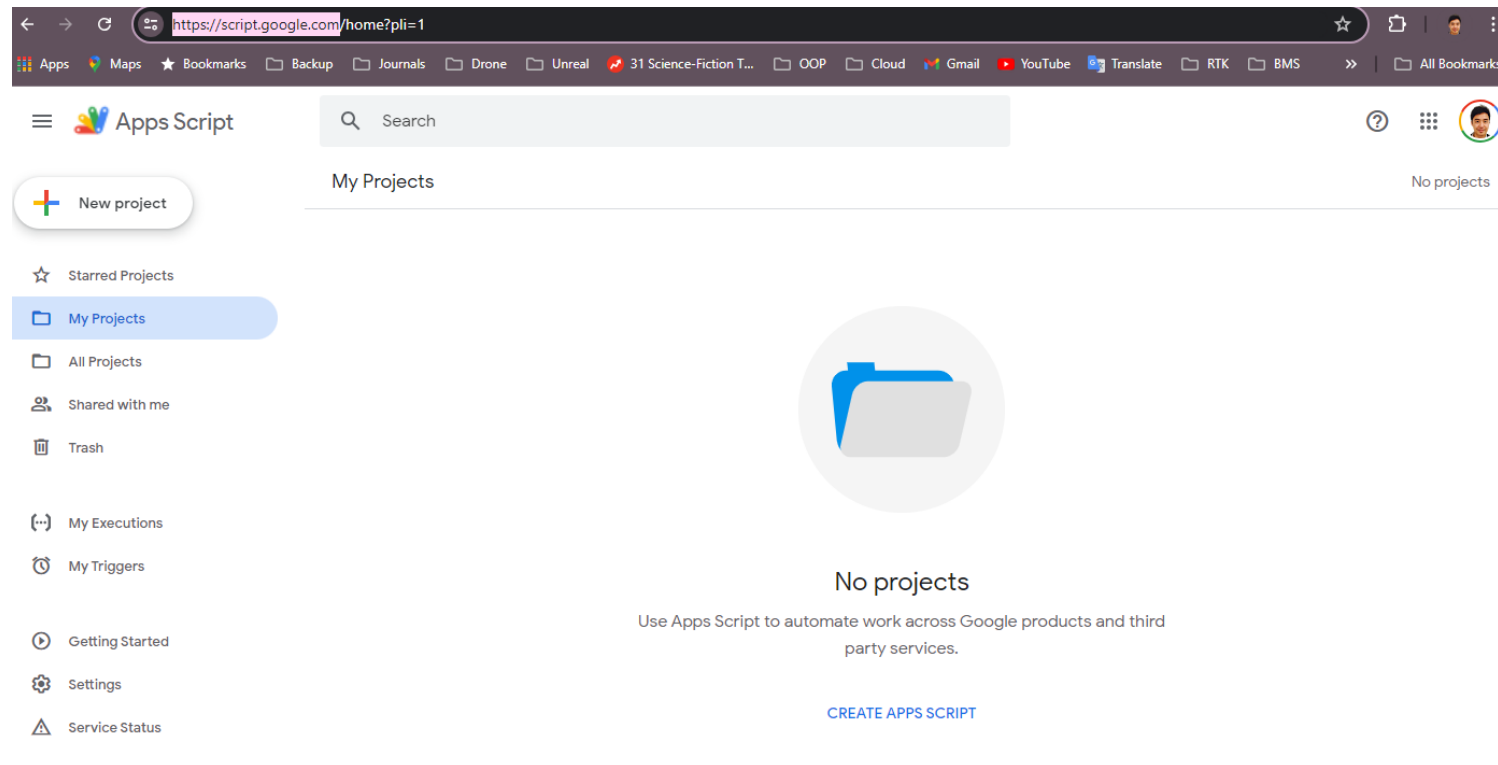
當用戶將本帳號加為好友時，可自動傳送訊息內容。  
[開啟「加入好友的歡迎訊息」設定畫面](#)

當用戶傳送訊息給本帳號或將本帳號加為好友時，從LINE平台傳送Webhook事件至Webhook網址。  
[開啟Messaging API的設定畫面](#)

可使用事先設定好的訊息內容進行自動回覆。  
[開啟自動回應訊息的設定畫面](#)

