Wittgenstein

语言的边界，就是世界的边界
而在今天，我们可以用语言去生成世界

Wittgenstein 让文本模型通过 Harness 直接产出真实多模态文件
MAKE TEXT MODELS SPEAK IN MEDIA

WTF

IS

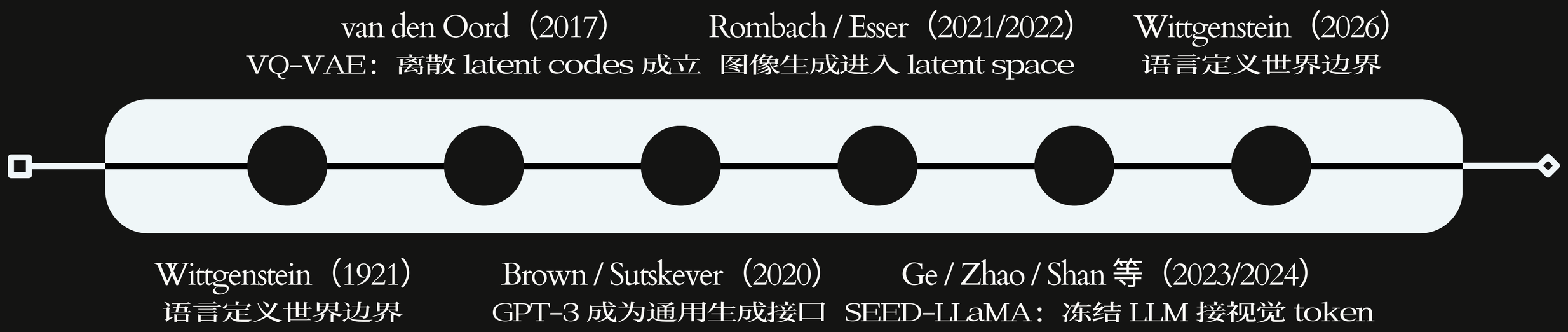
Wittgenstein
Wittgenstein
Wittgenstein
Wittgenstein
Wittgenstein

Wittgenstein

↗ wittgenstein.wtf

The modality harness for text-first models

一切来源于火花

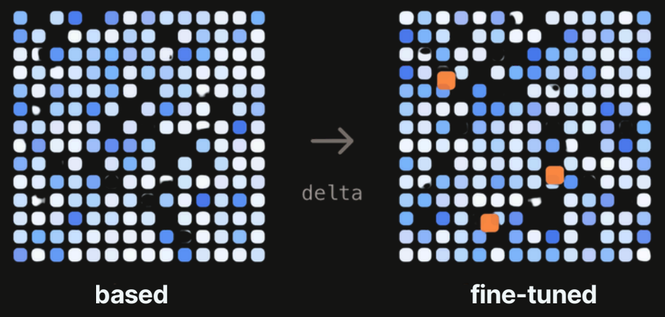


We build 1921 **differently.**

多模态输出仍然成本高昂，
但文本模型本身就是出色的推理器

Ilya 的论断是："一个足够好的 LLM 就是一个逼近 Kolmogorov 最优压缩的系统"

*给定数据 X 和 Y，足够好的压缩器会自动用 X 里的结构去压缩 Y



Text LLM 已经在参数里压缩了大量图像的结构信息

如 Minimax：LLM 越大，压缩率越高。压缩即理解

我们的 Hypothesis 和结论

我们在重新定义 multimodality

- 模态的差异只存在于 codec 层
- model 层可以完全 modality-agnostic

过去十年业界以为多模态必须依赖 VQ-VAE + fusion training，必须重训整个 model

我们通过极小的后训练成本以及规则编解码

快速赋予纯文本 llm多模态生成能力

and the COST?

$0.00

Additional investment.

