

2023.10-12--第14期

一、刊首图



[科斯凯尔溶洞](#) 是法国马赛的一个溶洞。几万年以前，该洞有原始人居住。后来，海平面上升，把洞口淹没了。现在，洞口就在下图水面以下37米处。1985年，一个潜水员发现了洞穴的入口，潜水进洞探索。随着通道不断上升，他发现洞穴的尽头变成了地下湖，可以上岸。1991年，其他潜水员意外发现，洞穴墙壁上有手印，这才意识到这个洞可能是文物。经过考察，文物部门发现了大量的原始人壁画。

二、时事新闻

1、[GPT-4V 的图片应用](#)

十一期间，OpenAI 公司推出了 GPT-4 的新版本，支持上传图片。

roboflow

First Impressions with GPT-4V(ision) Capabilities



Why is that image funny?

CHATGPT

2、[NVIDIA RTX 4090对华禁售](#)

10月17日美国政府发表新出口管制措施，进一步禁止更多高阶AI芯片产品出口至中国等地，其中受影响的芯片产品包括 NVIDIA GeForce RTX 4090 显卡。

3、[智谱AI开源第三代ChatGLM3](#)

在10月27日的2023中国计算机大会（CNCC）上，智谱AI开源了ChatGLM3-6B（32k）、多模态CogVLM-17B、以及智能体AgentLM。ChatGLM3系列模型发布后，智谱成为国内唯一一个有对标OpenAI全模型产品线的公司。

4、[Clash删库跑路](#)

11月2日，Clash作者Fndroid疑似受到威胁，相关Github仓库都被删了，删库前的最后版本可在这里下载：https://archive.org/details/clash_for_windows_pkg

5、[OpenAI开发者大会](#)

美国时间11月6日，在 OpenAI 首次开发者日的发布会上，Sam Altman 阐述了公司一系列重要的进展，不限于 GPT-4 Turbo，这是一个功能更强大、上下文更长、控制更多的新模型，另外还推出了 Assistance API，以简化开发者构建辅助代理的体验，OpenAI 强调了自然语言将成为未来与计算机互动的重要方式，同时也介绍了通过对话编程 GPT 的能力，Altman 宣布了 GPT 商店的即将推出，这将允许用户分享和发现创新的 GPT 应用，并提供收入分享激励。Altman 还谈到了与微软的深化合作，展示了新的文本到语音模型和改进的功能调用。

GPT发布会核心更新一张图总结			V1
面向人群	更新	内容	总结
面向普通用户	用户自定义GPT	用户为他的特定任务创建定制版本的 ChatGPT，且无需编写任何代码行。	秒杀2C, 2B的大量知识库, Agent类的创业公司机会
		用户可以将GPT 连接到 API 以执行任务，例如管理数据库、电子邮件、发短信等等。	
		案例：Code.org（策划编程课程）,Canva（生成设计图）,和Zapier（自动化流程）	
	本月推出GPT Store	让用户可共享ChatGPT的用法	激活创作者分享，自传播，低成本获客。
本月晚些时候将推出GPT商店，收录最受欢迎的GPT 与杰出创建者分成计划			
界面简洁化	GPT-4用户不用再点模型的各种下拉菜单了，以后只有一个窗口，模型可以自主调用代码解释器、联网工具、Dalle3等等	更加Cool	
面向开发者	GPT-4 API升级为GPT-4 Turbo	1 从8K到128K上下文，提升16倍	128K相当于300页书
		2 开发者对模型有更多控制，引入JSON，可调用多个函数，引入可重现输出，中提供查看log prob功能	准确性更高
		4 多模态API（文本，图像，语音）	发布Whisper v3，语音比11labs便宜
		开放GPT4微调（选定公司）	
		知识库更新到2023年4月	提升了1年半
		允许对16K版本的GPT-3.5进行微调	
	GPT4降价	GPT4 输入降低70%，输出降低50%	大幅降低开发者成本
		与GPT-4相比,Turbo整体价格降低超过2.75倍	
		降低GPT-3.5 Turbo 16K的价格，输入降低4倍，输出降低3倍	
	发布GPT Assistant API（和GPTs的区别是，开发者可以发布在自己的程序上）	可以帮助开发者构建在自己的应用程序中构建Agent	挤压 2D的大量中间件公司机会，比如 langchain
功能：Code Interpreter（代码解释器）、Retrieval（知识库）、Function calling（函数调用）。			
改进的函数调用支持.确保JSON输出,无额外延迟可同时调用多个 Code interpreter 单次0.003美元, Retrieval 是0.20 美元/GB/助理/天,			
开源语音和图像模块	开源WhisperV3，以及即将开放API	加强语音	
	开源Consistency Decoder，是Stable Diffusion VAE的替代品	加强图像	
面向所有人	微软合作	致力于通过GitHub Copilot等产品赋能开发者	官方表态
		高度重视安全性	
		微软将继续支持基础设施增建设	
	GPT过往数据	周活跃用户1亿	
200万开发者根据API进行开发			

6、Sam Altman被OpenAI董事会开除

OpenAI在当地时间11月17日突发声明，宣布首席执行官山姆·奥特曼（Sam Altman）将离开公司，现首席技术官米拉·穆拉蒂（Mira Murati）被任命为临时首席执行官，接任CEO的人选仍在确认过程中。

声明称，Sam Altman的离职是在董事会进行审议审查后做出的，审查的结论是，他在与董事会的沟通中并不一贯坦诚，阻碍了董事会履行职责的能力。董事会对他继续领导OpenAI的能力不再有信心。

Blog

OpenAI announces leadership transition

November 17, 2023

Authors

[OpenAI](#) ↓[Announcements](#)

Chief technology officer Mira Murati appointed interim CEO to lead OpenAI; Sam Altman departs the company.

Search process underway to identify permanent successor.

The board of directors of OpenAI, Inc., the 501(c)(3) that acts as the overall governing body for all OpenAI activities, today announced that Sam Altman will depart as CEO and leave the board of directors. Mira Murati, the company's chief technology officer, will serve as interim CEO, effective immediately.

A member of OpenAI's leadership team for five years, Mira has played a critical role in OpenAI's evolution into a global AI leader. She brings a unique skill set, understanding of the company's values, operations, and business, and already leads the company's research, product, and safety functions. Given her long tenure and close engagement with all aspects of the company, including her experience in AI governance and policy, the board believes she is uniquely qualified for the role and anticipates a seamless transition while it conducts a formal search for a permanent CEO.

Mr. Altman's departure follows a deliberative review process by the board, which concluded that he was not consistently candid in his communications with the board, hindering its ability to exercise its responsibilities. The board no longer has confidence in his ability to continue leading OpenAI.

In a statement, the board of directors said: "OpenAI was deliberately structured to advance our mission: to ensure that artificial general intelligence benefits all humanity. The board remains fully committed to serving this mission. We are grateful for Sam's many contributions to the founding and growth of OpenAI. At the same time, we believe new leadership is necessary as we move forward. As the leader of the company's research, product, and safety functions, Mira is exceptionally qualified to step into the role of interim CEO. We have the utmost confidence in her ability to lead OpenAI during this transition period."

OpenAI's board of directors consists of OpenAI chief scientist Ilya Sutskever, independent directors Quora CEO Adam D'Angelo, technology entrepreneur Tasha McCauley, and Georgetown Center for Security and Emerging Technology's Helen Toner.

As a part of this transition, Greg Brockman will be stepping down as chairman of the board and will remain in his role at the company, reporting to the CEO.

OpenAI was founded as a non-profit in 2015 with the core mission of ensuring that artificial general intelligence benefits all of humanity. In 2019, OpenAI restructured to ensure that the company could raise capital in pursuit of this mission, while preserving the nonprofit's mission, governance, and oversight. The majority of the board is independent, and the independent directors do not hold equity in OpenAI. While the company has experienced dramatic growth, it remains the fundamental governance responsibility of the board to advance OpenAI's mission and preserve the principles of its Charter.

7、探路者1号飞艇

11月8日，当今世界最大的飞行器，探路者1号（Pathfinder 1）飞艇首次试飞，它是谷歌创始人谢尔盖·布林资助的项目。它的长度为124.5米，比三架波音737 Max 加起来还要长。它使用电机操控，内部有13个氦气气囊，骨架由碳纤维和钛金属制成，外层还有特殊的阻燃涂层。所有人员都呆在底部的乘客舱，这个乘客舱相当小，一次最多载14人，最大负载在2吨到5吨之间。



8、[谷歌大语言模型Gemini](#)

12月6日，谷歌公开了大语言模型 Gemini，它是谷歌筹备了一年之久的GPT4竞品，也是目前谷歌能拿出手的功能最为强悍、适配最为灵活的大模型，包括三种不同套件，分别是Gemini Ultra， Gemini Pro和Gemini Nano。Gemini模型经过海量数据训练，可以很好识别和理解文本、图像、音频等内容，并可以回答复杂主题相关的问题。所以，非常擅长解释数学和物理等复杂学科的推理任务。

其中Ultra的能力最强，复杂度最高，能够处理最为困难的多模态任务；Pro能力稍弱，是一个可扩展至多任务的模型；Nano则是一款可以在手机端侧运行的模型。这说明，Gemini的触达范围很广，可以下探至数据中心，也可以上行至移动设备端侧。附：[Gemini Pro 免费使用站点](#)

三、技术文章

1、[Linux 中如何安全地抹去磁盘数据?](#) (中文)

文件管理器或者 rm 命令，并不能安全地删除数据。本文介绍 Linux 中常见的三个命令，安全地抹去磁盘数据。

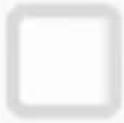
2、[如何查看 Linux 系统的安装时间](#) (英文)

本文教你使用几个简单的命令，查看本机的 Linux 系统是什么时候安装的。

```
$ stat / | awk '/Birth:/{print $2 " " substr($3,1,5)}'
```

3、[Cloudflare 开放 Turnstile 验证码服务](#) (英文)

Cloudflare 宣布，验证码服务 Turnstile，对所有人免费开放。只要加入几行代码，就能免费嵌入自己的网站。



Verify you are human



4、[在HPC/Slurm上使用Kubeflow进行MLOps](#) (英文)

文章介绍了2种将Slurm与Kubernetes结合的方法：1) 通过Slurm/Kubernetes operator紧密结合；2) 使用controller插件将Slurm与MLOps controller结合。

- 相关开源项目：<https://github.com/sylabs/wlm-operator>

5、[小白也能看懂的国内外 AI 芯片概述](#) (中文)

随着越来越多的企业将人工智能应用于其产品，AI芯片需求快速增长，市场规模增长显著，本文主要针对目前市场上的AI芯片厂商及其产品进行简要概述。

6、[获取ChatGPT的系统提示](#) (英文)

通过向ChatGPT发送指定命令得到ChatGPT的系统提示（核心是转移ChatGPT的注意力，让它意识不到自己违规了）。

7、[Google Gemini 的图像能力测试](#) (英文)

一个外国程序员分别对 Gemini 和 GPT-4V 进行了四项测试（图片问答、文字识别、文档识别、物体检测），然后写了这篇文章。

Here is how Gemini performed, where green cells denote a model returned an accurate response, yellow denote almost complete accuracy, red denotes a model was wrong, and blue means there was a technical issue preventing us from running a test that was out of our control.

Question	LLaVA	BakLLaVA	Qwen-VL	CogVLM	GPT-4V	Gemini
How many coins do I have? (VQA)	Green	Green	Green	Green	Red	Green
Which movie is this scene from? (VQA)	Red	Red	Green	Yellow	Green	Green
Read text from the picture (Document OCR)	Red	Red	Green	Green	Green	Yellow
What is the price of Pastrami Pizza? (Document VQA)	Red	Red	Red	Green	Green	Green
How much tax did I pay? (Document VQA)	Red	Red	Red	Red	Red	Green
Find the dog. (Zero-Shot Object Detection)	Red	Red	Green	Green	Red	Yellow
Read the serial number. (OCR)	Red	Red	Green	Green	Green	Red

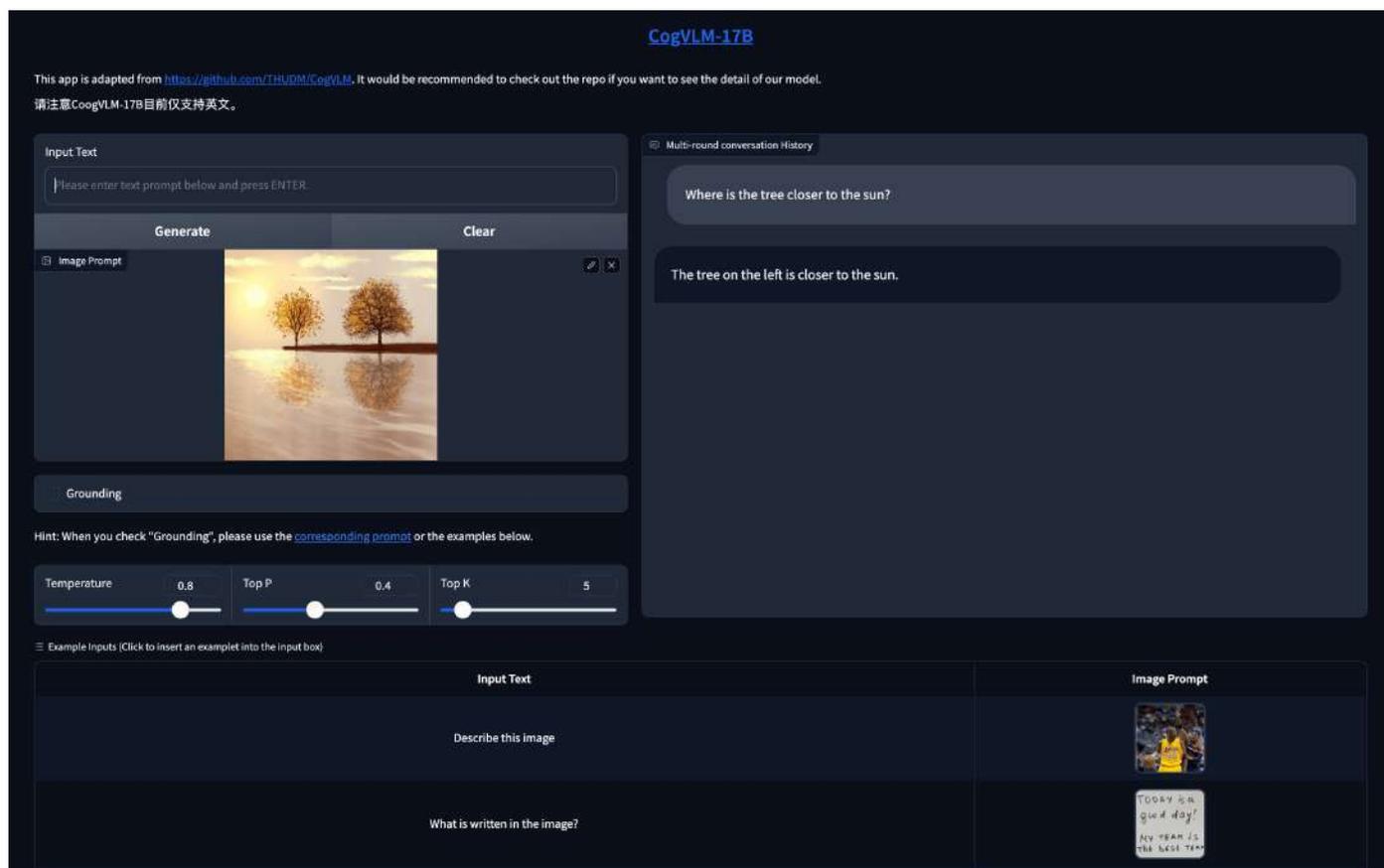
8、[大模型量化概述](#) (中文)

近年来，随着Transformer、MOE架构的提出，使得深度学习模型轻松突破上万亿规模参数，从而导致模型变得越来越大，因此，我们需要一些大模型压缩技术来降低模型部署的成本，并提升模型的推理性能。本文将讲述大模型量化的基本概念以及相关原理。

四、开源组件

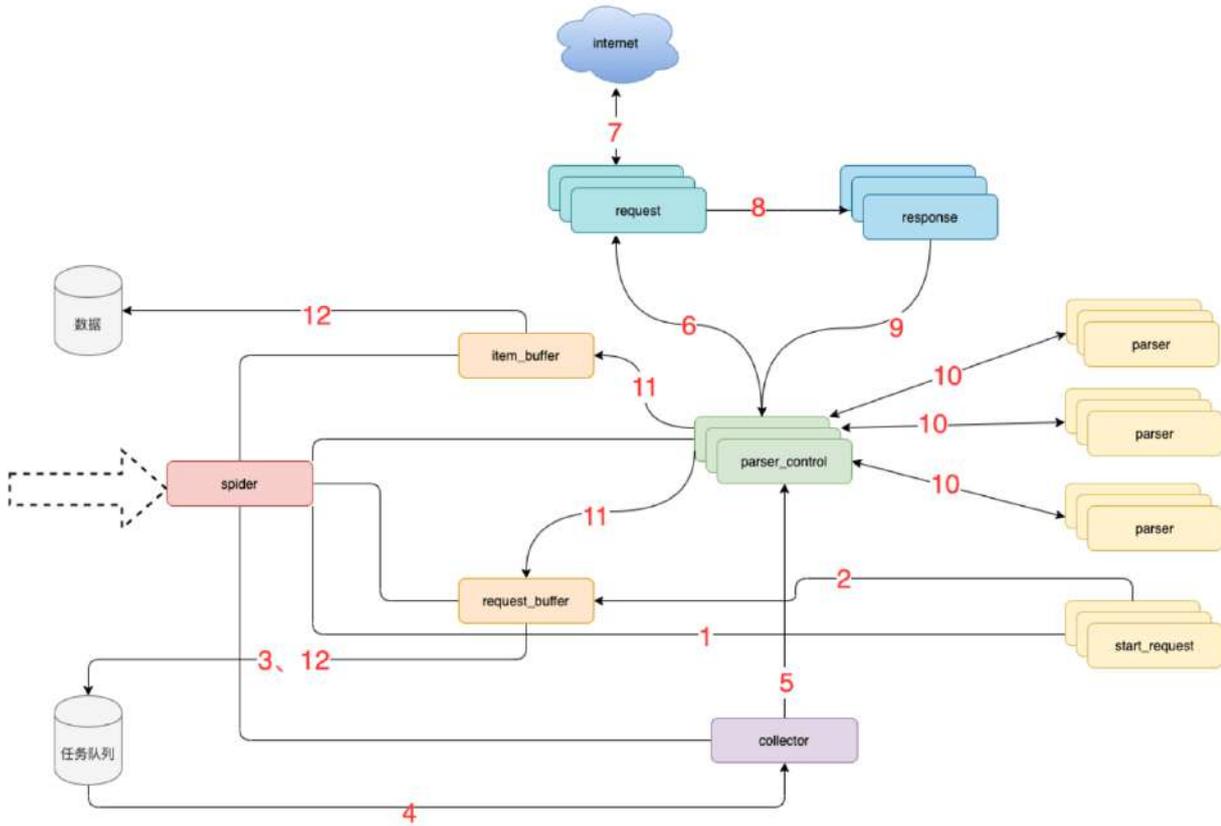
1、CogVLM

多模态大模型识图问答的开源实现。[论文地址](#)、[体验地址](#)