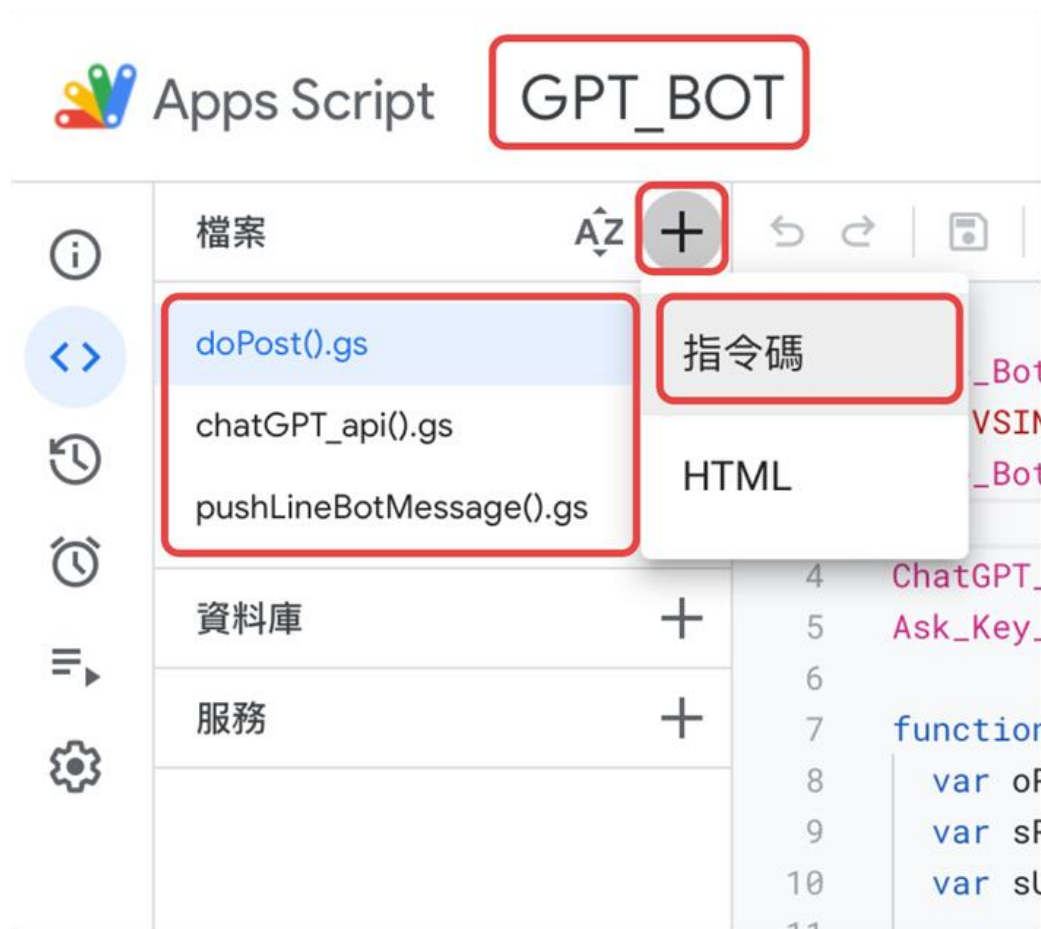
# Backend: Google App Script

- Free backend automation: <https://script.google.com/>
- Source code download: <https://destiny-panther-762.notion.site/CODE-d84f4cd6dd0e492d95c0b5200cceedd>



# Create new Script

- Create 3 scripts: doPost().gs, ChatGPT\_api().gs, pushLineBotMessage().gs



## 程式架構

- **doPost().gs**  
當有人 and LINEBOT 說話時自動會被觸發
- **chatGPT\_api().gs**  
負責 GPT 的串接
- **pushLineBotMessage().gs**  
負責 LINEBOT 訊息回覆

doPost().  
gs

```
1  LINE_CHANNEL_TOKEN = "您的LINE Channel Access Token";
2  LINE_BOT_API = "https://api.line.me/v2/bot/message/reply";
3
4  CHATGPT_ACCESS_TOKEN = "您的ChatGPT API TOKEN";
5  Ask_Keyword = "請問"
6
7  function doPost(e) {
8      var oPostData = JSON.parse(e.postData.contents);    //取得 Bot 收到的資訊
9      var sReplyToken = oPostData.events[0].replyToken
10     var sUserMsgText = oPostData.events[0].message.text
11
12     // 以下狀況不回覆
13     if (!sUserMsgText) {return;}
14     if (sUserMsgText.indexOf(Ask_Key_Word) === -1 ) {
15         sResponse = '如果想問問題的話，請開頭請用 ' + Ask_Keyword + ' 再接著問題內容即可哦!'
16     }
17     else{
18         // 呼叫ChatGPT API
19         sGPTReceive = chatGPT_api(sUserMsgText.replace(Ask_Key_Word, ""))
20         sResponse = JSON.parse(sGPTReceive).choices[0].message.content
21     }
22
23     pushLineBotMessage(sResponse, sReplyToken);
24 }
```

