



大模型应用开发经验 分享

白宦成 · Linux 中国

分享目录

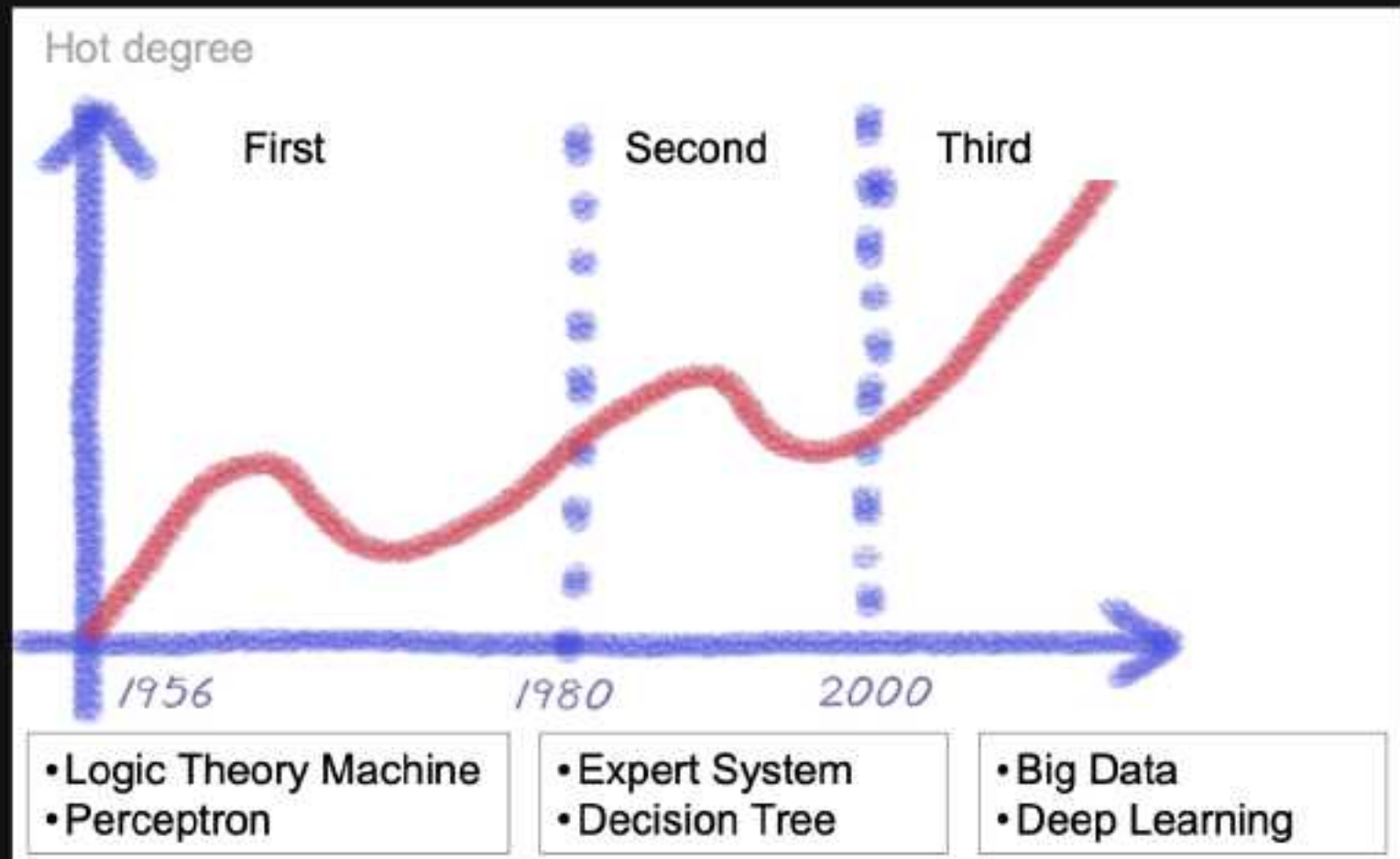
- AI 历史
- 为什么生成式 AI 这么🔥?
- 什么是 Prompt & 如何练习 Prompt
- 常见的 AI+ 应用是咋玩的?
- 大模型应用开发经验
- 除了软实力，大模型还可以...