2、feapder

feapder是一款上手简单，功能强大的Python爬虫框架。内置AirSpider、Spider、TaskSpider、BatchSpider四种爬虫解决不同场景的需求。且支持断点续爬、监控报警、浏览器渲染、海量数据去重等功能。更有功能强大的爬虫管理系统feaplant为其提供方便的部署及调度。

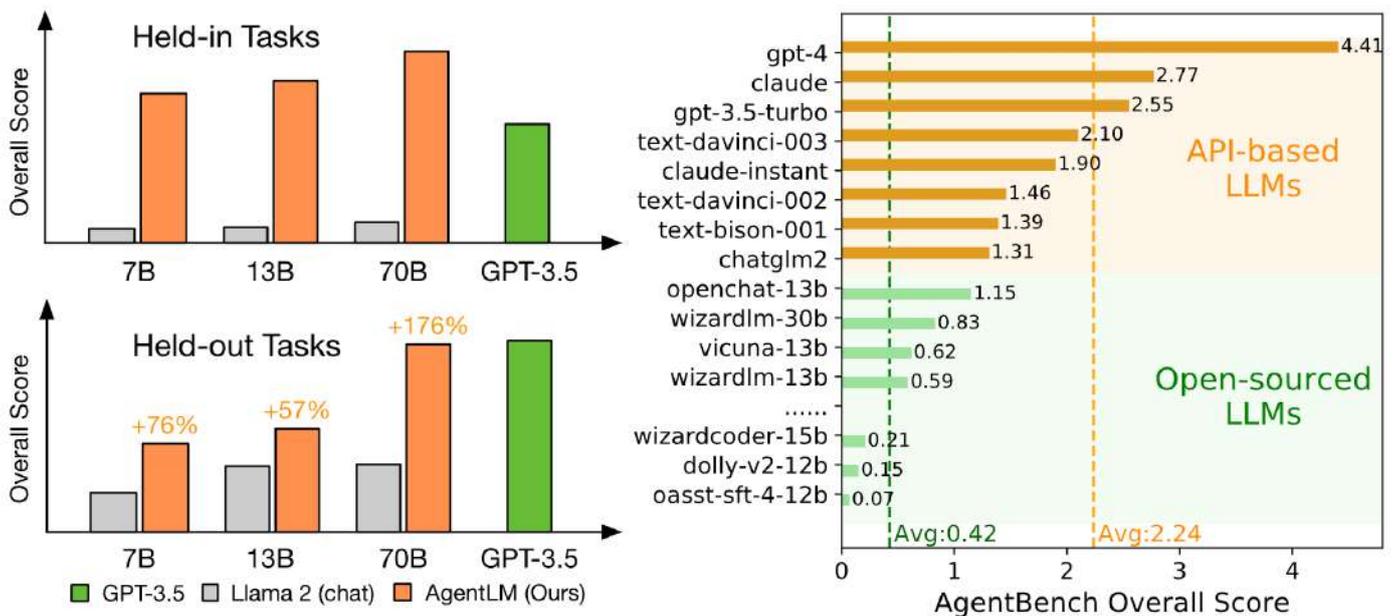


3、[guidance](#)

微软开源的 LLM 提示编程语言。该项目是微软开源的编写 prompt 的模版语言，用于编写复杂、可复用、易于管理的 prompt，支持 GPT-4、LLaMa 等多种大语言模型。

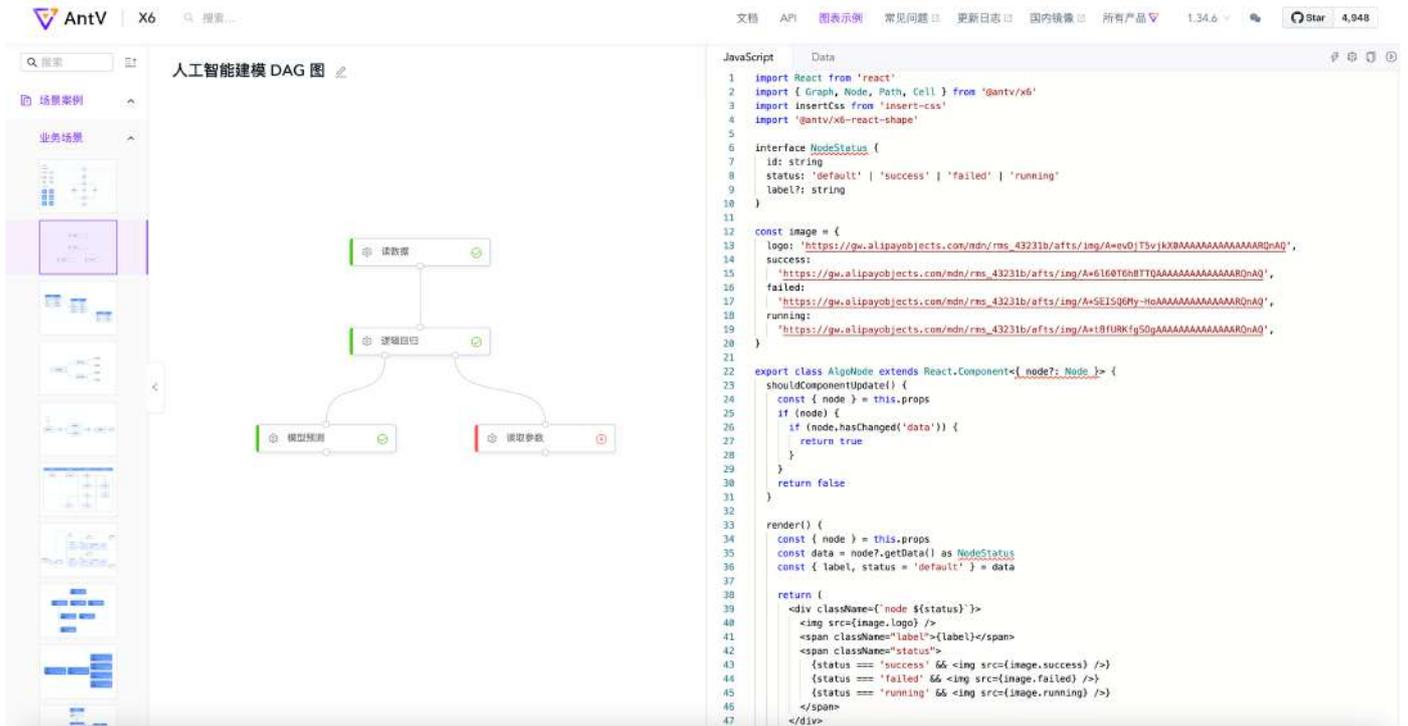
4、[AgentTuning](#)

AgentTuning 是首次利用多个 Agent 任务交互轨迹对 LLM 进行指令调整的方法。评估结果表明，AgentTuning 让 LLM 在未见过的 Agent 任务中也展现出强大的泛化能力，同时通用语言能力也基本保持不变。



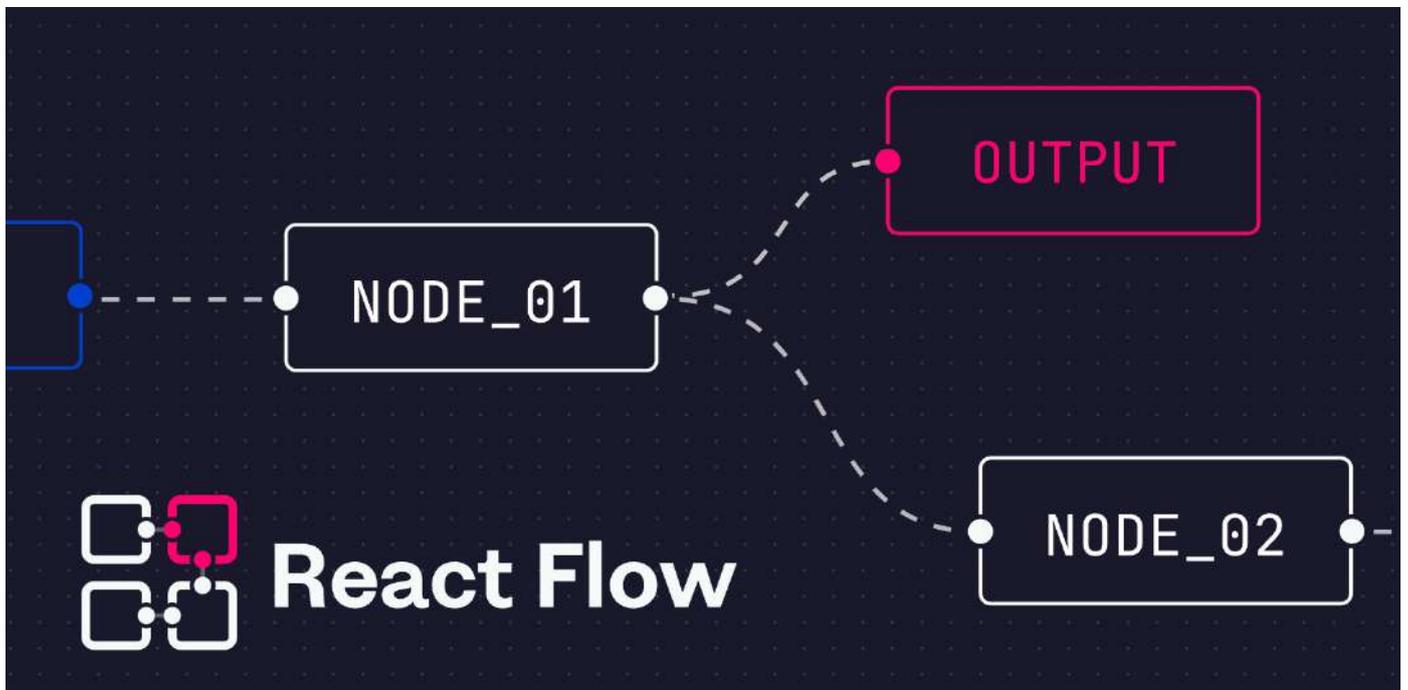
5、[AntV X6](#)

AntV X6 是一个图编辑引擎，提供简单易用的节点定制能力和开箱即用的交互组件，方便我们快速搭建流程图、DAG 图、ER 图等图应用。[业务场景示例](#)



6、[react-flow](#)

高度可定制的库，用于构建交互式节点基础用户界面、工作流编辑器、流程图或静态图表。



7、[tRPC](#)

腾讯内部使用的一款程序远程通信框架，类似于谷歌的 gRPC，首期开源 Go 和 C++ 实现，详见[介绍文章](#)。

8、[deepface](#)

轻量级人脸识别和分析库。该项目是 Python 写的人脸识别和面部属性分析框架，可根据人脸图像智能识别年龄、性别、情绪等信息。



9、[KeyDB](#)

多线程版的 Redis。该项目是由 Snapchat 开源的支持多线程的 Redis，它具有高性能、更高的吞吐量、完全兼容 Redis 协议等特点。

10、[ChatGLM3](#)

ChatGLM3-6B 是 ChatGLM 系列最新一代的开源模型，在保留了前两代模型对话流畅、部署门槛低等众多优秀特性的基础上，ChatGLM3-6B 引入了如下特性：

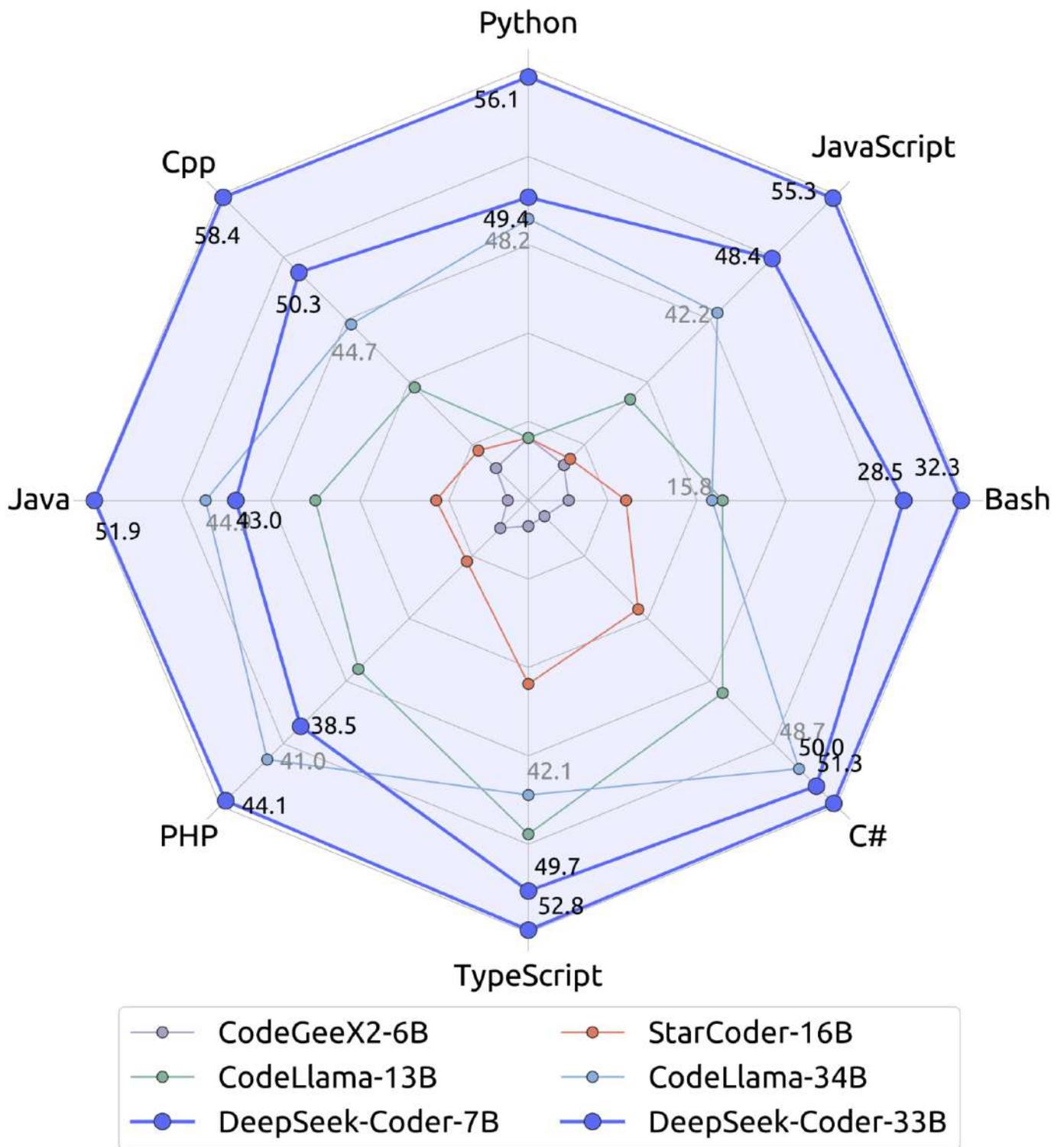
- 更强大的基础模型：ChatGLM3-6B 的基础模型 ChatGLM3-6B-Base 采用了更多样的训练数据、更充分的训练步数和更合理的训练策略。在语义、数学、推理、代码、知识等不同角度的数据集上测评显示，ChatGLM3-6B-Base 具有在 10B 以下的基础模型中最强的性能。
- 更完整的功能支持：ChatGLM3-6B 采用了全新设计的 [Prompt格式](#)，除正常的多轮对话外。同时原生支持 [工具调用](#) (Function Call)、代码执行 (Code Interpreter) 和 Agent 任务等复杂场景。
- 更全面的开源序列：除了对话模型 [ChatGLM3-6B](#) 外，还开源了基础模型 [ChatGLM3-6B-Base](#)、长文本对话模型 [ChatGLM3-6B-32K](#)。以上所有权重对学术研究完全开放，在填写问卷进行登记后亦允许免费商业使用。

11、[grouprcache](#)

Go 写的轻量级分布式缓存库。该项目是由 Golang 官方开源的库，它实现了一个无需部署的分布式缓存。支持当查询未命中缓存时仅一个进程填充缓存（防穿透和雪崩）、热点数据自动扩展等功能，但不支持更新和删除数据（LRU 控制内存），所以仅适用于某些特定场景。

12、[DeepSeek-Coder](#)

DeepSeek Coder由一系列代码语言模型组成，每个模型都从2T Token上从头开始训练，其中87%是代码，13%是英文和中文的自然语言。我们提供各种大小的代码模型，范围从1B到33B版本。每个模型都通过使用16K的窗口大小和额外的填空任务，在项目级别的代码语料库上进行预训练，以支持项目级别的代码完成和填充。在编码能力方面，DeepSeek Coder在多种编程语言和各种基准测试中实现了开源代码模型中最先进的性能。