How WE BUILD

the shit

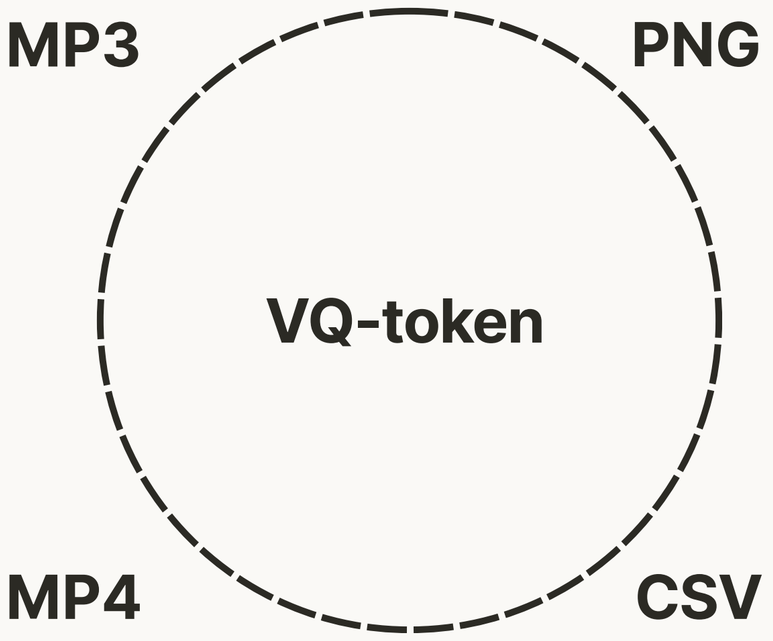
Wittgenstein 是一个多模态 Agent Harness

llm

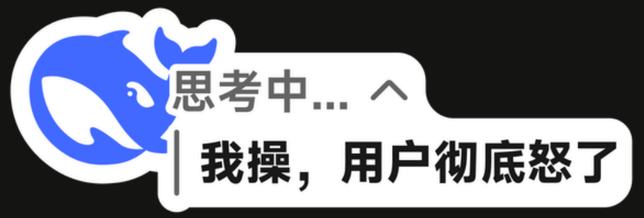
照样用文本「想事情」

Harness

负责把结果变成：



多模态输出仍然成本高昂，
但文本模型本身就是出色的推理器

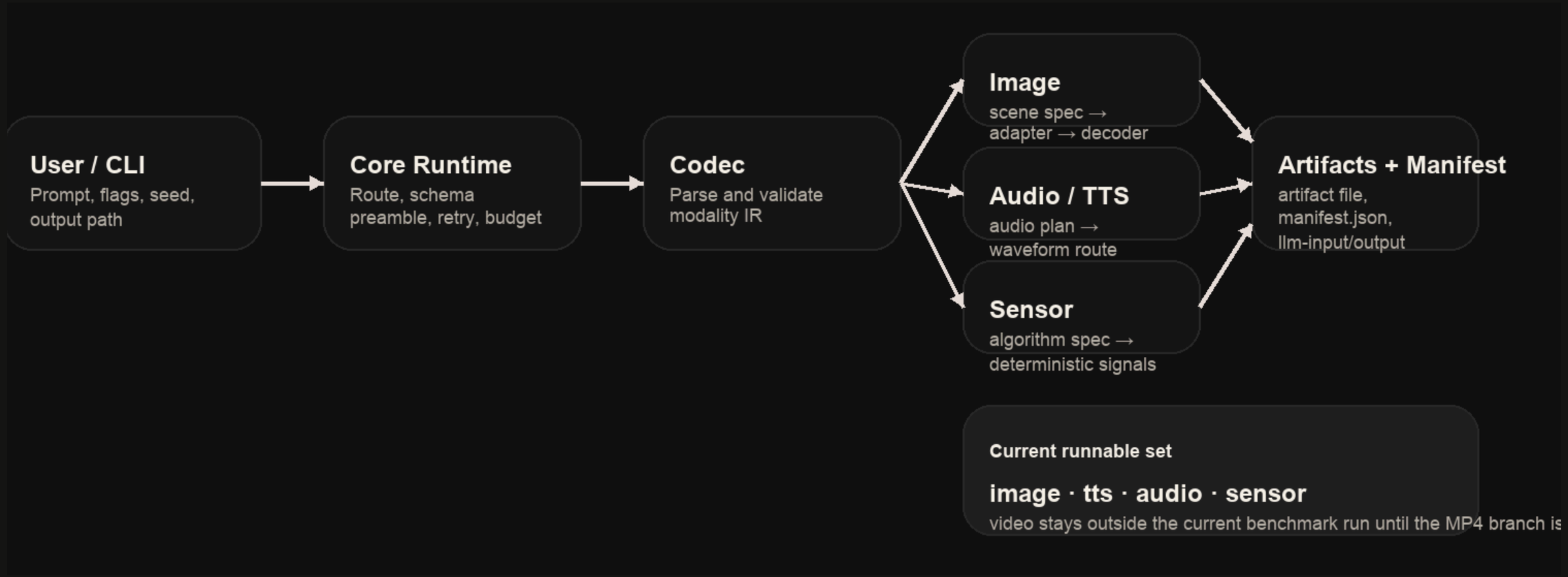


但是

好饭不怕晚

代价是 <5% 的推理额外 token 开销，且
无需任何 model 侧的介入

Architecture



Strict, strict image path

Text LLM
Returns scene semantics, composition, style, and decoder hints.

→

Scene Spec JSON
Validated structured output. No SVG, no Canvas, no raw pixels.

→

Adapter
Small learned scene → latent bridge.

→

Frozen Decoder
Deterministically reconstructs raster pixels.

→

PNG
Final image artifact

Optional bypass

If a provider can return valid discrete `providerLatents` in the JSON object, the runtime can skip the learned adapter and decode those tokens directly.

Why this matters

The repo keeps image generation aligned with a VQ-style discrete latent framing: decoder not generator, adapter not full image model.

Some use cases.

别听我们之前讲得很牛逼，但是接下来你会看到的效果

求，别骂

Everything built in 6 hours.

一些我们跑过的案例

虽然简陋，但是保真

重复一遍：

别骂

6 hours.