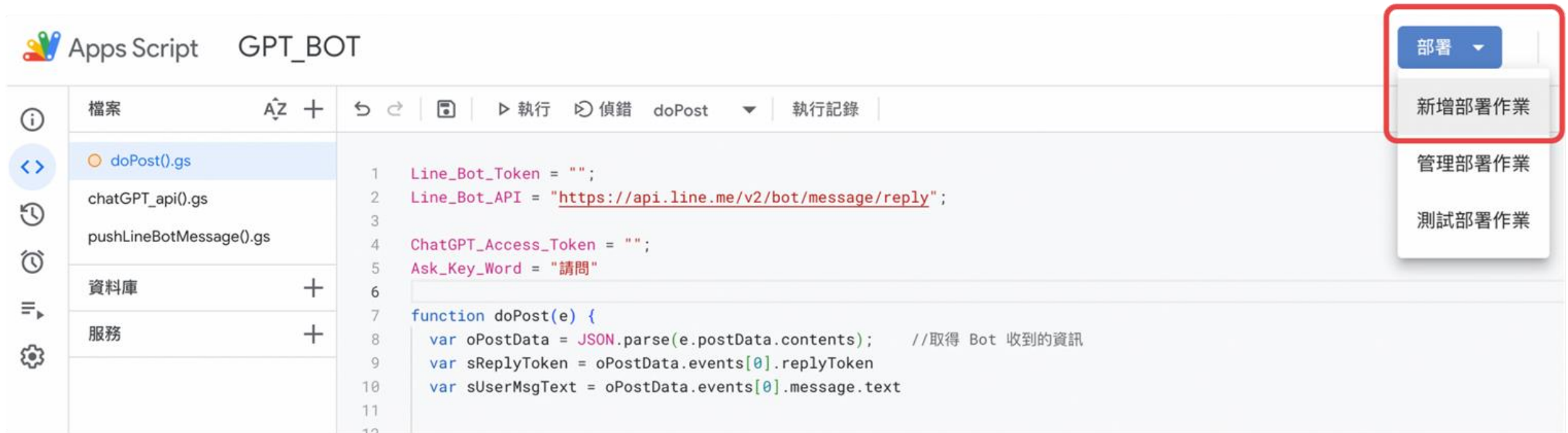
# ChatGPT\_api().gs

```
1 function chatGPT_api(sMsg) {  
2   return UrlFetchApp.fetch('https://api.openai.com/v1/chat/completions', {  
3     'method': 'post',  
4     'headers': {  
5       'Content-Type': 'application/json',  
6       'Authorization': 'Bearer ' + CHATGPT_ACCESS_TOKEN,  
7     },  
8     'payload': JSON.stringify({  
9       "model": "gpt-3.5-turbo",  
10      "messages": [{"role": "user", "content": sMsg}],  
11      "temperature": 0.7,  
12      "max_tokens": 300  
13    })),  
14  })  
15 }
```

## pushLineBot Message().gs

```
1  function pushLineBotMessage(sMsg, sReplyToken){
2
3      var linePayload = {
4          'replyToken': sReplyToken,
5          'messages': [
6              {
7                  'type': 'text',
8                  'text': sMsg
9              }
10         ]
11     };
12
13     var lineOptions = {
14         'method': 'post',
15         'headers': {
16             'Content-Type': 'application/json; charset=UTF-8',
17             'Authorization': 'Bearer ' + LINE_CHANNEL_TOKEN
18         },
19         'payload': JSON.stringify(linePayload)
20     };
21
22     UrlFetchApp.fetch(LINE_BOT_API, lineOptions);
23 }
```

# Deploy Google Apps Script (1)



Apps Script GPT\_BOT

檔案 A-Z +

doPost().gs

chatGPT\_api().gs

pushLineBotMessage().gs

資料庫 +

服務 +

```
1 Line_Bot_Token = "";
2 Line_Bot_API = "https://api.line.me/v2/bot/message/reply";
3
4 ChatGPT_Access_Token = "";
5 Ask_Key_Word = "請問"
6
7 function doPost(e) {
8   var oPostData = JSON.parse(e.postData.contents); //取得 Bot 收到的資訊
9   var sReplyToken = oPostData.events[0].replyToken
10  var sUserMsgText = oPostData.events[0].message.text
11
12
```

部署

新增部署作業

管理部署作業

測試部署作業

# Deploy Google Apps Script (2)

### 新增部署作業

選取類型

設定

網頁應用程式

API 執行檔

外掛程式

資料庫

網頁應用程式

執行身分

我 (chitinnng@gmail.com)