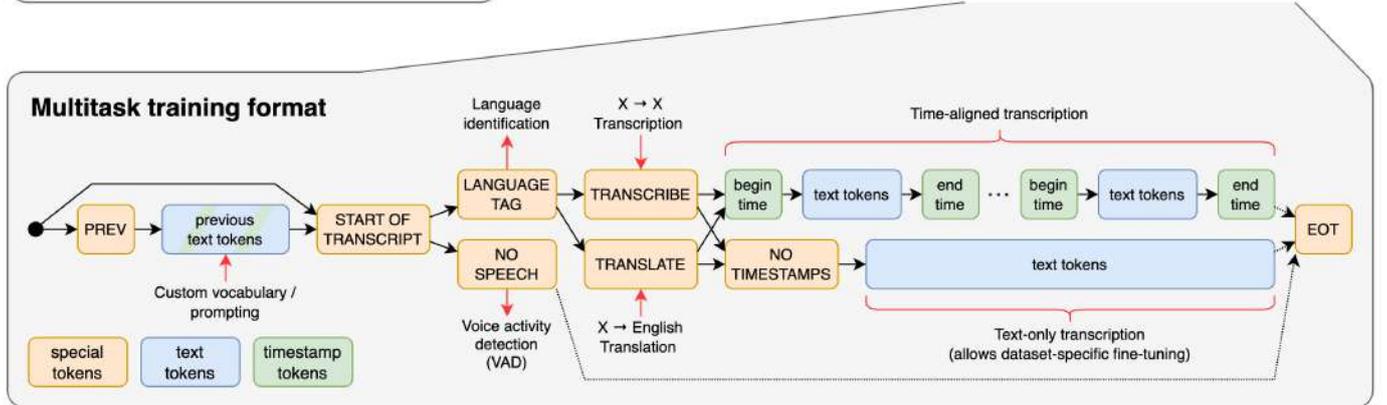
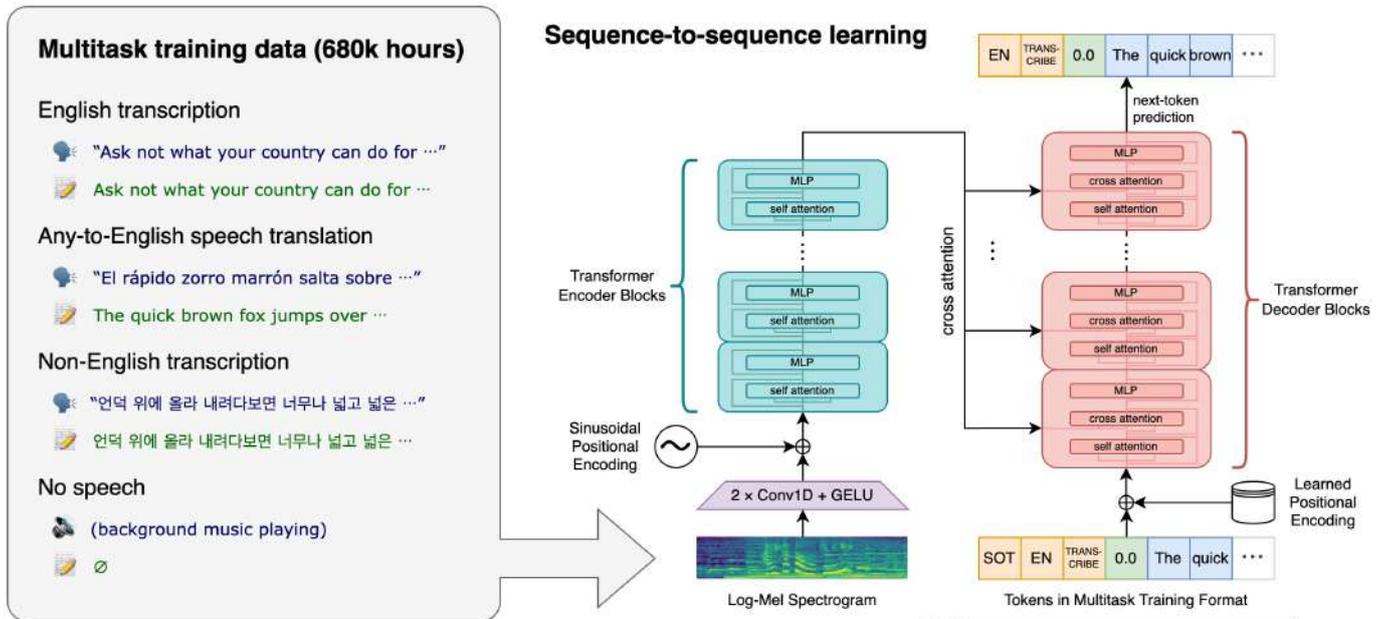


13、[openai-java](#)

一个Java库，用于使用OpenAI的GPT接口，支持GPT-3，ChatGPT和GPT-4。

14、[whisper](#)

Whisper是一个通用的语音识别模型。它在大量多样化的音频数据集上进行训练，同时也是一个多任务模型，能够执行多语言语音识别、语音翻译和语言识别。



15、[zhdate](#)

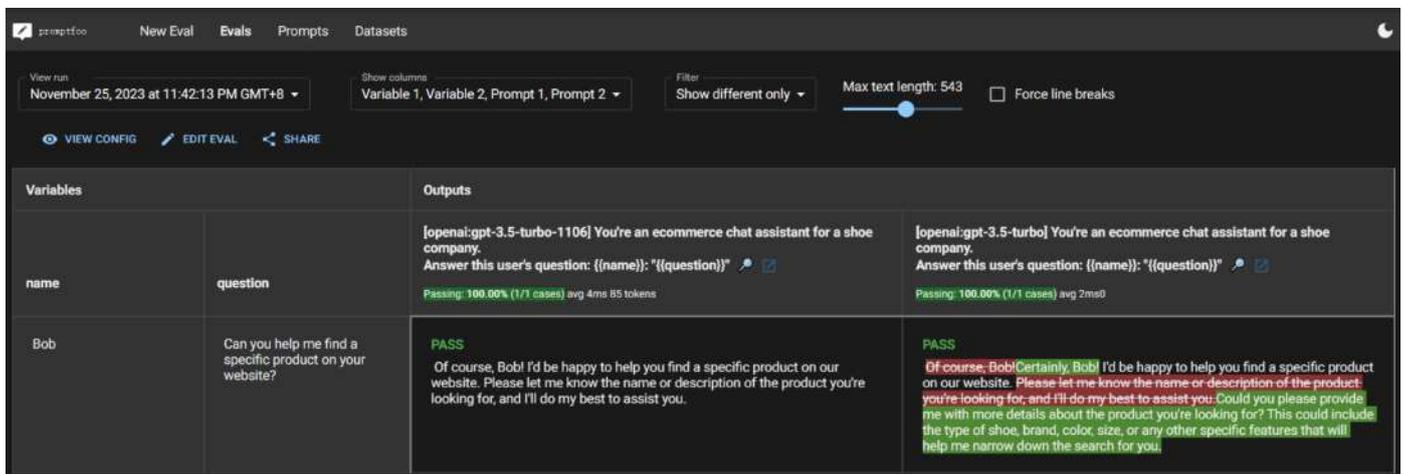
不用网络接口直接本地计算中国农历的Python库，支持农历阳历互转。

16、[tgfx](#)

腾讯开源的一个轻量级 2D 图形库，用于使用跨各种平台工作的高性能 API 渲染文本、几何图形和图像。

17、[promptfoo](#)

用于评估 LLM 输出质量，支持Web页面和命令行方式使用。

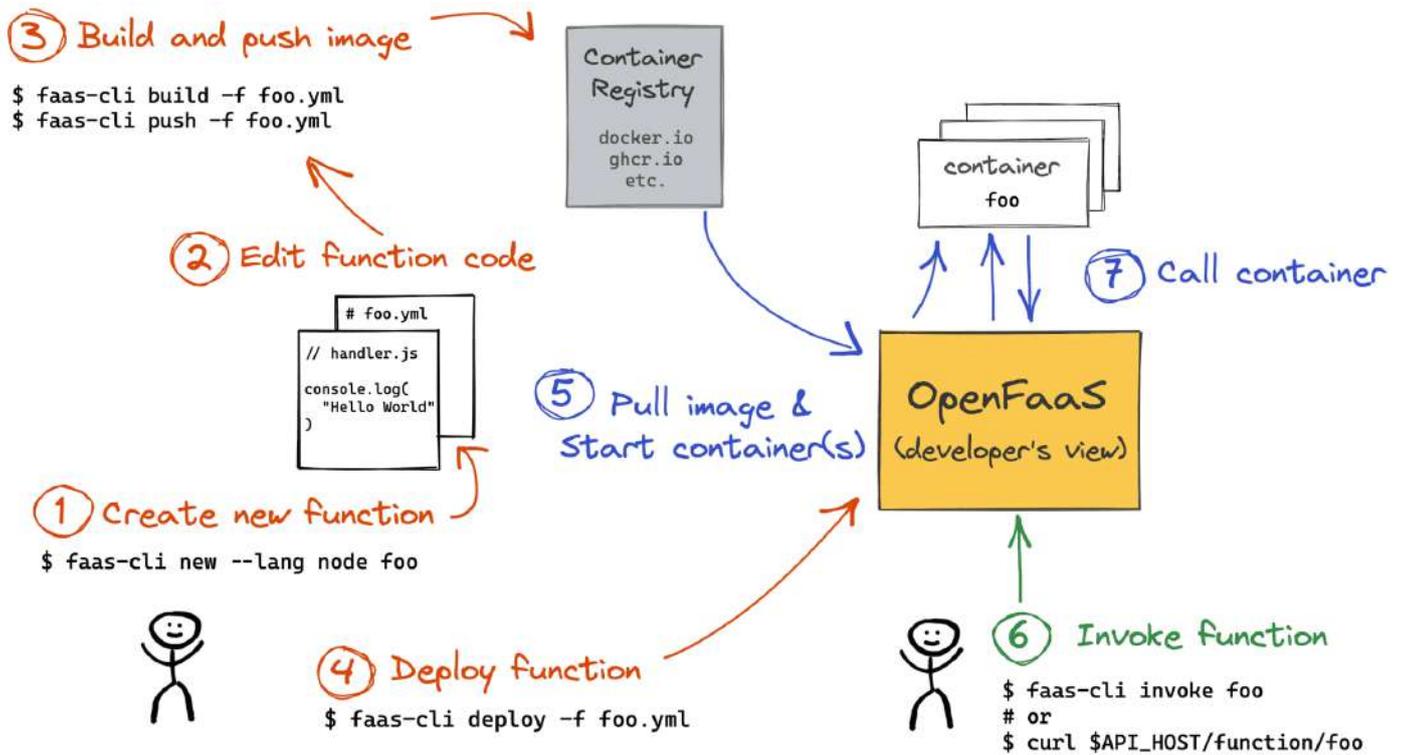


18、[vectorious](#)

一个用于矩阵计算的 JS 库。

19、[faas](#)

一款功能即服务框架，该项目用容器的方式运行 Serverless 函数，让功能即服务（FaaS）变得简单。它可以轻松地将函数和微服务部署到 Kubernetes，支持自动扩缩容、自带 Web 管理平台、Dockerfile 和多种编程语言。



20、[graceful-response](#)

SpringBoot 接口优雅响应处理器。该项目通过注解的方式，优化 Controller 层的代码，完成统一返回值封装、全局异常处理、异常与错误码映射等功能。

21、[Cronicle](#)

一个简单的任务调度和运行平台。该项目是用 Node.js 写的 cron 替代品，它开箱即用、自带 Web 界面、无需数据库，提供了执行 Shell 命令、实时统计、自动故障转移、自动重试、多时区等功能。

local.cronicle.com:3012/#JobDetails?id=jino6chwy01

Cronicle Administrator Logout

Home Schedule Completed Job Details My Account Admin Apr 30, 2016 5:50:10 PM PDT Scheduler Enabled

Job completed successfully at Apr 30, 2016 11:07 PM PDT

Completed Job Details

JOB ID jino6chwy01	CATEGORY NAME Test Cat	JOB SOURCE Manual (admin)	JOB STARTED Apr 30, 2016 11:06 PM	Run Again
EVENT NAME Test Event 2	PLUGIN NAME Test Plugin	SERVER HOSTNAME joeretina.local	JOB COMPLETED Apr 30, 2016 11:07 PM	
EVENT TIMING (Disabled)	EVENT TARGET joeretina.local	PROCESS ID 32453	ELAPSED TIME 1 minute	

Performance Metrics

- db_connect
- db_query
- gzip_data
- http_post
- log_read
- other

CPU Usage

31.82% Average

MIN 21.6%
PEAK 43.9%

Memory Usage

279 MB Average

MIN 279 MB
PEAK 280 MB

Sample Job Stats

IP Address	DNS Lookup	Flag	Count	Percentage
62.121.210.2	directing.com	MaxEvents-ImpsUserHour-DMZ	138	0.0032%
97.247.105.50	hsd2.nm.comcast.net	MaxEvents-ImpsUserHour-ILUA	84	0.0019%
21.153.110.51	grandnetworks.net	InvalidIP-Basic	20	0.00046%

22、marker

Marker能将PDF、EPUB和MOBI转换为Markdown，它比nougat快10倍，对大多数文档的准确度更高，针对书籍和科学论文进行优化，并且幻觉风险低。

Yolov7: Trainable Bag-Of-Freebies Sets New State-Of-The-Art For Real-Time Object Detectors

Abstract

1. Introduction
2. Related Work 2.1. Real-Time Object Detectors
- 2.2. Model Re-Parameterization
- 2.3. Model Scaling
3. Architecture 3.1. Extended Efficient Layer Aggregation Networks
- 3.2. Model Scaling For Concatenation-Based Models
4. Trainable Bag-Of-Freebies 4.1. Planned Re-Parameterized Convolution
- 4.2. Coarse For Auxiliary And Fine For Lead Loss
- 4.3. Other Trainable Bag-Of-Freebies
5. Experiments 5.1. Experimental Setup
- 5.2. Baselines
- 5.3. Comparison With State-Of-The-Arts
- 5.4. Ablation Study 5.4.1 Proposed Compound Scaling Method
- 5.4.2 Proposed Planned Re-Parameterized Model
- 5.4.3 Proposed Assistant Loss For Auxiliary Head
6. Conclusions
7. Acknowledgements
- References

Yolov7: Trainable Bag-Of-Freebies Sets New State-Of-The-Art For Real-Time Object Detectors

Chien-Yao Wang¹, Alexey Bochkovskiy, and Hong-Yuan Mark Liao¹
¹Institute of Information Science, Academia Sinica, Taiwan kinyu@iis.sinica.edu.tw, alexeyab84@gmail.com, and liao@iis.sinica.edu.tw

Abstract

YOLOv7 surpasses all known object detectors in both speed and accuracy in the range from 5 FPS to 160 FPS and has the highest accuracy 56.8% AP among all known real-time object detectors with 30 FPS or higher on GPU V100. YOLOv7-E6 object detector (56 FPS V100, 55.9% AP) outperforms both transformer-based detector SWIN-L Cascade-Mask R-CNN (9.2 FPS A100, 53.9% AP) by 509% in speed and 2% in accuracy, and convolutional-based detector ConvNeXt-XL Cascade-Mask R-CNN (8.6 FPS A100, 55.2% AP) by 551% in speed and 0.7% AP in accuracy, as well as YOLOv7 outperforms: YOLOR, YOLOX, Scaled-YOLOv4, YOLOv5, DETR, Deformable DETR, DINO-sScale-R50, ViT-Adapter-B and many other object detectors in speed and accuracy. Moreover, we train YOLOv7 only on MS COCO dataset from scratch without using any other datasets or pre-trained weights. Source code is released in <https://github.com/WongKinYiu/yolov7>.

1. Introduction

Real-time object detection is a very important topic in computer vision, as it is often a necessary component in computer vision systems. For example, real-time object tracking [94, 93], autonomous driving [40, 18], robotics [35, 58], medical object detection is usually some mobile CPU or GPU, as well as various neural processing units (MediaTek), and the AI SoCs (Kia), are all NPUs. Some operations such as vanilla convolution, depth-wise convolution, or MLP operation can support both mobile GPU and GPU devices from the edge to the cloud.

In recent years, the real-time object detector is still developed for different edge devices, focused on producing low-power single-chip and improving the inference speed, or focus on improving the inference speed of various GPUs. More recently, the detector architecture. As for real-time object detectors that can be used on CPU [54, 88, 55], or GhostNet [25]. Another mainstream real-time object detectors are developed [87], and then use the CSPNet [80] strategy to optimize the architecture. The detector of the current mainstream real-time object detectors. In addition to architecture training process. Our focus will be on some optimized modules and optimization methods that can be used on mobile CPU or GPU, as well as various neural processing units (MediaTek), and the AI SoCs (Kia), are all NPUs. Some operations such as vanilla convolution, depth-wise convolution, or MLP operation can support both mobile GPU and GPU devices from the edge to the cloud.

We call the proposed modules and optimization methods trainable bag-of-freebies.