重复一遍：

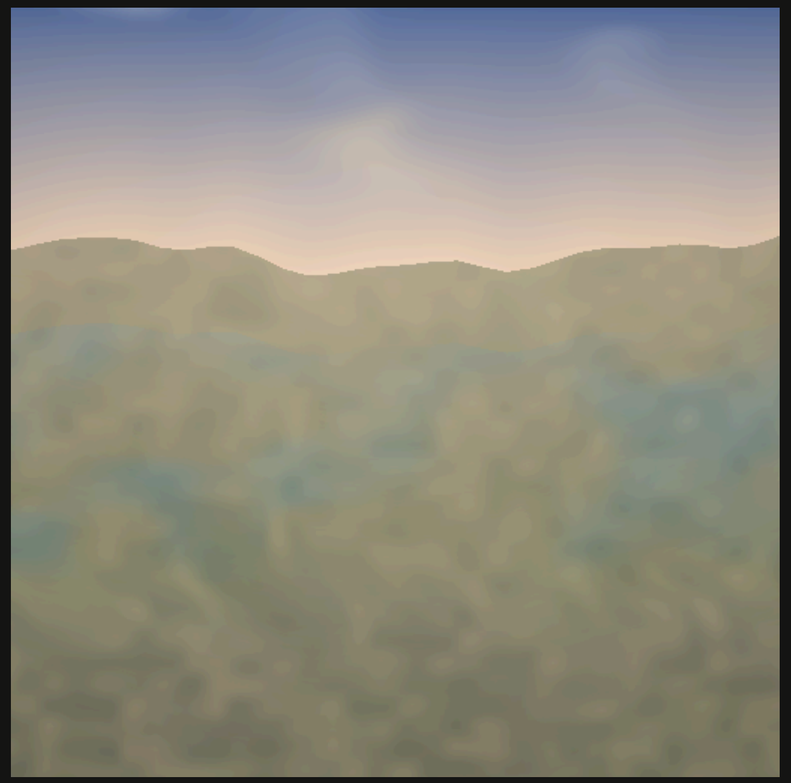
求，别骂

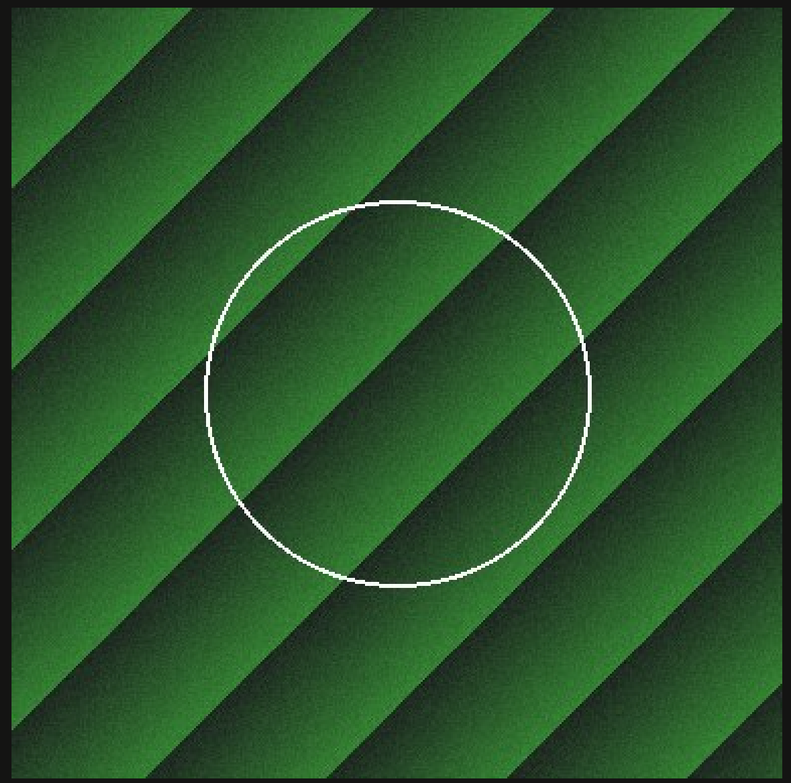
Everything built in 6 hours.

PNG

llm -image

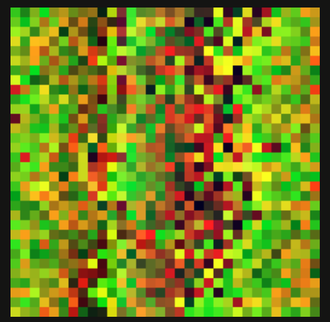
coast / forest / lake / mountain / meadow