AI 历史

AI 历史



「AI 四小龙」



商汤
sensetime

MEGVII 旷视



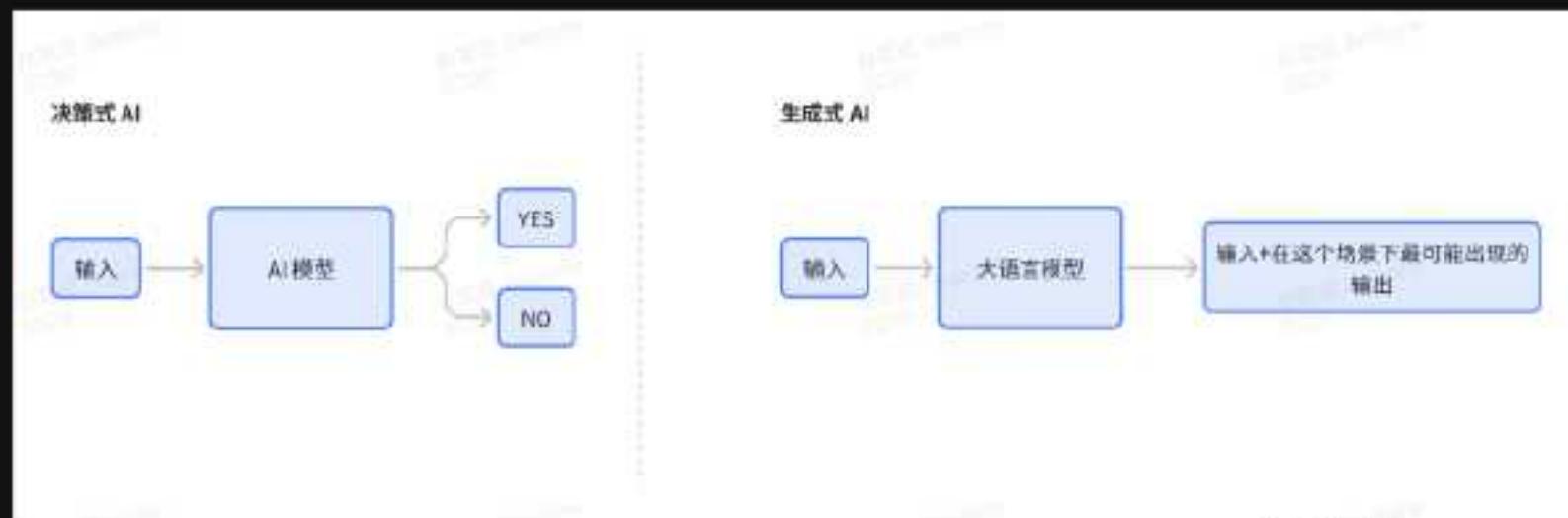
云从科技
CLOUDWALK



依图 | YITU



决策式 AI vs 生成式 AI



大模型到底在干嘛？

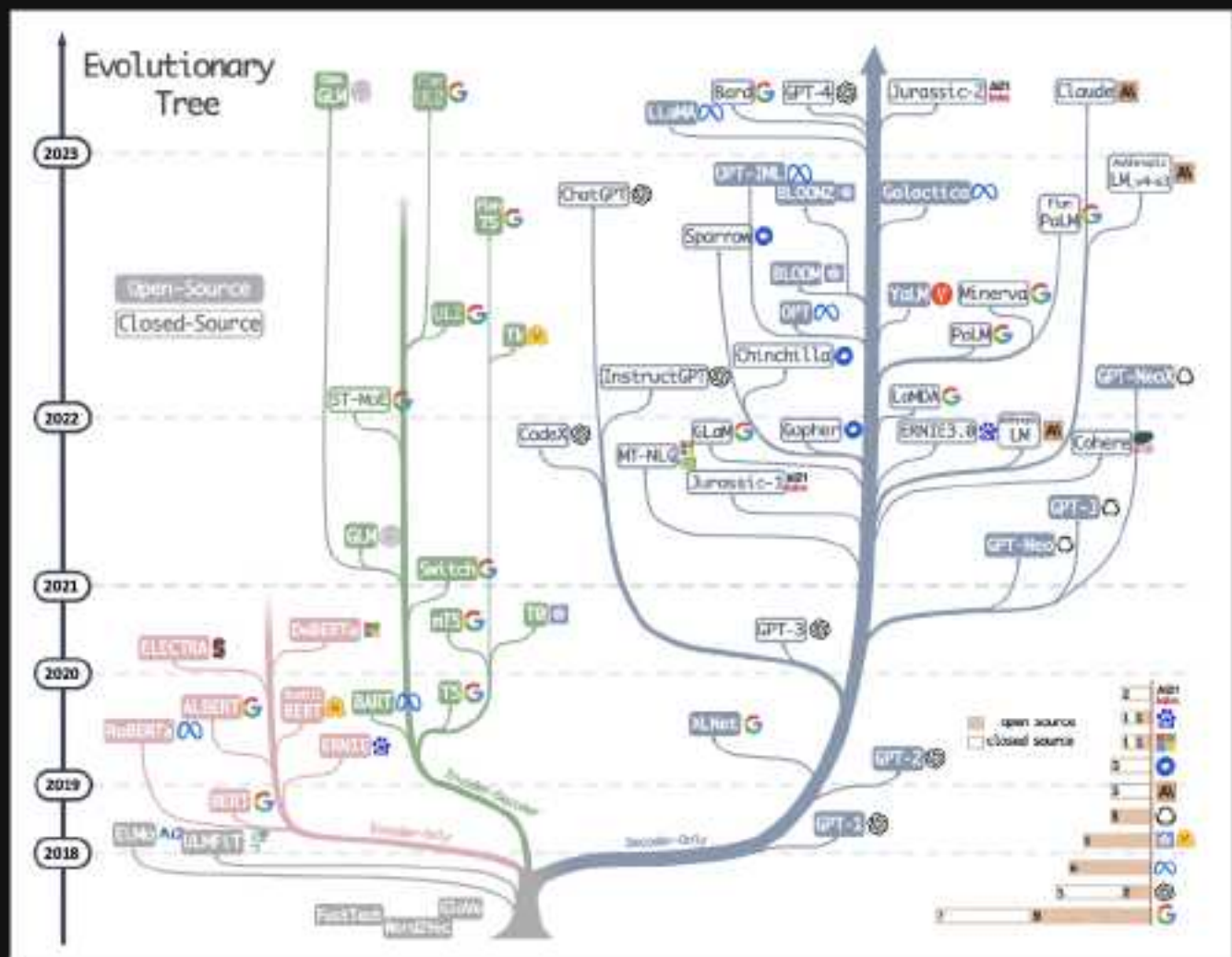
The cat sat on the ->



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.

edge
bench
table
couch
sofa
floor
ground
bed
other
back

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 visualization showing various words that could complete 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.

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

单词首先会被拆分为 **Tokens**，一种能够被编码的基础单元。虽然 Token（词元）常常只代表单词的一部分，我们这里简化处理，把一个完整单词当成一个 Token。

We go to work by train

要理解 **work** 这个词，LLMs 需要在大量的训练数据中观察它是如何使用的，看看它**周围的单词** 都有哪些。这些数据来自互联网上的大量文本，新的 LLMs 通过学习数十亿个单词来训练。

in	the	kitchen	last		
at	of	to	do		
reviewed	her	before	submitting		
where	they	and	entered		
a	nasty	environment	like		
on	her	in	quantum		
I	went	into	work	yesterday	morning