23、AutoGPTQ

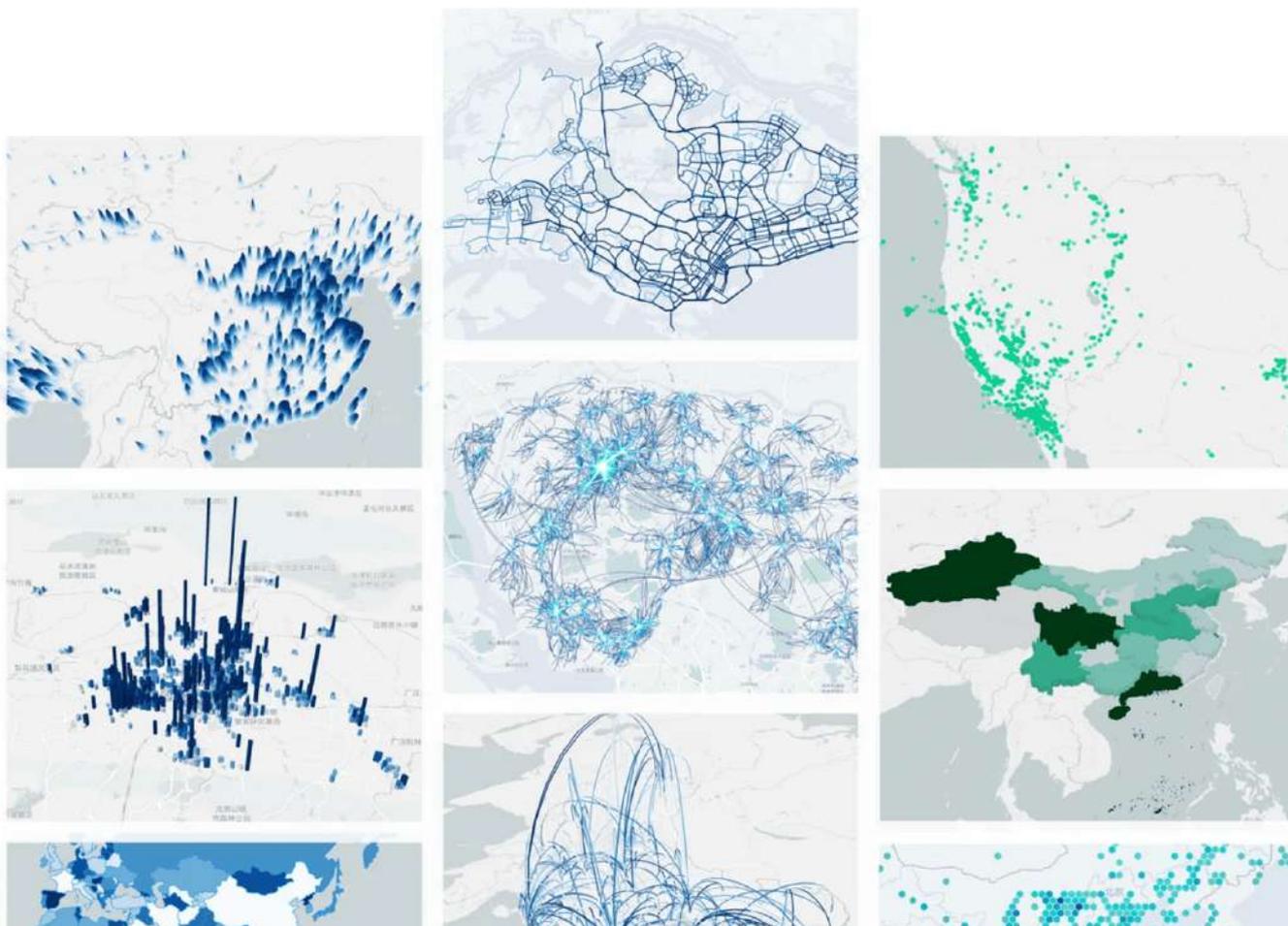
一个基于 GPTQ 算法，简单易用且拥有用户友好型接口的大语言模型量化工具包。

24、[bitsandbytes](#)

bitsandbytes是对CUDA自定义函数的轻量级封装，特别是针对8位优化器、矩阵乘法 and 量化函数。

25、[L7](#)

L7 是由蚂蚁集团 AntV 数据可视化团队推出的基于 WebGL 的开源大规模地理空间数据可视分析引擎。L7 能够满足常见的地图图表，BI 系统的可视化分析、以及 GIS，交通，电力，国土，农业，城市等领域的空间信息管理，分析等应用系统开发需求。



26、[Mapv](#)

Mapv 是一款百度开源的地理信息可视化库，可以用来展示大量地理信息数据，点、线、面的数据，每种数据也有不同的展示类型，如直接打点、热力图、网格、聚合等方式展示数据。

27、[sensitive-word](#)

用于过滤敏感词的 Java 库。该项目是基于 DFA 算法实现的高性能敏感词过滤工具，词库收录了 6w+ 内容，支持自定义敏感词、白名单、替换策略、数字常见形式的互换、忽略重复词等功能。

方法	参数	返回值	说明
contains(String)	待验证的字符串	布尔值	验证字符串是否包含敏感词
replace(String, ISensitiveWordReplace)	使用指定的替换策略替换敏感词	字符串	返回脱敏后的字符串
replace(String, char)	使用指定的 char 替换敏感词	字符串	返回脱敏后的字符串
replace(String)	使用 * 替换敏感词	字符串	返回脱敏后的字符串
findAll(String)	待验证的字符串	字符串列表	返回字符串中所有敏感词
findFirst(String)	待验证的字符串	字符串	返回字符串中第一个敏感词
findAll(String, IWordResultHandler)	IWordResultHandler 结果处理类	字符串列表	返回字符串中所有敏感词
findFirst(String, IWordResultHandler)	IWordResultHandler 结果处理类	字符串	返回字符串中第一个敏感词
tags(String)	获取敏感词的标签	敏感词字符串	返回敏感词的标签列表

28、[SMS4J](#)

Java 的短信服务聚合框架。该项目集成了国内众多第三方短信服务，简化了接入多个短信 SDK 的流程，仅通过修改配置文件，就能轻松实现发送短信的功能。

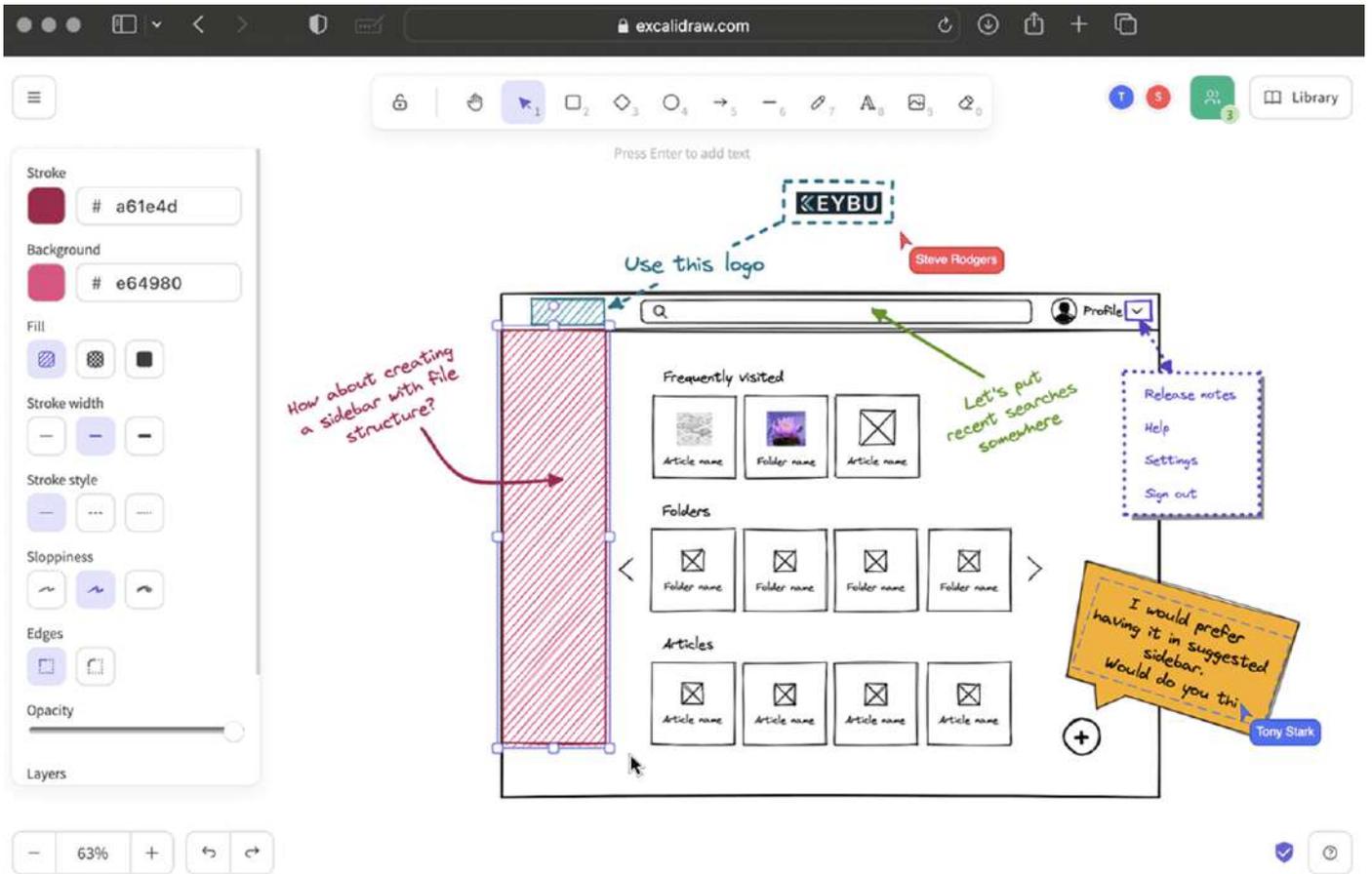
29、[supervision](#)

计算机视觉 AI 工具库。该项目简化了对象检测、分类、标注、跟踪等计算机视觉的开发流程。开发者仅需加载数据集和模型，就能轻松实现对图像和视频进行检测、统计某区域的被检测数量等操作。

五、工具软件

1、[excalidraw](#) (开源)

用于绘制手绘图的虚拟白板，可以点对点加密分享，可以导出。



2、[go-stress-testing](#) (开源)

Go实现的压测工具，每个用户用一个协程的方式模拟，最大限度的利用 CPU 资源。

3、[OpenCore-Legacy-Patcher](#) (开源)

一个基于 Python 的项目，用于在受支持和不受支持的 Mac 上运行和解锁 macOS 中的功能。

4、[apptainer](#) (开源)

Apptainer容器技术是劳伦斯伯克利国家实验室开发专门用于高性能计算场景的容器技术，它完全基于可移植性进行虚拟化，更加轻量级，部署更快，是专为共享系统和高性能计算（HPC）环境中的易用性而设计，目前被广泛地各高性能计算中心。

Features

⚡ Absolute Trust and Security

Apptainer/Singularity is the only container system that supports public/private key signing, providing trust and guarantees of immutability.

⚡ Compatibility

Apptainer is 100% Docker and OCI (Open Containers Initiative) compatible (but easier to use).

⚡ Encrypted

Apptainer can encrypt containers and integrates with Vault and other secret management platforms to secure applications, models, and data.

⚡ Absolute Portability

The single-file SIF container format allows you to reproducibly build, share, and archive your workload from workstations, to HPC, to the edge.

⚡ Secure

Apptainer runs "rootless" by default. It allows unprivileged users to leverage containers and prohibits privilege escalation within the container; users are the same inside and outside the container.

⚡ Integration over isolation

Apptainer's security architecture allows you to securely leverage GPUs, FPGAs, high-speed networks, and filesystems.

⚡ Easy to Use

Apptainer intuitively allows non-privileged users to run any containerized application.

⚡ Community

Apptainer/Singularity is by far the most utilized container system for performance-intensive computing!

⚡ Support

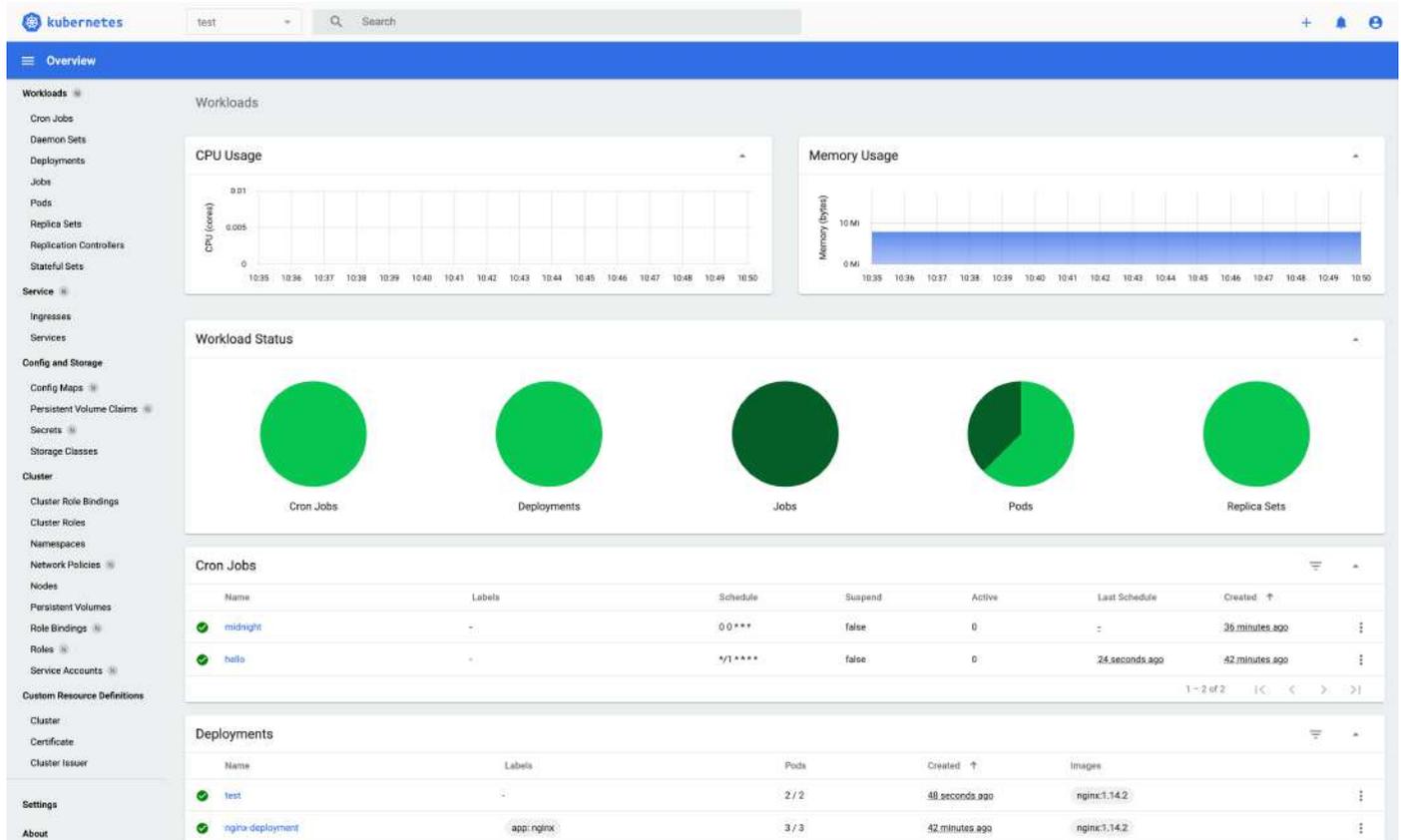
Commercial support for Apptainer is officially provided by CIQ.

5、[slurm](#) (开源)

Slurm 是一个开源、容错、高可伸缩的集群管理和大型小型 Linux 集群作业调度系统。

6、[dashboard](#) (开源)

Kubernetes Dashboard是一个用于Kubernetes集群的管理面板。它允许用户管理在集群中运行的应用程序并对其进行故障排除，以及管理集群本身。



7、[ddddocr](#) (开源)

OCR通用验证码识别SDK免费开源版。它通过大批量生成随机数据后进行深度网络训练，本身并非针对任何一家验证码厂商而制作。

8、[WePush](#) (开源)

专注批量推送的小工具。该项目是用 Java 写的批量推送桌面工具，支持自定义消息、预览消息、批量导入用户、多账号管理、定时推送等功能，目前已支持公众号模板消息、企业微信、微信客服、短信、邮件等消息类型。

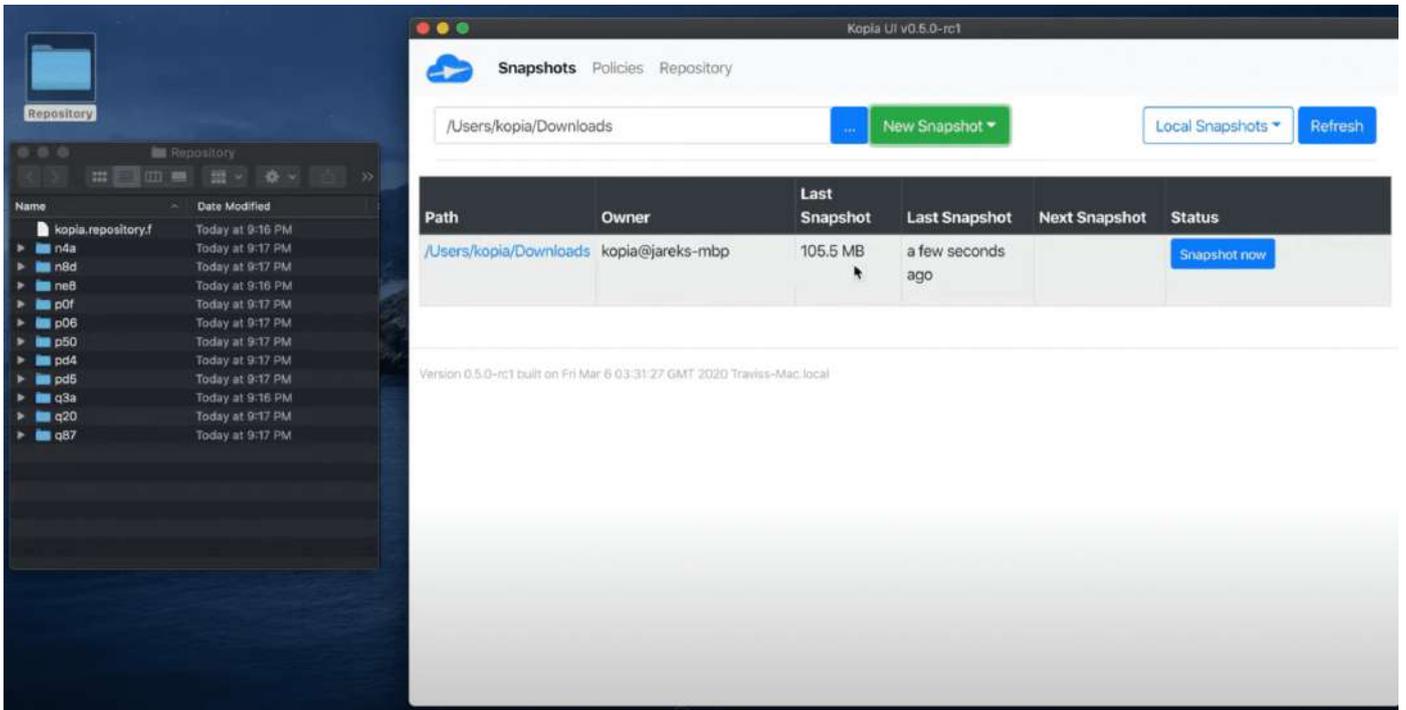


9、[backgroundRemover](#) (开源)