forest / leaf

这是开始的效果
丑不拉几的



tree

llm -audio

WAV

Music sample: 01-launch-minimal.wav

竖耳朵听

Soundscape sample: 02-forest-morning.wav

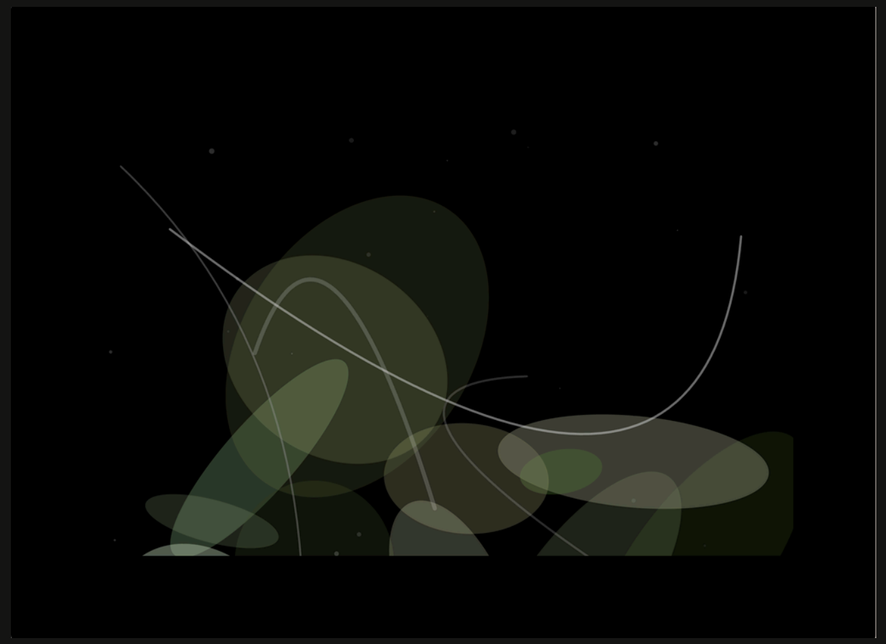
就好

真好

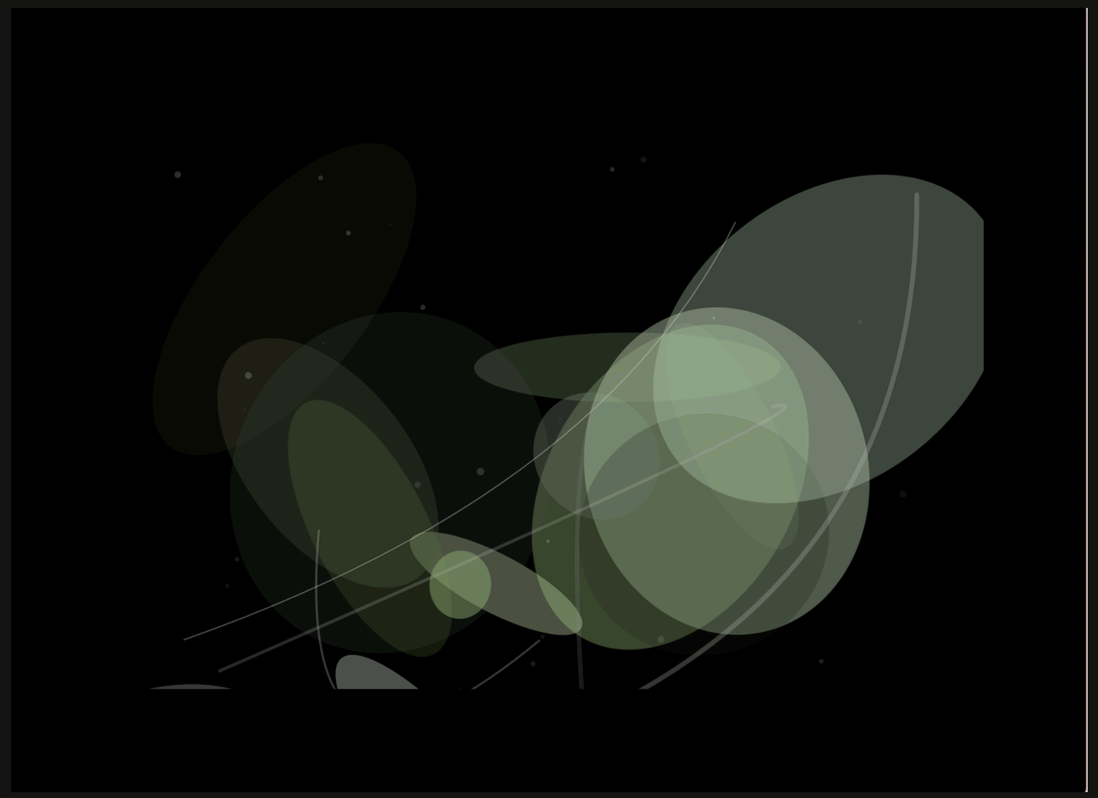
llm-video

MP4

这个也是 Tree



9s



llm-sensor

基于 loupe 框架，把 CSV 等表格类数据可视化呈现

loupe saas_metrics.csv · 9 cols · 30 rows

Search all columns...

