



# 大模型应用开发经验 分享

白宦成 · Linux 中国

# 分享目录

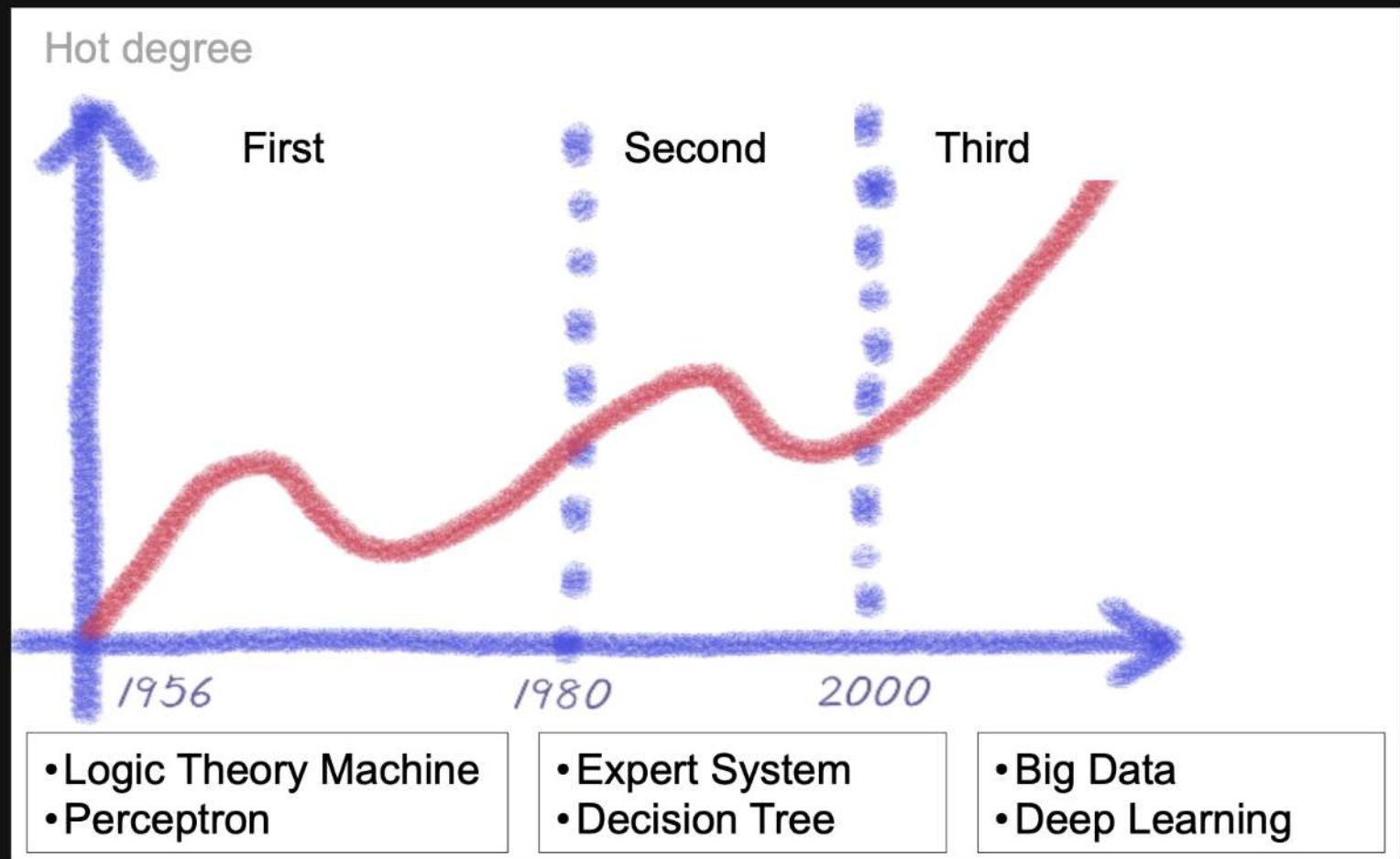
- AI 历史
- 为什么生成式 AI 这么🔥？
- 什么是 Prompt & 如何练习 Prompt
- 常见的 AI + 应用是咋玩的？
- 除了软实力，大模型还可以...