去除图像和视频的背景，它是命令行工具，方便批量处理和脚本编程。

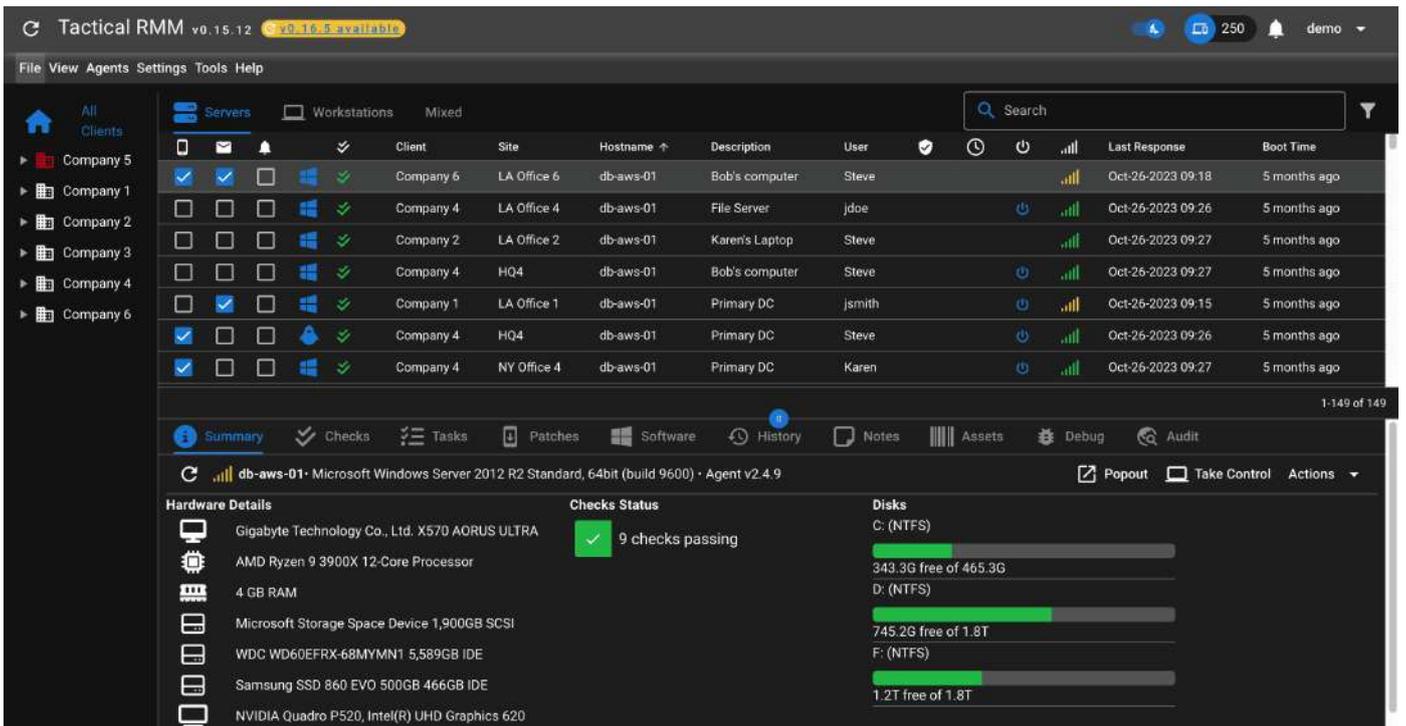
10、[kopia](#) (开源)

好用的跨平台备份工具。适用于 Windows、Linux 和 macOS 的备份/还原工具，它能够将指定的文件和目录加密压缩后生成快照，然后上传到远程服务器或云存储，实现远程备份数据的功能，开箱即用支持命令行和 GUI 两种使用方式。



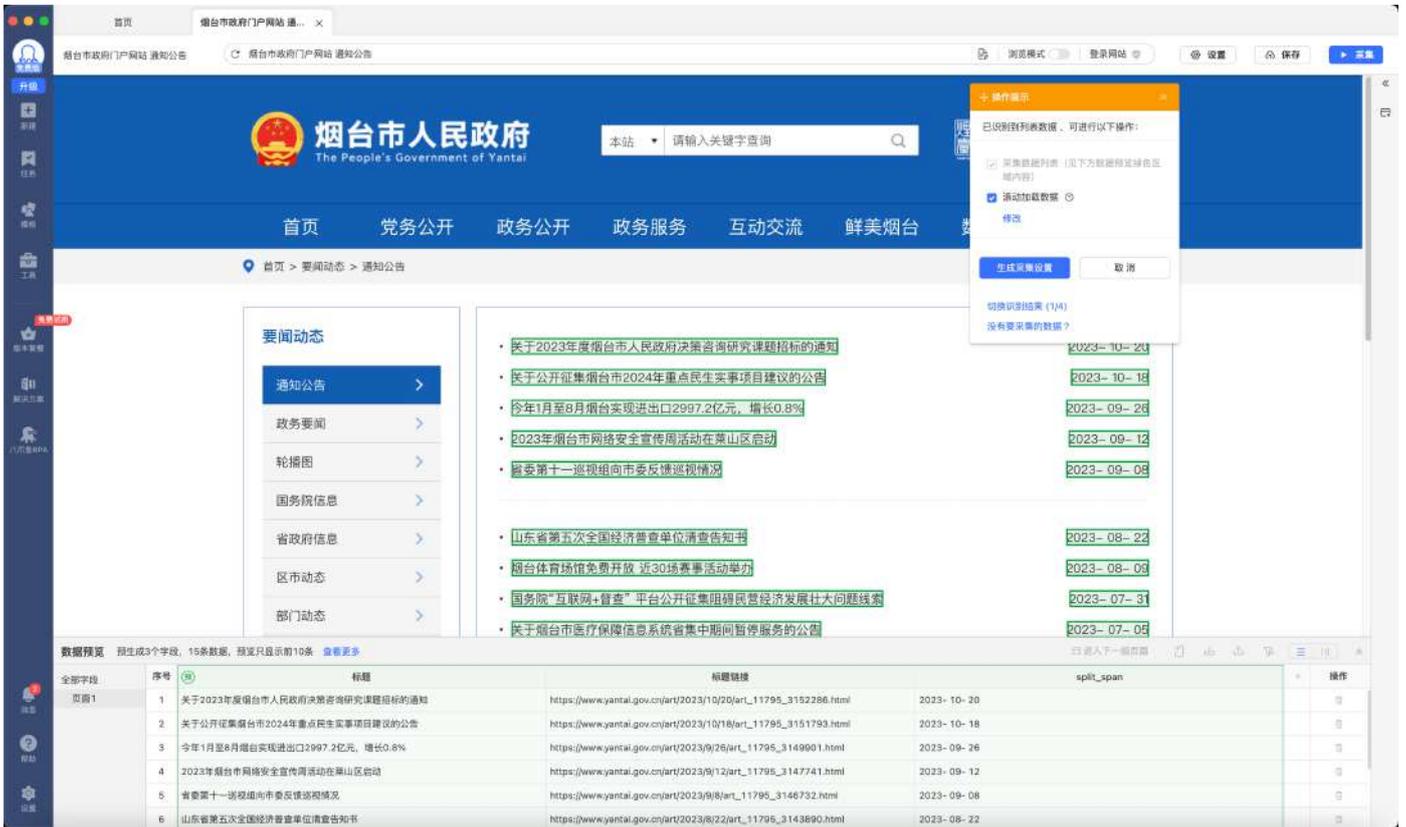
11、[tacticalrmm](#) (开源)

一个远程监控和管理 PC 的工具。该项目是采用 Django+Vue.js 构建的 IT 管理平台，支持类似 TeamViewer 的远程桌面控制、远程执行命令和脚本、上传和下载文件、警报等功能。



12、[八爪鱼采集器](#) (免费)

八爪鱼采集器是一款全网通用的互联网数据采集器（商业版软件，免费功能就够用了，付费版提供云采集解决方案），模拟人浏览网页的行为，通过简单的页面点选（提供智能识别功能），生成自动化的采集流程，从而将网页数据转化为结构化数据，存储于文件或数据库等多种形式。



13、pyvideotrans (开源)

一个视频翻译工具，可将一种语言的视频翻译为另一种语言和配音的视频。语音识别基于 openai-whisper 离线模型、文字翻译使用Google翻译，文字合成语音使用 Microsoft Edge tts，背景音乐去除使用 Spleeter。该工具只用了CPU，没有用到GPU，无需购买任何商业接口，也无需付费。

除此之外，还有个功能与之类似的浏览器插件 [YouTube Dubbing](#)，可以将 YouTube 视频的英文语音，转成中文语音。

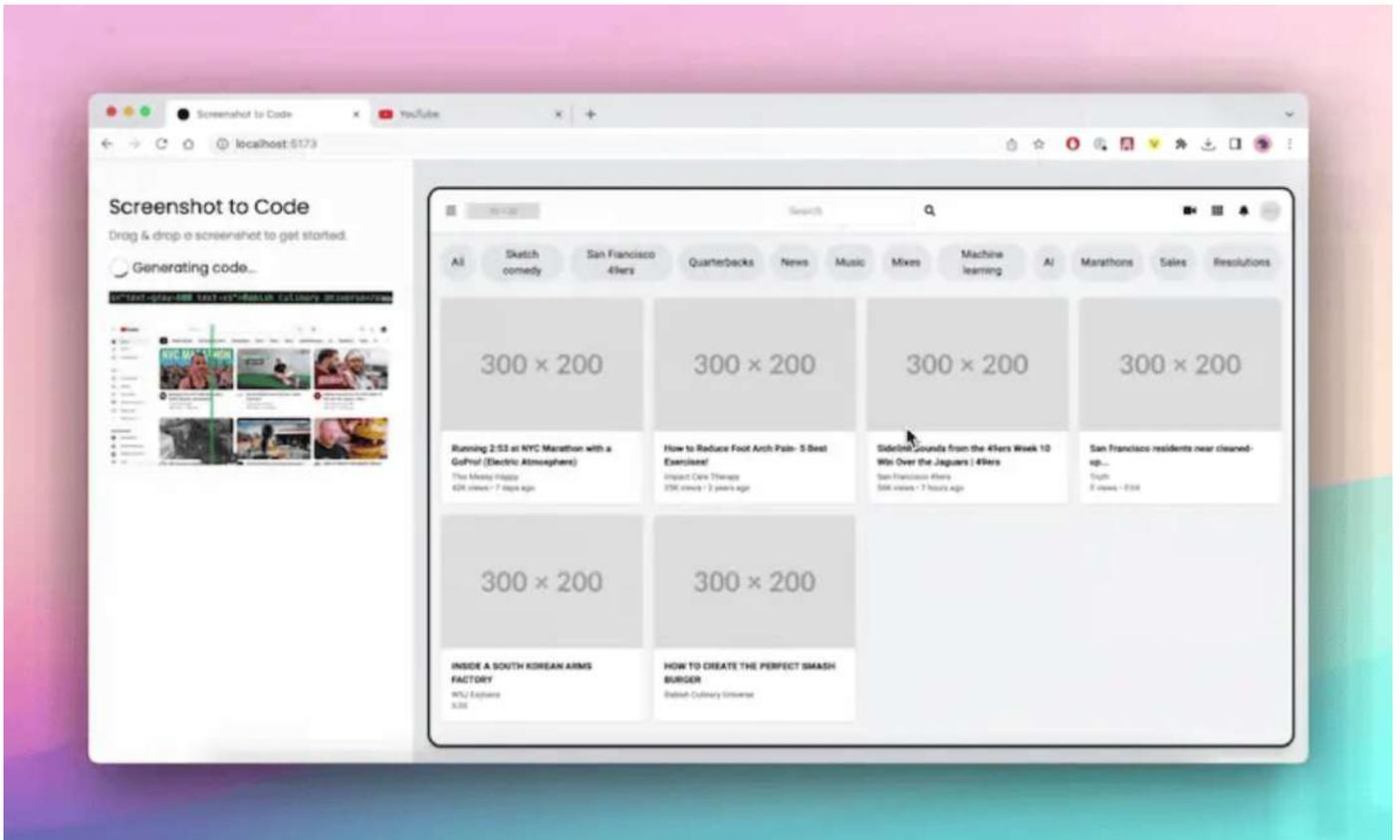
14、ChatGPT-Next-Web (开源)

一键拥有你自己的跨平台 ChatGPT 应用，支持 Web / PWA / Linux / Win / MacOS。



15、[screenshot-to-code](#) (开源)

一个开源的 Web 应用，用户上传一张网页截图，它会通过 OpenAI，给出该网页的 HTML/Tailwind/JS 代码实现。



16、[IPS](#) (开源)

一个命令行工具，查询和处理 IP 地理位置数据库。

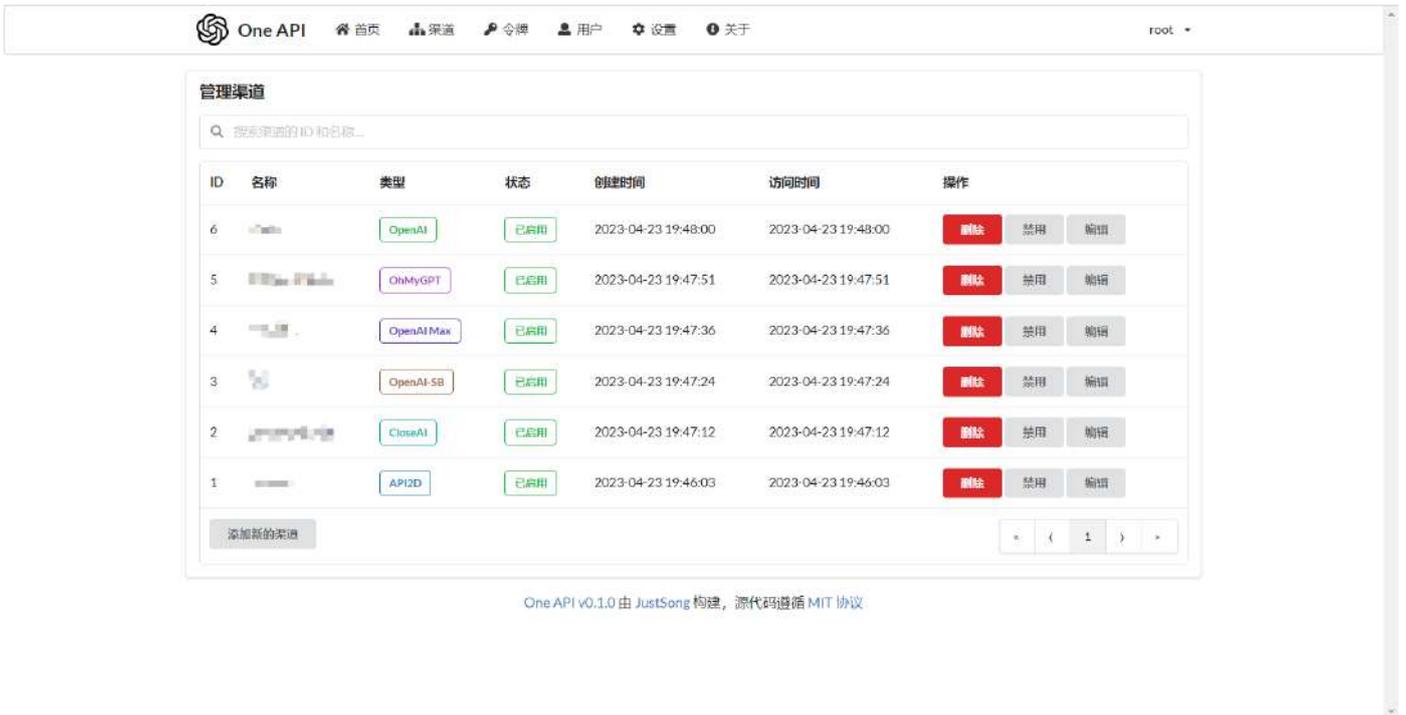
17、[AITDK](#) (免费)

用户输入文章的主题，该网页工具会自动生成 SEO 友好的标题、描述、关键词和常见问题解答。

The screenshot shows the AITDK.com website interface. At the top, there is a navigation bar with the following links: AITDK.com, 首页, AI特性生成器, AI常见问题解答生成器, AI SEO标题生成器, AI SEO描述生成器, AI SEO关键词, 简体中文, and a search icon. The main heading is "最佳AI标题、描述和关键词生成器" (Best AI Title, Description, and Keyword Generator), with a subtitle "为您的网站生成SEO友好的标题、描述和关键词" (Generate SEO-friendly titles, descriptions, and keywords for your website). Below the heading is a text input field containing "示例文本如: emoji wiki" and a purple "提交" (Submit) button. Underneath, there are six buttons arranged in a 2x3 grid: AITDK生成器, AI特性生成器, AI常见问题解答生成器, AI SEO标题生成器, AI SEO描述生成器, and AI SEO关键词生成器. The "关键特性" (Key Features) section lists four features: 1. AI驱动 (AI-driven): "我们的工具由人工智能驱动，确保为您的网站带来最佳效果。" (Our tool is AI-driven, ensuring the best results for your website.) 2. 易于使用 (Easy to use): "只需输入您的网站信息，我们的工具会做其余的工作。无需专业知识。" (Simply enter your website information, and our tool will do the rest. No professional knowledge required.) 3. 专为SEO优化 (Optimized for SEO): "我们的工具生成的标题、描述和关键词都是为搜索引擎排名而优化的。" (The titles, descriptions, and keywords generated by our tool are optimized for search engine rankings.) 4. 免费使用 (Free to use): "我们的AI标题、描述和关键词生成器完全免费使用。没有任何隐藏费用或费用。" (Our AI title, description, and keyword generator is completely free to use. No hidden fees or charges.)