網頁應用程式將有權在執行時使用您的帳戶資料。

誰可以存取

所有人

這項專案也可以當做程式庫使用。[瞭解詳情](#)

取消

部署

# Deploy Google Apps Script (3)

## 新增部署作業

已成功更新部署作業。

3 版 (2024年5月9日凌晨3:17)

部署作業 ID

AKfycbyooLvmq9Frvl3VWU6gFkdaKoHe\_vQemnYHJ3oXlAs7CHr-q\_5Sw9\_c\_Zz7WeRdPCUW6w

 複製

網頁應用程式

網址

[https://script.google.com/macros/s/AKfycbyooLvmq9Frvl3VWU6gFkdaKoHe\\_vQemnYHJ3oXlAs7CHr-q\\_5Sw9\\_c\\_Zz7We...](https://script.google.com/macros/s/AKfycbyooLvmq9Frvl3VWU6gFkdaKoHe_vQemnYHJ3oXlAs7CHr-q_5Sw9_c_Zz7WeRdPCUW6w)

 複製

完成

# Setup LINE BOT's Webhook

Console home

TOP > chatGPT > assistant > Messaging API



Scan this QR code with LINE to add your LINE Official Account as a friend. You can share the code with others.

Webhook settings

Webhook URL ⓘ  
  

✓ Don't leave this empty

✓ Enter a valid HTTPS URL

✓ Enter no more than 500 characters

Update

Cancel

Webhook settings

Webhook URL ⓘ  
  

Verify

Edit

Use webhook ⓘ

☒

# Talk to Our Assistant

AI小幫手



9:07 PM 你好



如果想問問題的話，請開頭請用 請問再  
接著問題內容即可哦!

9:07 PM

9:07 PM 請問你是誰?



我是一個AI助手，可以回答你的問題和提  
供幫助。有什麼可以幫到你的嗎?

9:07 PM

9:24 PM 請問要怎麼學習AI知識?



要學習AI知識，首先可以從以下幾個途徑  
入手：

1. 在線課程：許多線上教育平台提供了AI  
相關的課程，如Coursera、edX、Udemy  
等，可以通過這些課程學習AI的基礎知識  
和應用技術。

2. 學術書籍：可以閱讀相關的專業書籍，  
如《人工智能：一種現

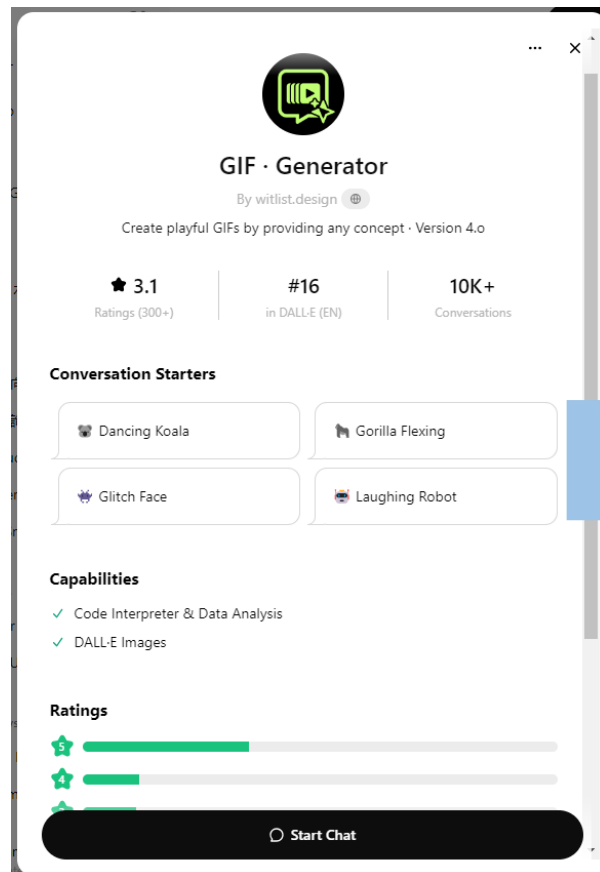
9:24 PM

Enter a message



# Generate Gif

- Choose “Explore GPTs”,  
Enter “gif”



generate the smiling actions of a cute cartoon girl from the future.  
anime-style.

Send 'A' to redo (It should be a grid)

Time to count the GIF frames!

Tell me the grid size like:  $\leftrightarrow \times \updownarrow$   
( Width x Height ) to continue...

For example: '4x4'.. or '3x5'..



Message GIF · Generator



Thank You!