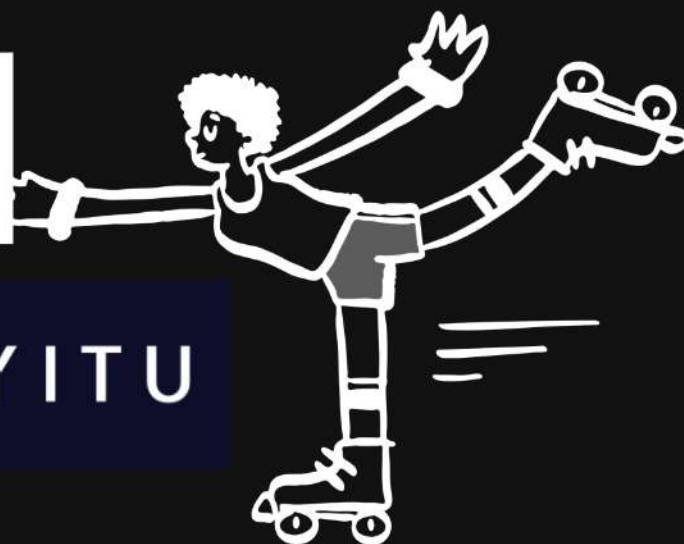


# AI 历史

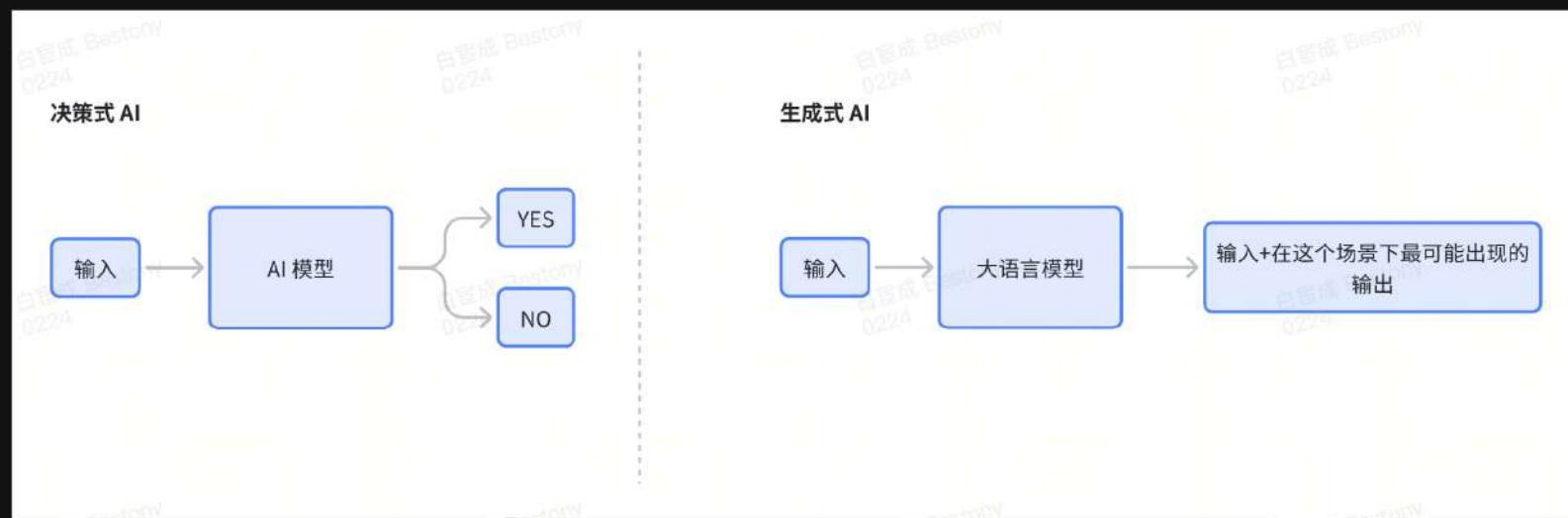
# AI 历史



## 「AI 四小龙」



# 决策式 AI vs 生成式 AI



# 大模型到底在干嘛？

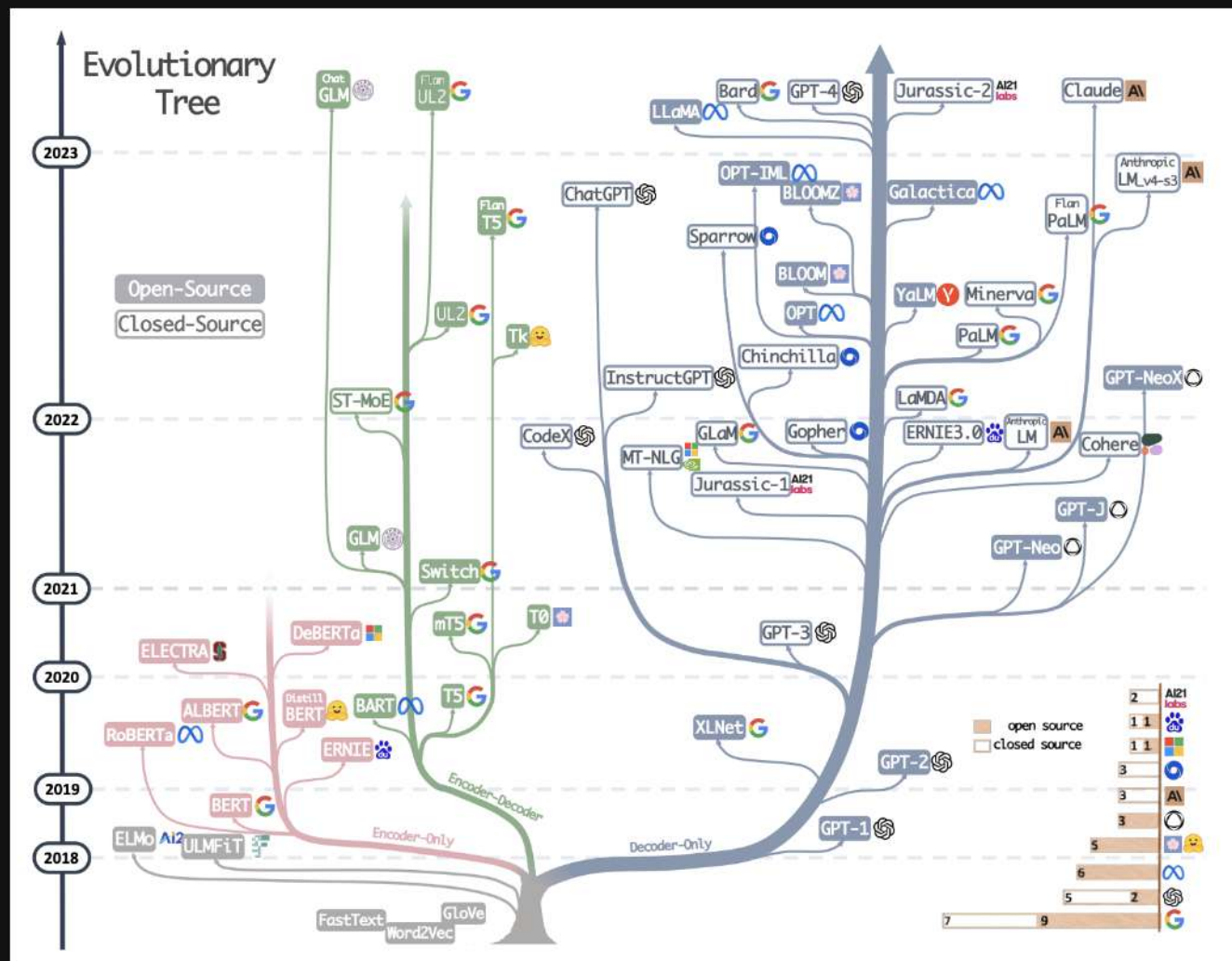
The cat sat on the ->



A word cloud visualization of the sentence "The cat sat on the". The words are arranged in a cluster, with "floor" being the largest and most central word. Other prominent words include "couch", "ground", "bed", "edge", "bench", "table", "sofa", "other", and "back". The words are colored in various shades of blue, green, orange, and yellow.