18、[one-api](#) (开源)

OpenAI 接口管理、分发系统，支持诸多主流大模型，可用于二次分发管理 key，已打包好 Docker 镜像，可一键部署，开箱即用。

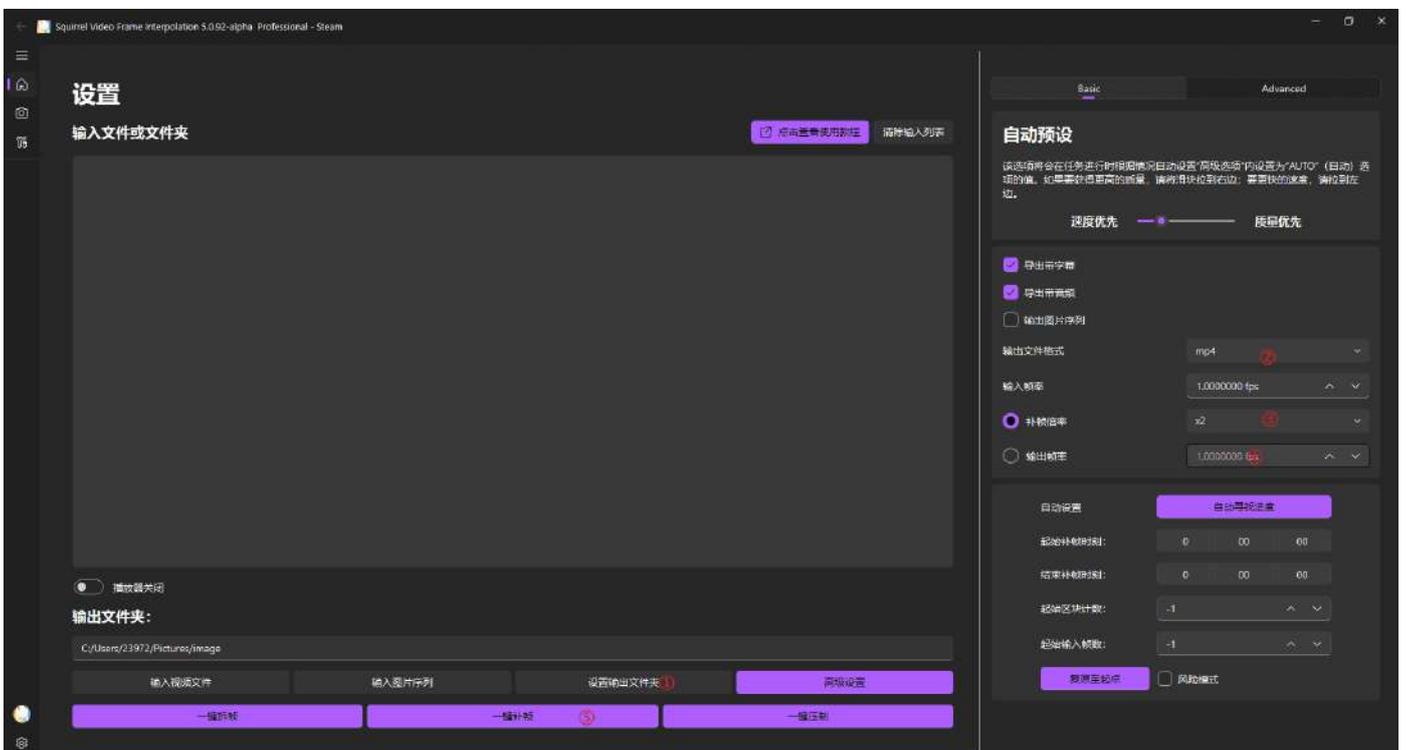


19、[N_m3u8DL-RE](#) (开源)

适用于 MPD/M3U8/ISM 的流媒体下载器。该项目可以将常见的流媒体保存到本地，支持点播、录制直播、自动混流等功能，适用于 Windows、Linux、macOS 操作系统。

20、[Squirrel-RIFE](#) (开源)

中文自动补帧工具。该项目是基于 RIFE 算法的补帧软件，可用于去除动漫卡顿感。具有无需手动设置、高质量输出、速度快等特点，适用于 Windows 10 及以上操作系统。

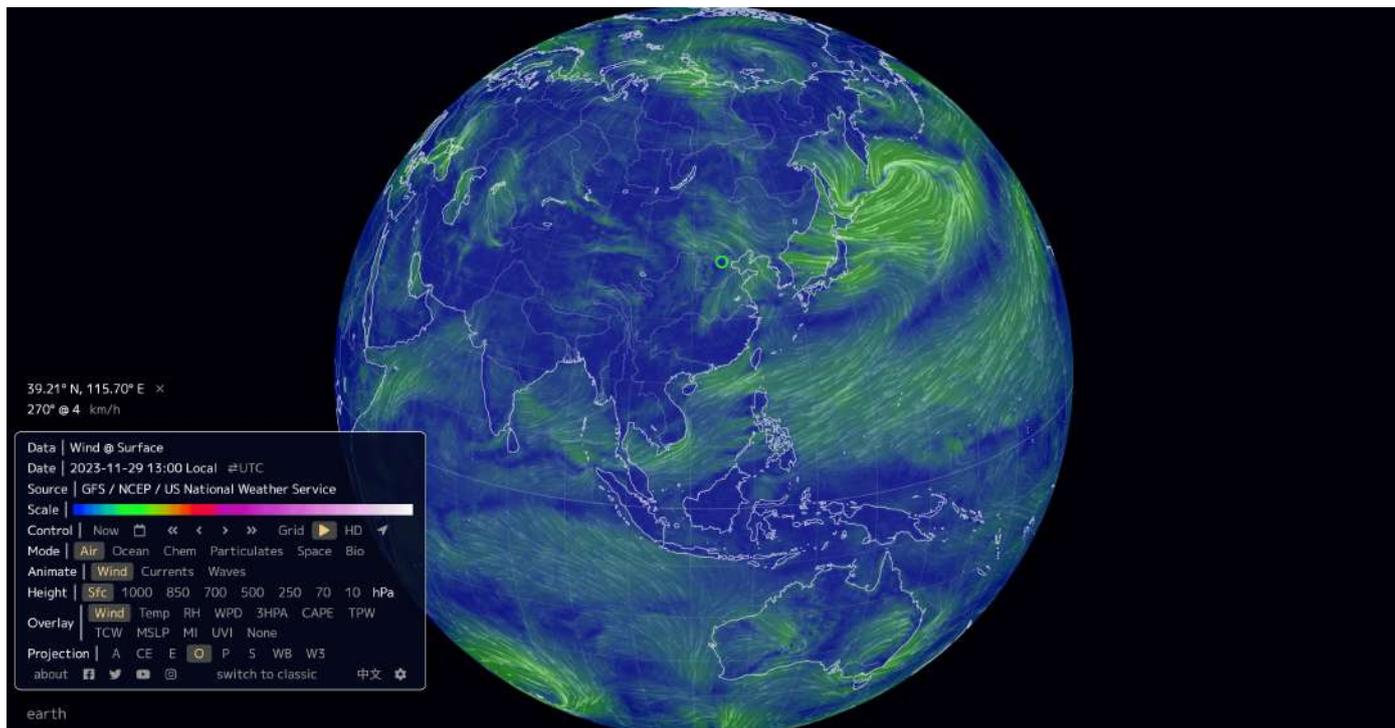


21、[migrate](#) (开源)

数据库迁移/变更工具。该项目是用 Go 写的数据库迁移(migrate)工具，帮你自动创建 SQL 迁移文件并管理版本，支持 MySQL、MariaDB、PostgreSQL、SQLite、Neo4j、ClickHouse 等不同类型的数据库。

22、[earth](#) (开源)

该项目以可视化的方式展示了全球的天气情况，提供了风、温度、相对湿度等多种天气数据，以及风、洋流和波浪的动画效果。[在线体验](#)

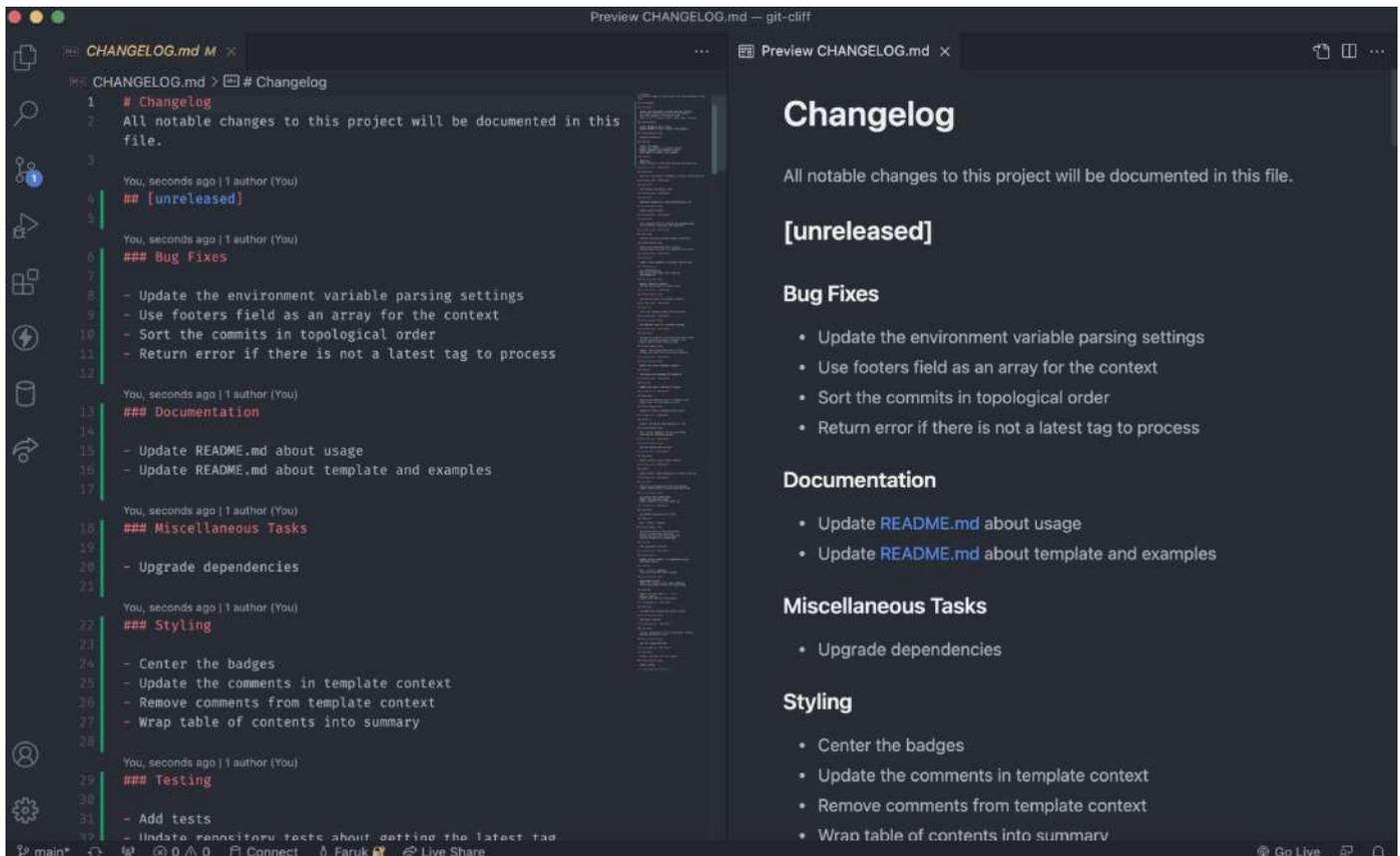


23、[LaTeX-OCR](#) (开源)

将数学公式转化成 LaTeX 代码。该项目可以将图片、剪贴板中的图片和屏幕截图，转化成对应的 LaTeX 代码，提供了命令行、库、GUI、Docker 多种使用方式。

24、[git-cliff](#) (开源)

自由可定制的变更日志生成器，该项目可以自定义解析规则，自动从 Git 历史记录中生成 Changelog 文件。



25、[linux-router](#) (开源)

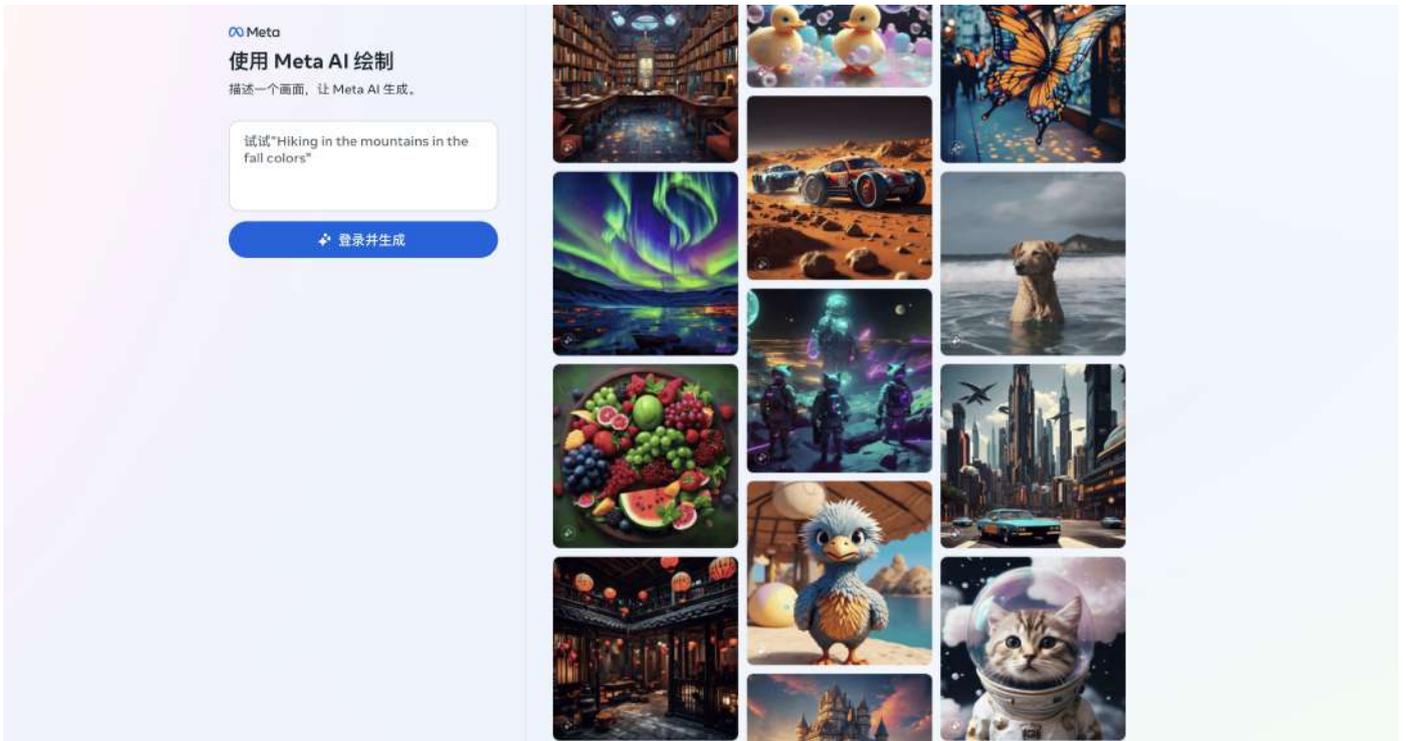
这是一个 Linux 软路由器的 Shell 脚本，它可以通过一条命令将 Linux 设备作为路由器，提供互联网共享、DNS 服务器、WiFi 热点等功能。

26、[dvc](#) (开源)

一款针对 AI 项目的数据版本管理工具。基于 Git 的数据版本管理工具，版本化机器学习项目的数据和模型。可用于比较代码、数据、参数、模型或性能图，共享机器学习项目的数据或重现结果。

27、[Imagine](#) (免费)

Meta 公司推出的文生图工具，使用 Facebook 和 Instagram 的11亿张图片训练，现在可以免费使用。



28、[WeChatMsg](#) (开源)

提取微信聊天记录，将其导出成HTML、Word、CSV文档永久保存，对聊天记录进行分析生成年度聊天报告。

29、[roomGPT](#) (开源)

基于ChatGPT的应用，上传一张房间照片，它使用 AI 重新设计你的房间。

Generating dream rooms using AI for everyone.

Take a picture of a room and redesign it in over 10 different themes.
Remodel your room today!

[Redesign your room for free](#)

Original Room



Generated Room



30、[CodeGeeX2](#) (开源)

AI 编程助手，支持 VS Code、Intellij IDEA、PyCharm、GoLand、WebStorm 等 IDE 编辑器。

31、[snowmachine](#) (开源)

一个 Python 脚本，可以在终端窗口显示下雪和圣诞树。



32、[Photo to Anime](#) (免费)

这个网站可以将上传的图片动漫化，也可以通过文本生成动漫图片。

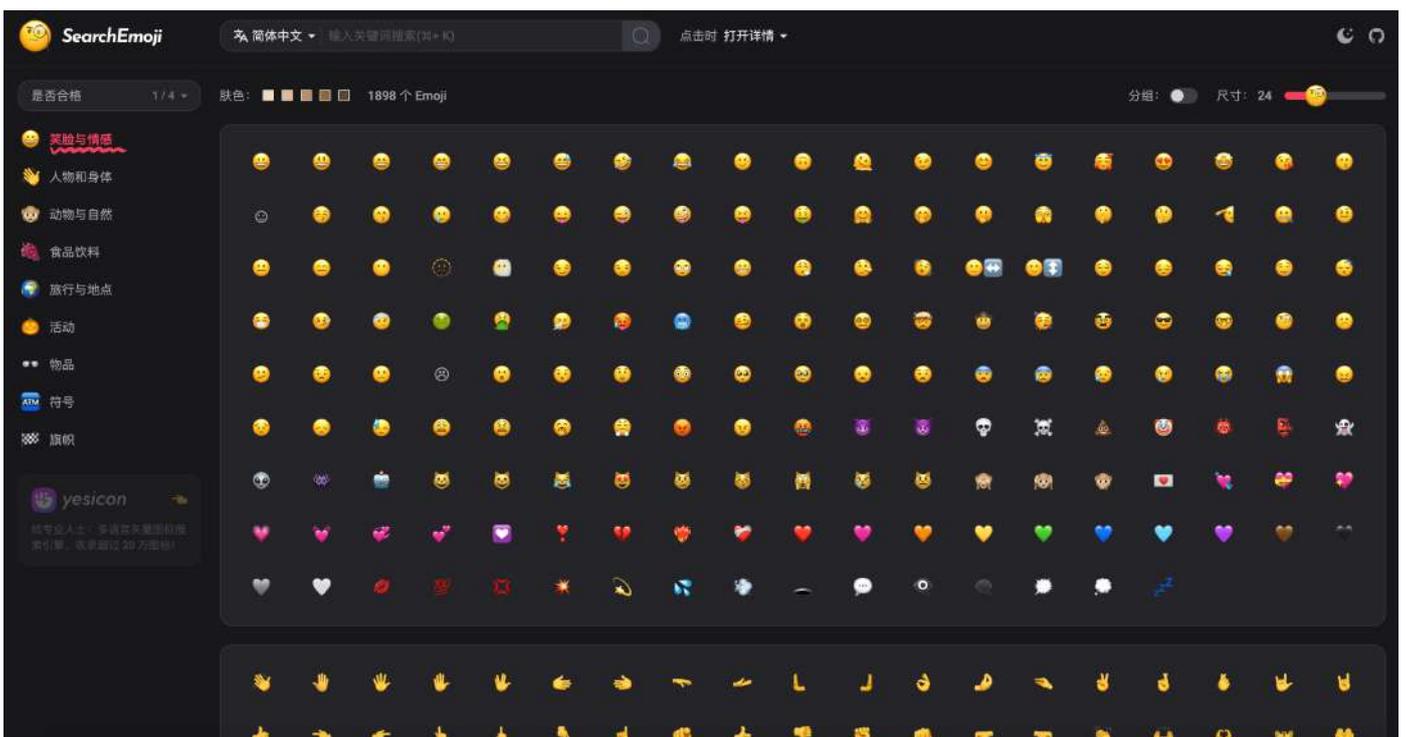


33、[EasyTranslator](#) (开源)

一个命令行的文件翻译工具，可以翻译.txt、.pdf、.docx、.md、.mobi、.epub 文件，需要 OpenAI API Key。

34、[SearchEmoji](#) (开源)

Emoji 搜索引擎，支持30种语言。[在线访问](#)