I can't believe how much this project requires

Many people go downtown to in office buildings

The roof at has a great view

He takes pride in his hard and effort

The accountant meticulously analyzed related financial documents

My commute is easy because I in my neighbourhood

It's essential to balance and personal life effectively

经过这样的训练，我们获得了一个巨大的单词库，里面有与 `work` 一同出现的单词，还有哪些 没在 `work` 附近出现的。

work	roof
------	------

work	at
------	----

work	a
------	---

work	his
------	-----

work	has
------	-----

work	hard
------	------

work	zebra
------	-------

work	effort
------	--------

work	atmosphere
------	------------

work	dove
------	------

work	in
------	----

work	to
------	----

work	and
------	-----

work	office
------	--------

work	polka
------	-------

work	downtown
------	----------

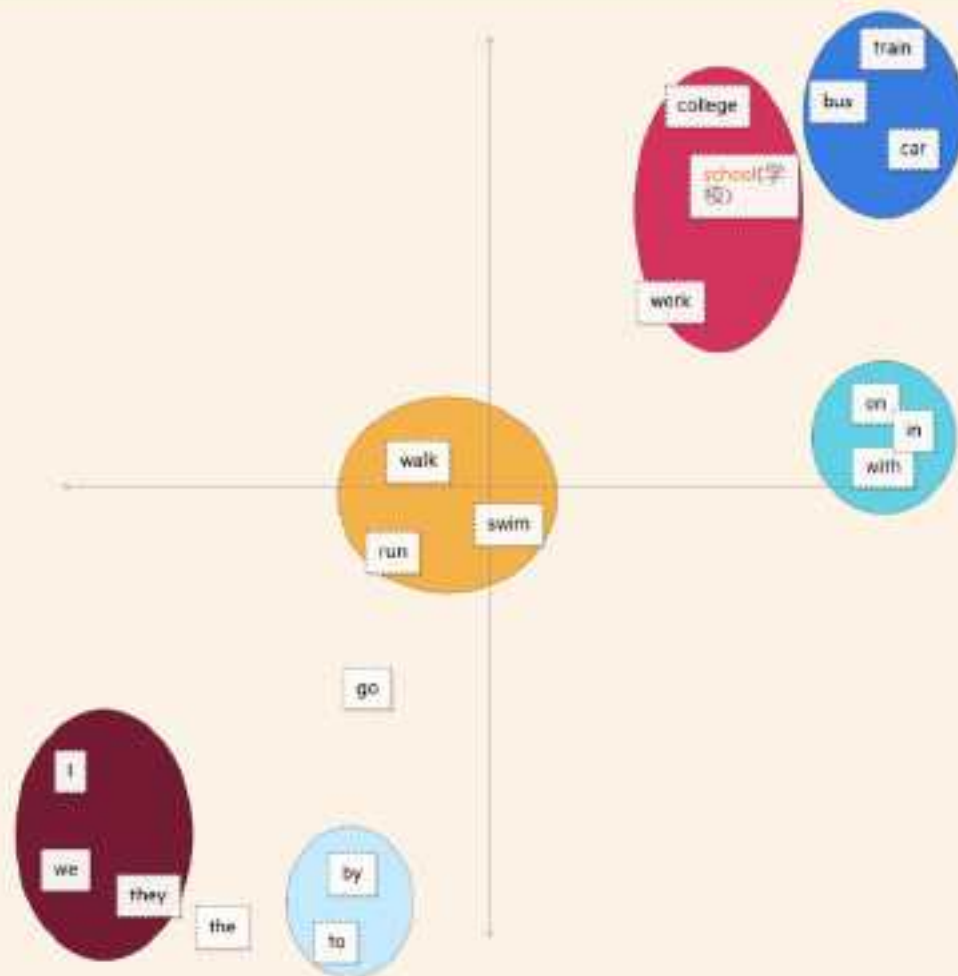
处理 这些单词时，模型会生成一个包含多个值的向量（或者说一系列数值），并根据每个单词与 **work** 在训练数据中的距离调整这些值。这个向量就是所谓的词嵌入（Word Embedding）。

work	roof	work	atmosphere
work	at	work	dove
work	a	work	in
work	his	work	to
work	has	work	and
work	hard	work	office
work	zebra	work	polka
work	effort	work	downtown

0  1



这样，我们就能发现一些词群，比如 **代词** 群，**交通工具** 群。而能量化单词，正是让模型成功生成文本的第一步。







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





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



ChatGPT - GPTs





Non-CUI大模型应用开发经验

Non-CUI / CUI大模型应用靠什么？

大模型推理能力 + 行业 Know How + 行业数据 + Prompt + 多模态 + 业务逻辑 Code

大模型推理能力

- 大模型推理能力越强，越「聪明」，对于 Prompt 的理解能力越强

行业 Know How

- 行业 Know How 决定了你能解决多么深刻的问题，产生多大的业务价值。
- 单纯的 General 的 Q & A 提供的价值有限，行业问题和答案才能提供价值的加成

行业数据

- 大模型自身的数据集是稳定的，没有差异，差异来自于各自的行业数据，最终形成护城河和城墙。

Prompt

- Prompt 本身是为了矫正大模型补全时的能力。
- 通过不断添加 Prompt 来找到黄金分割点



多模态

- 更多的模态提供了信息压缩的能力 — 将二维的图片信息压缩成一维的文本，帮助 AI 更好的理解上下文，提供更加丰富的能力。

业务逻辑

- 内嵌在业务系统之中的部分，根据业务场景，选择最合适的交互方式。
- 可能是按钮、可能是自动触发

举个🍌

假设我们要做一个活动的大模型应用：

- 大模型推理能力：我应该选通义千问还是选文心一言还是选 OpenAI?
- 行业 Know How：活动组织者关注的是什么？是流程？还是成本？我应该做哪个方向的大模型应用？
- 行业数据：假设我要做成本这个方向，我需要哪些数据？哪里有这些高质量的数据？
- Prompt：如何写 Prompt，才能让大模型更好的为我提供服务。
- 多模态：大模型能理解我的手绘需求么？还是一定要写成文字才行？
- 业务逻辑 Code：展现形态是什么样？是一个独立的 App 还是一个小程序？



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