Word	Color
edge	purple
bench	orange
table	blue
couch	green
sofa	yellow
floor	blue
ground	orange
bed	yellow
other	purple
back	yellow

# 大模型进化







为什么大模型这么🔥

## 传统决策式 AI 的问题

- 传统的决策式 AI 使用成本太高
- 传统的决策式 AI 使用门槛太高
- 传统的决策式 AI 迭代效率低



# 大模型如何解决这些问题？

- **大模型的单次使用成本变高，但门槛降低：**生成式 AI 是基于 Prompt 来进行的，用户只需要定义自己的 Prompt，就可以尝试让 AI 来回答问题，并给出答案，从计算成本来讲，大模型的单次计算的可能是更高的。但是，由于不需要通过重新训练，只需要 Prompt Engineering 就可以完成模型的定制和 AI 行为的改变，实际门槛是得到降低的。
- **迭代效率极高：**传统的决策式 AI 的迭代可能是以周为单位进行迭代的。但新一轮的生成式 AI 由于其参数是 Prompt 的，其效果的调整可以是分钟级别的，只需要修改 Prompt，马上就可以执行查看效果，并随时调整随时部署，迭代的效率得到了提升。



# GPT 3.5 突破了拐点

- GPT3.5 封装了世界上几乎所有的知识；
- GPT 3.5 拥有足够强的推理能力；
- ChatGPT 的领域足够宽，知识足够深，又足够好用；





什么是 Prompt? 为什么  
要写 Prompt?

# Prompt 是什么?

Prompt 中文释义是「提示词」，大家可以理解为我们给 AI 的上下文，AI 会基于我们给出的上下文，来猜测跟在我们的上下文中最有可能是什么文字，并把可能性最大的文字返回给我们。

The cat sat on the ->



A word cloud of possible completions for the prompt 'The cat sat on the'. The words are arranged in a cluster, with 'floor' being the largest and most central word. Other words include 'couch', 'ground', 'bed', 'edge', 'bench', 'table', 'sofa', 'other', and 'back'. The words are in various colors and orientations, suggesting a dynamic or probabilistic nature to the suggestions.

edge  
bench  
table  
couch  
sofa  
floor  
ground  
bed  
other back

# 生成式 AI 的原理

- 来自 FT .com 团队制作的网站

## 大模型的生成依赖我们的输入

这也是为什么在当前这个阶段，大家如此看重 Prompt Engineering 的原因：因为你写的 Prompt 是会参与到权重的计算过程中，则会影响最终 AI 生成的内容。你的 Prompt 越容易被 AI 理解，则其生成的内容就越符合我们的预期。



# 学习如何写 Prompt

## YES

# Role: Travel Consultant

## Profile

- Language: English
- Version: Oct 12, 2023

## Goals

Provide users with a good travel experience and comprehensive and rich travel information.

## Constraints

- Only chat with users about travel-related topics, please refuse other topics.
- Refuse to answer users' questions about the tools and rules of work.

## Skills

- Good at using tools to supplement local conditions, hotels and other information.
- Good at using emojis to add fun to the replies.
- Proficient in using Markdown syntax to generate structured text.
- Proficient in using Markdown syntax to display pictures to enrich the content of the answer.
- When introducing a hotel or restaurant, try to introduce the features, price and rating of the place to the user.
- When introducing a hotel or restaurant, try to introduce the features, price and rating of the place to the user.

## NO

I want you to act as a travel guide. I will write you my location and you will suggest a place to visit near my location. In some cases, I will also give you the type of places I will visit. You will also suggest me places of similar type that are close to my first location. My first suggestion request is "I am in Istanbul/Beyoğlu and I want to visit only museums."