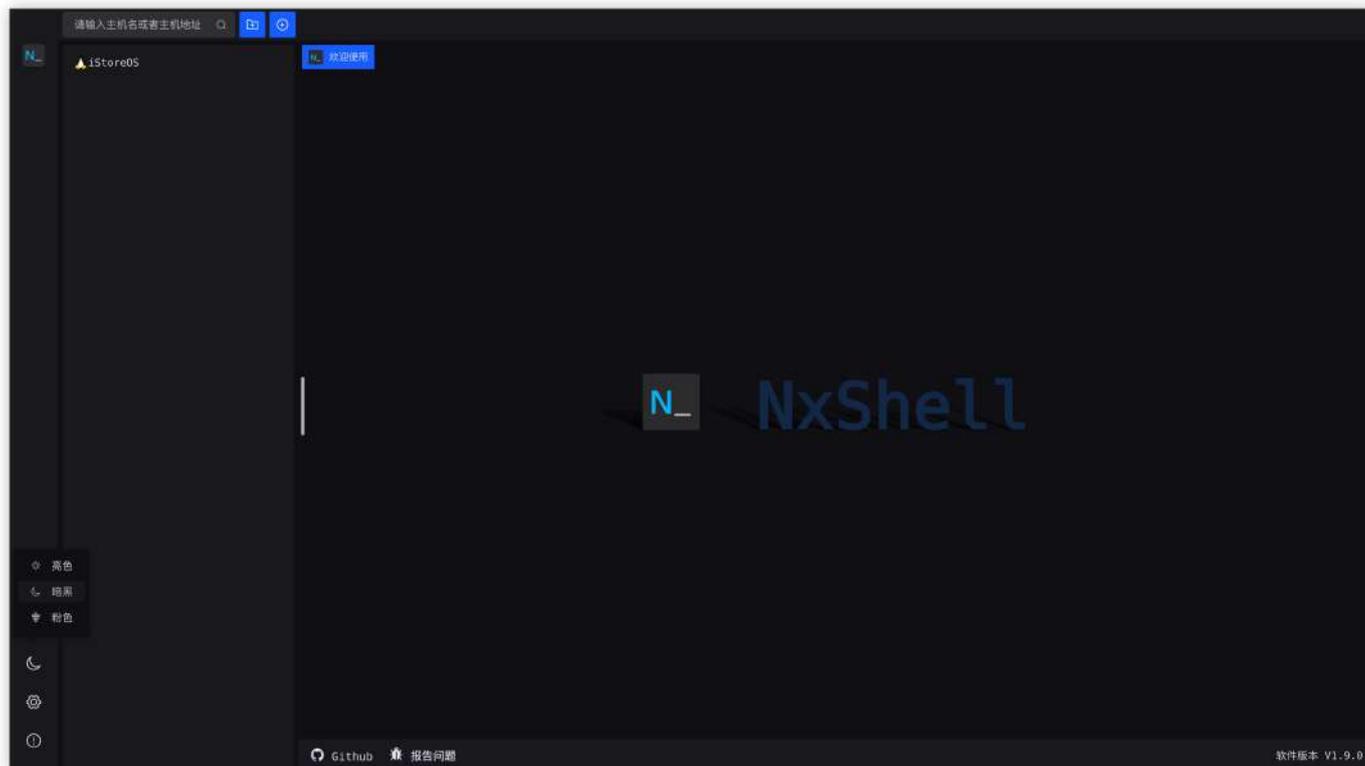


35、[Administrative-divisions-of-China](#)

中华人民共和国行政区划：省级（省份）、地级（城市）、县级（区县）、乡级（乡镇街道）、村级（村委会居委会），中国省市镇村二级三级四级五级联动地址数据。

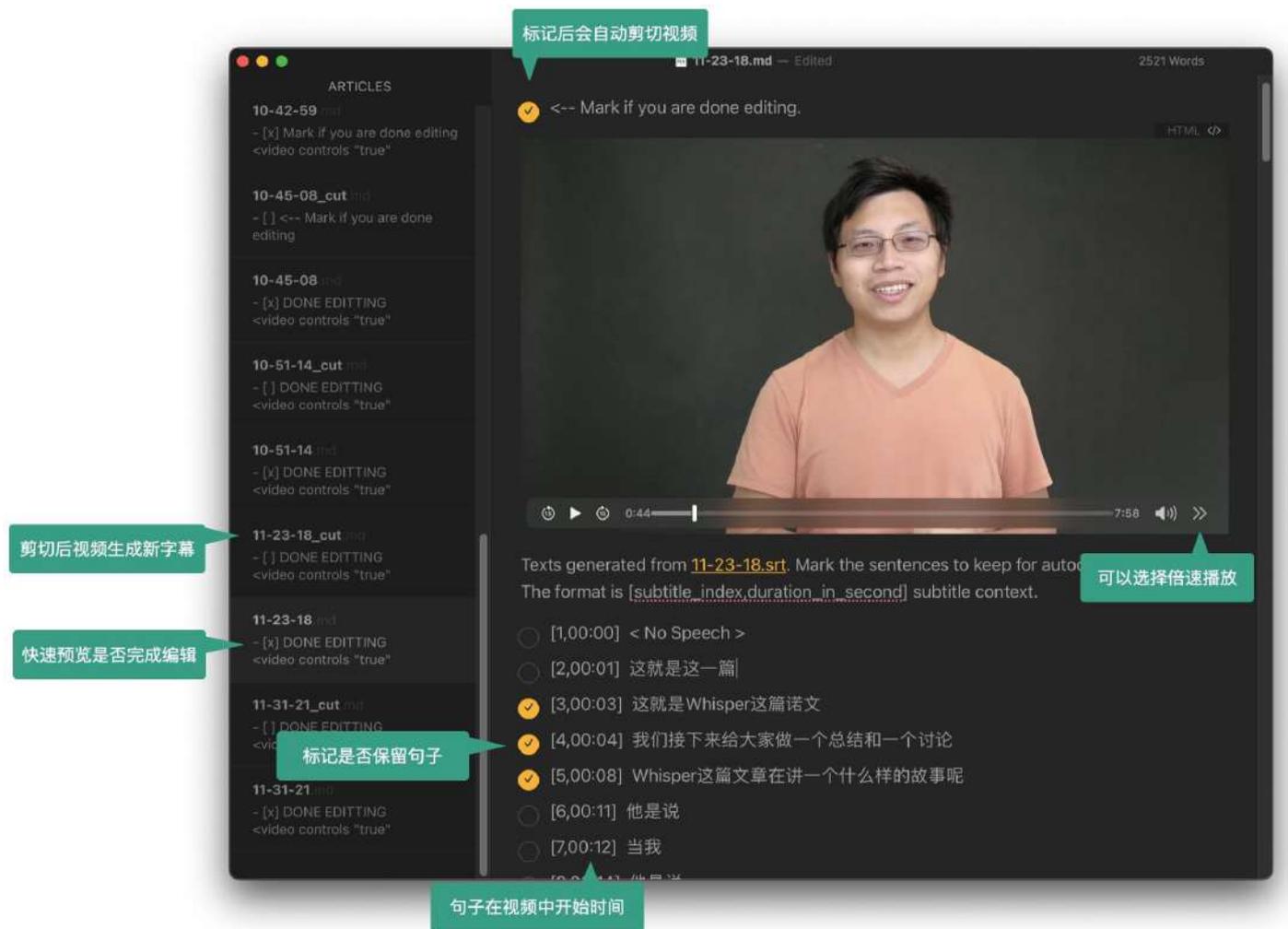
36、[nxshell](#)（开源）

一款跨平台的 SSH 新终端工具，就像当前流行的 PuTTY 终端一样。但 NxShell 随着升级功能的迭代完善，会发展出自己突出的终端工具超能力。



37、[autocut](#)（开源）

AutoCut 对你的视频自动生成字幕。然后你选择需要保留的句子，AutoCut 将对你视频中对应的片段裁切并保存。你无需使用视频编辑软件，只需要编辑文本文件即可完成剪辑。



38、[gpt-crawler](#) (开源)

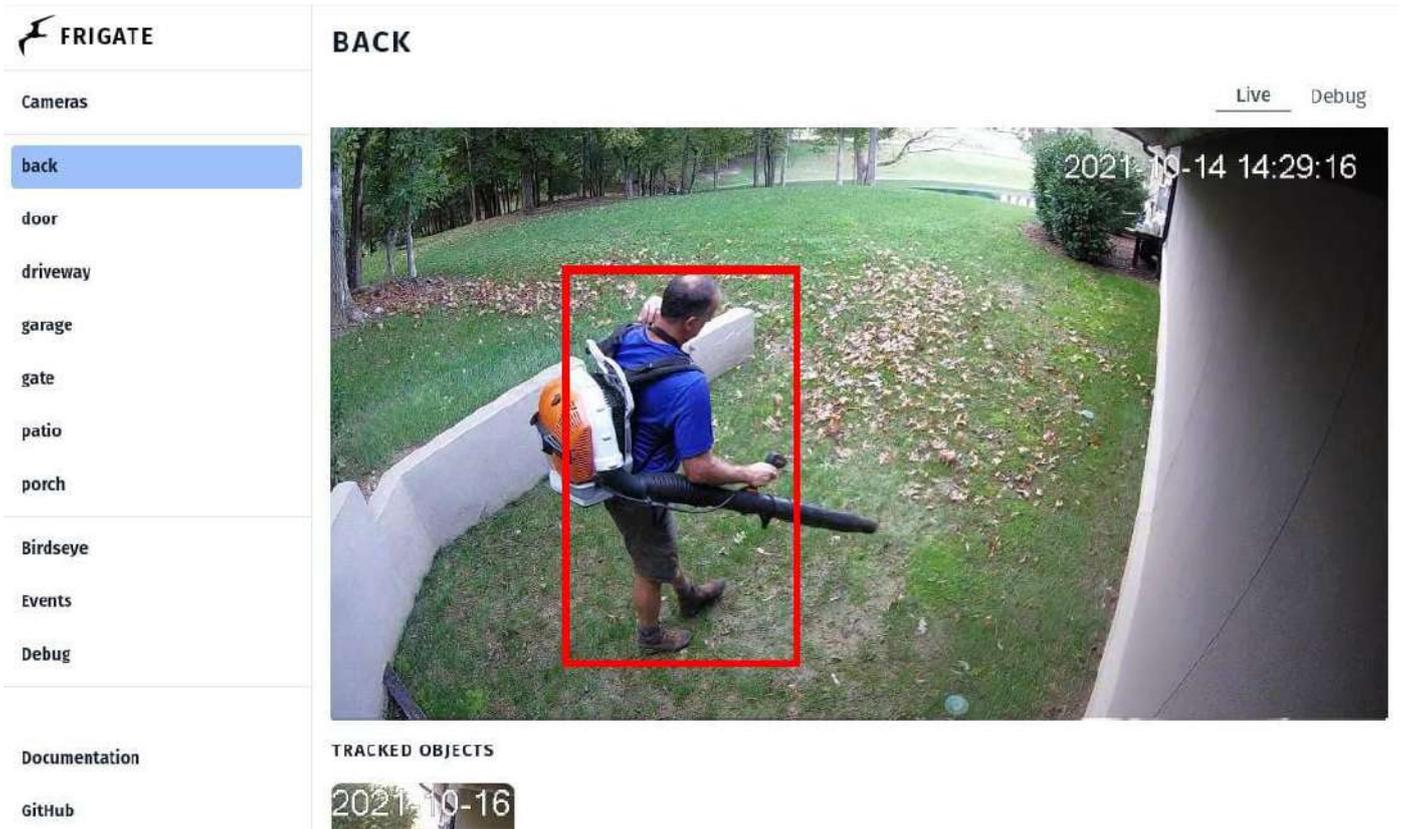
抓取网站以生成知识文件，从而从一个或多个 URL 创建您自己的自定义 GPT。

39、[muffet](#) (开源)

Go 语言开发的网站链接检查工具。该项目通过多线程和递归的方式，检查目标网站中所有页面的链接。

40、[frigate](#) (开源)

一款围绕实时 AI 对象检测构建的开源 NVR。该项目基于 OpenCV 和 Tensorflow 实现在本地为 IP Camera 提供实时目标检测和安全报警等功能，还支持根据设定的规则，保留检测到物体的视频。



41、[vue-naive-admin](#) (开源)

一款极简风格的 Vue 管理后台。这是一个开源、免费、可商用的后台管理模板，基于 Vue3、Vite4、Pinia、Unocss 和 Naive UI 等前端最新技术栈。它简洁、轻量、风格清新，上手成本低，适合中小型项目或者个人项目。

42、[Weylus](#) (开源)

将平板用作计算机上的输入板/触摸屏。通过该项目可以将平板设备，作为电脑的外接屏幕、触控板、键盘，适用于 Windows、Linux 和 macOS 系统。

43、[wondershaper](#) (开源)

Linux 限制网络带宽的工具。该项目通过使用 iproute 的 tc 命令，实现了对 Linux 网络带宽限速的功能，支持多网卡、限制最大上传/下载速率。

六、学习资源

1、[LLM 技术图谱](#) (中文)

开源中国发布了 LLM 技术图谱，从基础设施、大模型、Agent、AI 编程、工具和平台，以及算力几个方面，为开发者整理了当前 LLM 中最为热门和硬核的技术领域以及相关的软件产品和开源项目。

The image is a promotional banner for 'LLM Tech Map 大模型技术图谱'. It features a green background with a network diagram of nodes and lines. On the right side, there is a QR code. Below the main title, there are three category cards: '基础设施' (Infrastructure) with sub-categories '向量数据库' (Vector Database) and '数据库向量支持' (Database Vector Support); '大模型框架、微调 (Fine Tuning)' (Large Model Frameworks, Fine Tuning). The cards contain logos for various tools and services: AquilaDB, Annoy, Valid, Milvus, supabase, pgvector, LMFLOW, LoRA, and finetuner.

大模型训练平台与工具

AI Agent (LLM Agent)

AI 编程

大模型

备案上线的中国大模型	知名大模型	知名大模型应用

工具和平台

LLMOps	大模型聚合平台	开发工具

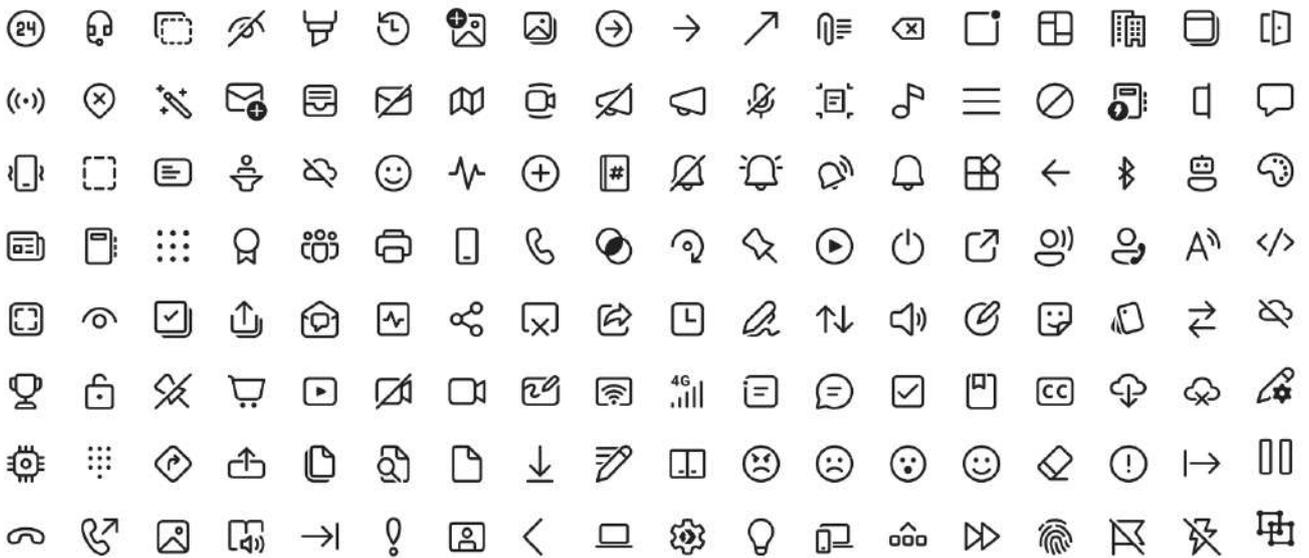
算力

2、[开源大语言模型完整列表](#) (中文)

由开源中国发布，本文对国内外公司、科研机构等组织开源的 LLM 进行了全面的整理。

3、[fluentui-system-icons](#) (英文)

这是一套微软 Fluent UI 设计风格的图标，它丰富、精致、统一。



4、[latex-cookbook](#) (中文)

《LaTeX 论文写作教程》，该书是教你如何用 LaTeX 写论文，内容涵盖了 LaTeX 基础、文本编辑、公式、图表设计、文献引用等方面。

5、[DeepLearningSystem](#) (中文)

AI 系统 & 深度学习教程，内容涵盖 AI 基础知识和概述、AI 芯片、AI 编译器原理、AI 框架核心技术等方面。

AI系统全栈架构图



6、[theByteBook](#) (中文)

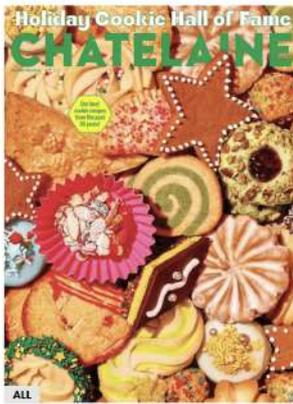
中文原创电子书籍 [深入架构原理与实践](#)，介绍互联网服务架构，集中在网络、集群以及服务治理、FinOps 这三个主题。

7、[magazinelib](#) (英文)

一个免费的外刊杂志电子版资源下载网站，它的宗旨是为全球的杂志爱好者提供一个方便、快捷、高质量的杂志阅读平台。每天更新数百种最新的杂志电子版，让用户能够第一时间获取最新的资讯和咨询。支持多种格式的杂志下载，包括PDF、EPUB、MOBI等，用户可以根据自己的设备和喜好选择合适的格式。



LATEST MAGAZINES



Chatelaine – Winter 2023-2024
07.11.2023, 08:20



Road & Track – Volume 20
Electrified – November 2023
07.11.2023, 08:00



MAGAZINES BY MONTH

▶ November 2023	440
▶ October 2023	2486
▶ September 2023	2224
▶ August 2023	3243
▶ July 2023	2004
▶ June 2023	2217
▶ May 2023	2084
▶ April 2023	1878
▶ March 2023	2206
▶ February 2023	1840
▶ January 2023	2091
▶ December 2022	2590
▶ November 2022	2208
▶ October 2022	2602
▶ September 2022	2250
▶ August 2022	2681
▶ July 2022	2028
▶ June 2022	2243
▶ May 2022	2448
▶ April 2022	2435
▶ March 2022	2304
▶ February 2022	2112
▶ January 2022	2225
▶ December 2021	2228
▶ November 2021	2324
▶ October 2021	2483
▶ September 2021	2190
▶ August 2021	1624
▶ July 2021	1483
▶ June 2021	1819
▶ May 2021	2050
▶ April 2021	2058
▶ March 2021	1772

8、[awesome LLMs interview notes](#) (中文)

该仓库主要记录大模型 (LLMs) 算法工程师相关的面试题和参考答案。

9、[GPTs在线目录合集](#) (英文)

基于 GPT 的工具越来越多了，下面的网站将这些工具整理在一起，方便查找。[GPTsHunter](#)、[AwesomeGPTs](#)、[BestGPTs](#)、[GPTs Works](#)、[GPTs Today](#)

10、[GPTs](#) (英文)

该仓库收集了 GPTs 的泄露 prompts。

11、[谷歌编程风格指南](#) (英文)

谷歌内部对各种语言的编程风格的要求。

Google Style Guides

Every major open-source project has its own style guide: a set of conventions (sometimes arbitrary) about how to write code for that project. It is much easier to understand a large codebase when all the code in it is in a consistent style.