MRR	USERS	CHURN_RATE	NPS_SCORE
$12,841	341	4.2%	66.3
avg · 350–41,200	avg · 7–1,102	avg · 0.6–13.5	avg · 36–92

COMPANY	COUNTRY	PLAN	MRR	USERS	CHURN_RATE	NPS_SCORE	ACTIVE
Acme Corp	USA	Enterprise	$12,400	342	2.1	72	✓
Borealis AI	Canada	Growth	$4,800	89	3.4	65	✓
Cirrus Cloud	UK	Starter	$890	14	8.2	48	✓
Drift Labs	Germany	Enterprise	$18,750	521	1.2	81	✓
Ember.io	Australia	Growth	$6,200	167	2.9	68	✓
Folio Tech	Japan	Starter	$350	8	12.1	41	✓
Galactic SaaS	USA	Enterprise	$31,000	890	0.8	88	✓
Helix Systems	France	Growth	$5,100	112	3.1	70	✓

Showing 1–30 of 30 rows

这里有个很优雅的小算法，
比 llm 生成的数据在收敛区间更合理的发挥
可以用作模拟的数据训练，token 还用得少！

Next
here is

Few
more things

Benchmark

Image/Editorial
PNG
6 ms
TTS/Launch
WAV/MP3
608 ms
Animation
MP4
507 ms
Sensor/ECG
CSV/HTML
24 ms

Lightweight proxy Benchmark

Case	Price	Latency	Quality
Image / Editorial PNG · 5.2 KB	**$0.00** local path	**6 ms** end-to-end wall clock	**85/100** structural proxy
TTS / Launch WAV · 107.7 KB	**$0.00** local path	**608 ms** end-to-end wall clock	**85/100** structural proxy
Audio / Music WAV · 107.7 KB	**$0.00** local path	**507 ms** end-to-end wall clock	**85/100** structural proxy
Sensor / ECG HTML · 490 B	**$0.00** local path	**24 ms** end-to-end wall clock	**85/100** structural proxy

Video is not in this run because the MP4 renderer is still on another branch. This kit only reports runnable main-branch cases.

llm + Harness = Agent

text-first LLM **Modality Harness** **?**

Thesis 1

过去二十年间，模拟流程逐渐被数字流程所取代
创新与杠杆效应密切相关，也就是以更少的资源创造更多价值
更细腻、更有创造力，以及更稳定的 infra，应该被创造和打磨

Thesis 2

我们今天看到的很多新产品都是升级换代的产品
我们认为，总是存在绝佳的机会，可以打造一些并非升级换代，而是完全不同的东西
总要有人像尼采点燃闪电， 总要有人像维特根斯坦长久如云漂泊

Thesis 3

The best ideas haven't even been though of yet.
Today, what is cheap, can be tomorrow, made smart.
总是 Build，会有好事发生

>_

Wittgenstein

产品发布：
wittgenstein.WTF

Thanks

presented by **+1** and **sym**

Researcher Builder Trier

以人为本的创新 以爱为基础的超级对齐

Wittgenstein

Wittgenstein

语言的边界，就是世界的边界
而在今天，我们可以用语言去生成世界

Wittgenstein 让文本模型通过 Harness 直接产出真实多模态文件
MAKE TEXT MODELS SPEAK IN MEDIA

New
release ‹/›

Moapacha/wittgenstein/releases

Built by:

Looking for collaboration!