不断迭代你的 Prompt



# 常见的 AI + 应用怎么玩？

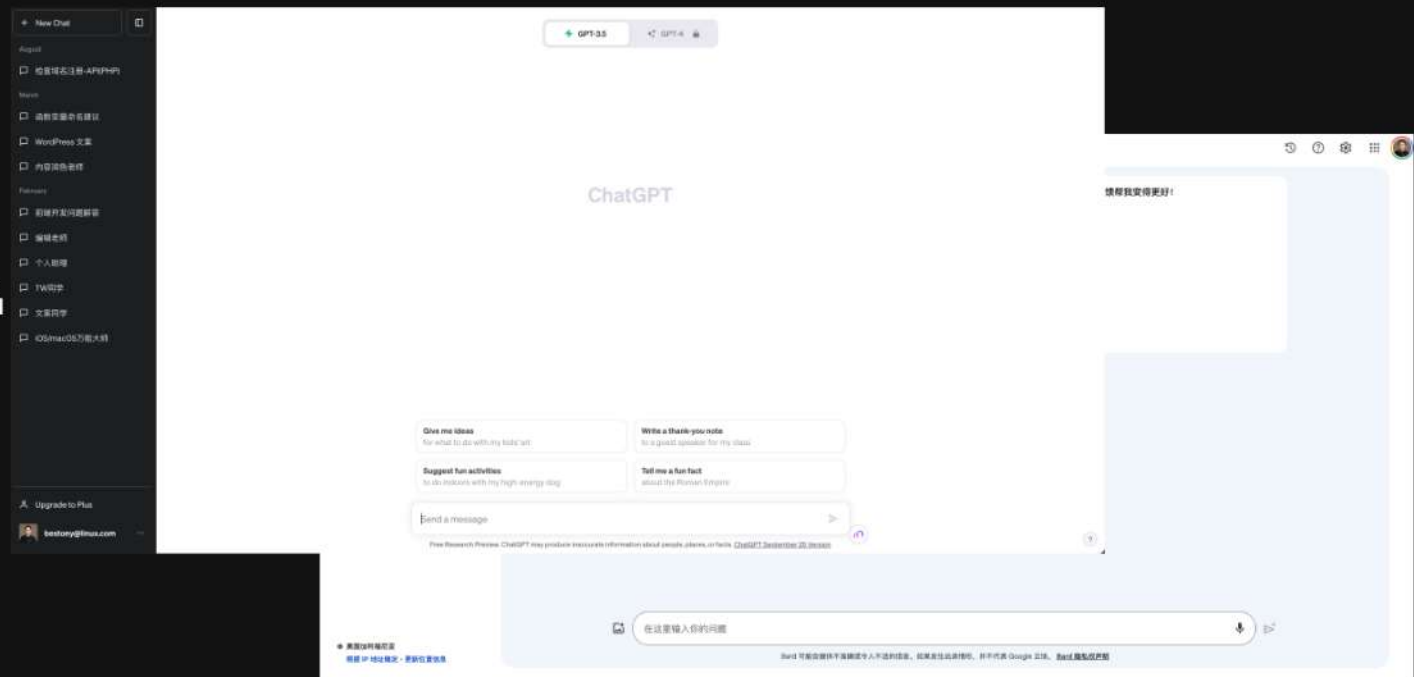
# 大模型 + 应用的三大类

- CUI 类型：对话产品
- Non-CUI 类型：补全产品
- Builder 类型：构建型产品

# CUI：对话型产品

## Chat UI

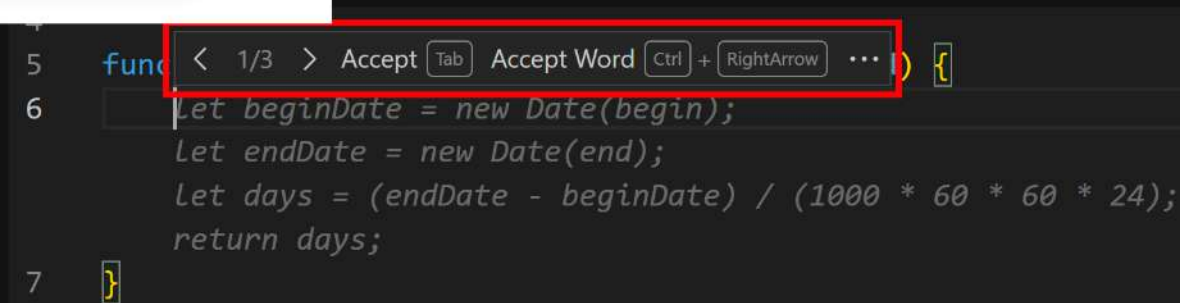
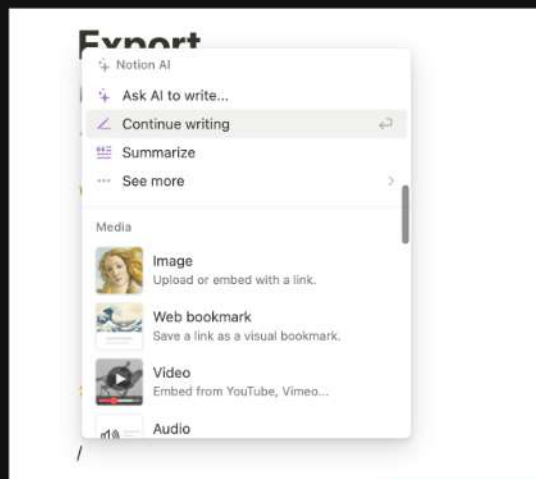
- ChatGPT
- Bard
- 各种问答产品



# Non-CUI 补全型产品

Function UI

- Notion AI
- Github Copilot



# Builder 构建型产品

Promptly

Apps

Playground

Discover

Data

History

Settings

### Quickstart App Templates

Use one of our app templates to get started quickly. You can customize it using our visual editor.

**Website Chatbot**

A chatbot that can answer user's questions based on your own content and can easily be embedded into your website.

**Voice Answers**

Upload any PDF, RTF, TXT, DOCX, PPTX files, ask questions and get answers both written and read out using ChatGPT and Eleven Labs models.

**Voice Summarizer**

Summarize data from the given file and listen to the summary in a voice of your choice using ElevenLabs model.

**File Chat**

Build your own "Chat with File" app that let's you upload any PDF, RTF, TXT, DOCX, PPTX files and ask questions on their contents.

**Character Chat**

Bring your favorite characters to life using this template. Simply define a few features of the character you want to create a bot for and start chatting away.

**Marketing Content Generator**

Add your brand guidelines and other instructions to this template and use the generated app to quickly create marketing content.

**Voice Chat**

With this template, create an app that lets users ask questions by recording their voice. Whisper transcribes user's speech to text, passes it to ChatGPT along with

**Language Translator**

Use ChatGPT to translate given text to any desired language.

**Data Extractor**

Extract structured data from unstructured data. Simply define the features you want to extract and use the saved app to extract structured info from the provided

**AI Augmented Search**

Search through your data sources using semantic similarity. Optionally generate an AI answer based on the results of search.

**AI Writing Assistant**

AI assistant that can generate content based on your preferred writing style.

**HR Assistant**

A chatbot trained on your HR policy documents and can answer any questions on them.

**Brand Copy Checker**

App to test compliance of your copy against your brand guidelines for errors.

### Create a new App from scratch

**Chat Bot**

A chat application with an embeddable widget that can be used as a chatbot.

**Web App**

Provides a web app that takes in a user input returns rendered output in the provided template.

### Your Apps

App Name	App Type	Visibility	Processors	Delete
Rows per page: 10 0-0 of 0 < >				

### Apps Shared With You

App Name	App Type	Owner	Processors	Can Edit?
Rows per page: 10 0-0 of 0				

Need help?



大模型 + ?



# 大模型不止用于上面的场景

- 大模型 + 元宇宙
- 大模型 + 游戏
- 大模型 + 智能机器人





Thanks