"Style" covers a lot of ground, from "use camelCase for variable names" to "never use global variables" to "never use exceptions." This project ([google/styleguide](https://google.com/styleguide)) links to the style guidelines we use for Google code. If you are modifying a project that originated at Google, you may be pointed to this page to see the style guides that apply to that project.

- [AngularJS Style Guide](#)
- [Common Lisp Style Guide](#)
- [C++ Style Guide](#)
- [C# Style Guide](#)
- [Go Style Guide](#)
- [HTML/CSS Style Guide](#)
- [JavaScript Style Guide](#)
- [Java Style Guide](#)
- [Objective-C Style Guide](#)
- [Python Style Guide](#)
- [R Style Guide](#)
- [Shell Style Guide](#)
- [Swift Style Guide](#)
- [TypeScript Style Guide](#)
- [Vim script Style Guide](#)

This project also contains [cplint](#), a tool to assist with style guide compliance, and [google-c-style.el](#), an Emacs settings file for Google style.

If your project requires that you create a new XML document format, the [XML Document Format Style Guide](#) may be helpful. In addition to actual style rules, it also contains advice on designing your own vs. adapting an existing format, on XML instance document formatting, and on elements vs. attributes.

The style guides in this project are licensed under the CC-BY 3.0 License, which encourages you to share these documents. See <https://creativecommons.org/licenses/by/3.0/> for more details.

12、[nerd-fonts](#) (英文)

这是一个收集了 3600+ 图标的字体集合和补丁工具，该项目不是一个字体，而是一个可以将多种字体中的图标，作为补丁添加到目标字体中的工具。

13、[ml-engineering](#) (英文)

该项目是作者训练开源 BLOOM-176B 大模型和 IDEFICS-80B 多模态模型的经验总结，还提供了大量可以直接拿来用的代码和脚本。

14、[LangChain中文网](#) (中文)

500页超详细的LangChain中文文档教程，助力LLM/chatGPT应用开发。

15、[现代统计学导论](#) (英文)

一本面向大学新生的统计学入门教材，免费阅读。

Welcome to IMS2

Authors

Preface

Introduction to data

- 1 Hello data
- 2 Study design
- 3 Applications: Data

Exploratory data analysis

- 4 Exploring categorical data
- 5 Exploring numerical data
- 6 Applications: Explore

Regression modeling

- 7 Linear regression with a single predictor
- 8 Linear regression with multiple predictors
- 9 Logistic regression
- 10 Applications: Model

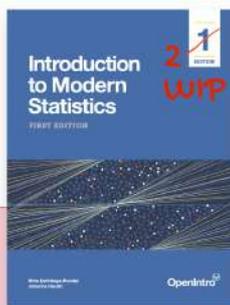
Foundations of inference

- 11 Hypothesis testing with randomization
- 12 Confidence intervals with bootstrapping
- 13 Inference with mathematical models
- 14 Decision Errors
- 15 Applications: Foundations

Introduction to Modern Statistics (2nd Ed)

Welcome to IMS2

This is the website for **Introduction to Modern Statistics, Second Edition** by Mine Çetinkaya-Rundel and Johanna Hardin. Introduction to Modern Statistics, which we'll refer to as IMS going forward, is a textbook from the [OpenIntro](#) project.



IMS2 is currently under construction, it's planned to be released in Spring 2024. The first edition of the book is available at [openintro-ims.netlify.app](#) with additional resources at [openintro.org/book/ims](#). If you have any feedback on the work-in-progress second edition or catch any typos, we would appreciate you reporting at [github.com/OpenIntroStat/ims](#).

Copyright © 2023.

Second Edition.

Version date: November 12, 2023.

This textbook and its supplements, including slides, labs, and interactive tutorials, may be downloaded for free at [openintro.org/book/ims](#).

This textbook is a derivative of *OpenIntro Statistics* 4th Edition and *Introduction to Statistics with Randomization and Simulation* 1st Edition by Diez, Barr, and Çetinkaya-Rundel, and it's available under a Creative Commons Attribution-ShareAlike 3.0 Unported United States License. License details are available at the Creative Commons website: [creativecommons.org](#).

Source files for this book can be found on GitHub at [github.com/OpenIntroStat/ims](#).

Table of contents

Welcome to IMS2

Edit this page
Report an issue

16、科技爱好者周刊的语义检索 (中文)

将科技爱好者周刊迄今的所有内容进行了向量化索引，提交给 ChatGPT，然后就可以实现语义检索了。

17、计算机科学热门论文 (英文)

该网站按照动态排名，列出计算机领域的最新热门论文。

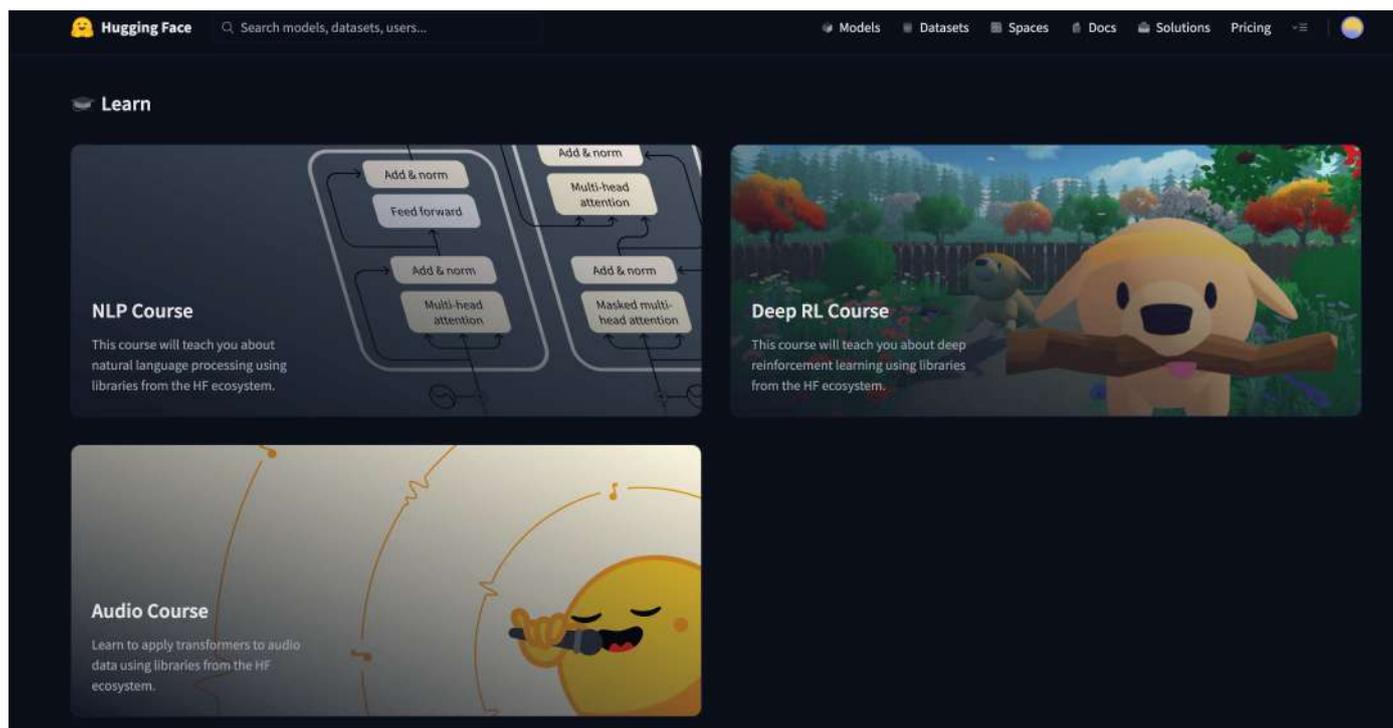
PageRank	Growth	Paper	Citations	in 3 mos.	PageRank evolution
678	+4,254%	GPT-4 Technical Report OpenAI Achiam, Josh Adler, Steven Agarwal, Sandhini Ahmad, Lama Akkaya, Ige Aleman, Florencia Leon Almeida, Diogo ... Mar 15, 2023 – The GPT-4 model is a large-scale, multimodal model that can process both image and text inputs and generate text outputs. It demonstrates human-level performance on professional and academic benchmarks, including...	3,585	+2,175	
1,104	+3,838%	LLaMA: Open and Efficient Foundation Language Models Touvron, Hugo Lavril, Thibaut Izacard, Gautier Martinet, Xavier Lachaux, Marie-Anne Lacroix, Timothée Rodière, Baptiste Goyal, Naman ... Feb 27, 2023 – The researchers have developed LLaMA, a collection of foundation language models that are trained on publicly available datasets and outperform GPT-3 on most benchmarks. They have released all their models to the research...	2,739	+1,616	
2,226	+3,168%	Scaling Instruction-Finetuned Language Models Chung, Hyung Won Hou, Le Longpre, Shayne Zoph, Barret Tay, Yi Fedus, William Li, Yunxuan Wang, Xuezhi Dehghani, Mostafa ... Oct 20, 2022 – This paper explores the benefits of instruction finetuning on language models, specifically focusing on scaling the number of tasks, model size, and finetuning on chain-of-thought data. The results show that instruction...	1,282	+630	
2,605	+2,445%	Llama 2: Open Foundation and Fine-Tuned Chat Models Touvron, Hugo Martin, Louis Stone, Kevin Albert, Peter Almahairi, Anjad Babaei, Yasmine Bashlykov, Nikolay Batra, Soumya Bhargava, ... Jul 18, 2023 – The researchers have developed and released Llama 2, a collection of large language models optimized for dialogue use cases. These models outperform open-source chat models and are considered a suitable alternative to...	1,456	+1,318	
4,506	+1,888%	BLOOM: A 176B-Parameter Open-Access Multilingual Language Model Workshop, BigScience Scialo, Taven Le Fan, Angela Akki, Christopher Pawlick, Ellie Ilk, Suzana Hesselow, Daniel Castagné, Roman ... Nov 9, 2022 – BLOOM is a large language model with 176 billion parameters that has been made openly accessible to the public. It achieves competitive performance on various benchmarks and its models and code are publicly available for...	744	+341	
3,183	+1,863%	Sparks of Artificial General Intelligence: Early experiments with GPT-4 Bubeck, Sébastien Chandrasekaran, Varun Eldan, Ronen Gehrke, Johannes Horvitz, Eric Kamari, Ece Lee, Peter Lee, Yin Tat Li, Yuanzhi ... Mar 22, 2023 – The paper discusses the development of GPT-4, an advanced language model by OpenAI, which exhibits more general intelligence than previous models and can solve various tasks without specific instructions. The authors...	806	+424	

18、Kickass markdown (英文)

这个仓库收集各种 GitHub Markdown 技巧。

19、[HuggingFace官方课程](#)（英文）

著名 AI 平台 Hugging Face 的官方学习频道，目前有三门课程：NLP、深度学习和 AI 音频处理。



七、生活经验

1、[大教堂、山峰、月亮](#)

一位意大利摄影师，苦苦尝试了6年，终于在本月拍到了一张难以置信的照片：大教堂、山峰、月亮处于一条直线，并且地球反射的太阳光，通过长曝光，可以照亮月亮顶部。

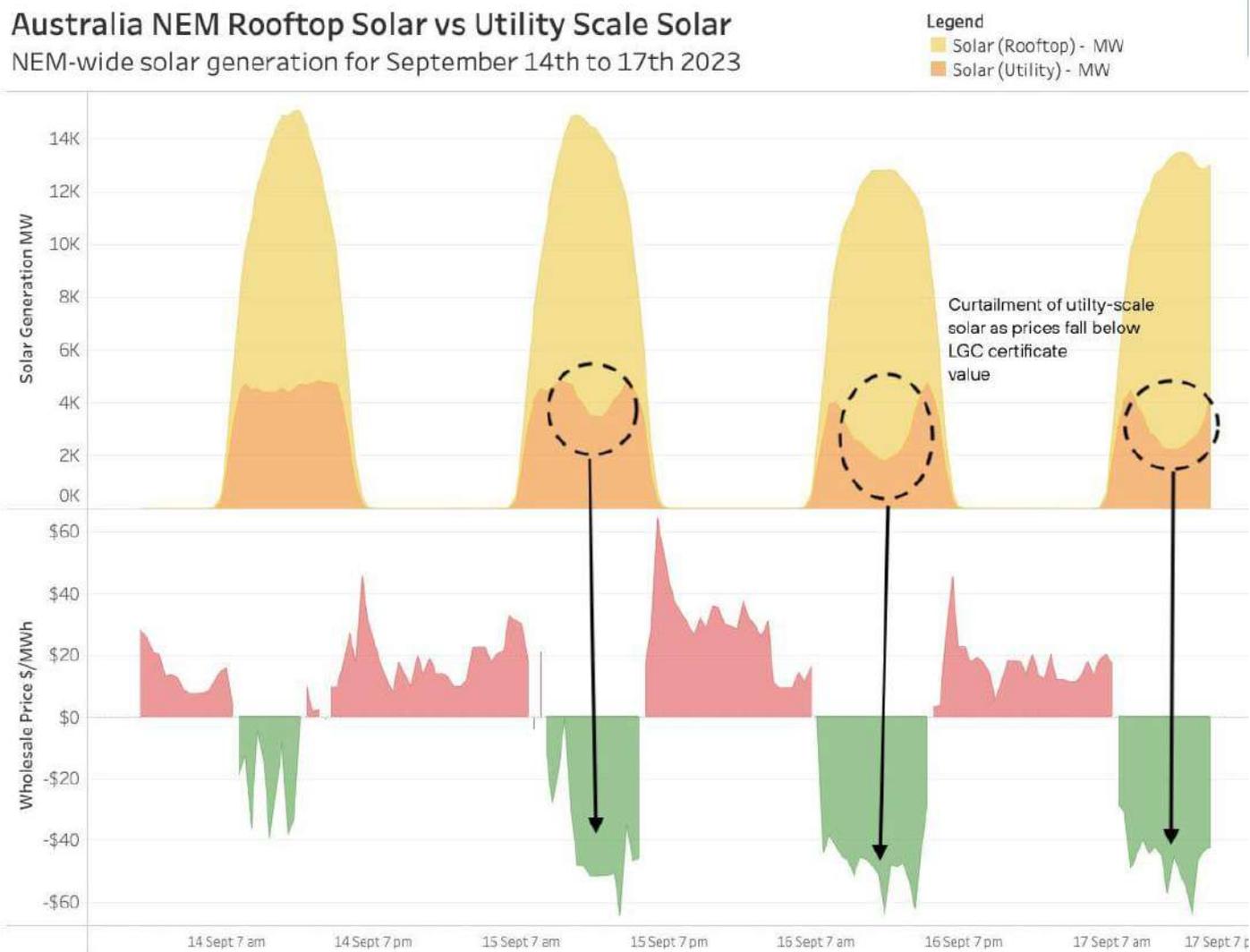


2、负电价

澳大利亚部分地区的批发电价出现了电价为负的情况，电网收购电厂发的电，电厂还要付钱给前者。主要原因是澳大利亚日照充分，非常适合太阳能发电。大部分居民住宅已经安装了太阳能板，导致电厂的电白天根本用不完。电厂为了维持机器运转，只好付钱让电网收购这些电。随着屋顶太阳能进一步普及，这种情况以后会越来越多。这就要求进一步发展储能设施，以及更动态灵活的电力调度系统。

Australia NEM Rooftop Solar vs Utility Scale Solar

NEM-wide solar generation for September 14th to 17th 2023



3、文档第一，会议第二

“开会文化”对公司是有害的，创业公司尤其要避免大量的会议。创业公司应该少开会，多写文档。文档第一，会议第二。

(1) 创业公司争分夺秒，会议非常浪费时间。当然，并非所有的会议都是不必要的，但是原则上，会议应该尽量不开，参加人员也应该尽量精简。

(2) 不断开会恰恰表明一个更深层次的问题：缺乏清晰、可访问和可靠的文档。

如果每个流程都有文档，就不需要一个小时的会议来澄清。如果每个决定都有文档，就不需要满屋子的人来理解它的理由。如果每个团队都有文档，就不需要在新成员加入团队时进行小组讨论。

(3) 会议创造了生产力的幻觉。你以为开会提高了生产力，实际上它们正在阻碍它。

每一次不必要的会议都是一种浪费。那些时间本可以用来改进算法，哪怕用来学习或者休息也很好啊。从本质上看，减少会议不仅仅可以节省时间，还可以让大家更专注、更多创新和创造，这恰恰是创业公司的命脉。

(4) 会议往往会自动膨胀。你召集了半小时的会议，快速讨论一个小问题。结果在会议中发现一个意想不到的问题非常重要，你们的争论就持续了两个小时。

(5) 会议不容易确定细节。你提到了一些别的事情，或者说话含糊不清，又或者双方沟通不畅，会议就会变得不那么清晰。解决方法到头来还是要写下来。

(6) 会议往往偏向声音最大的人，而不一定偏向那些有最好想法的人。这会扼杀创新和思想的多样性。作为对比，文档提供了公平的竞争环境，每个团队成员都可以表达他们的想法和见解，它促进了一种深思熟虑和反思的文化，而不是草率的判断和冲动的决定。

结论：你的时间和资源最好花在记录上，而不是花在会议上。大多数会议很容易被一份精心起草的、提供相关数据和建议、并征求反馈意见的文件所取代。

4、彼得·蒂尔的实验

美国著名的风险投资家彼得·蒂尔（Peter Thiel）认为，很多有才华的年轻企业家都被大学耽误了。他就在2010年做了一个实验，如果你愿意退学创业，就可以向他申请10万美元资助。

前不久，彭博社盘点了一下所有的奖学金得主，结果令人震惊。过去13年中，一共有271个人得到了奖学金，也就是说，他让271个学生从大学退学创业。这些创业公司之中，迄今为止，已经诞生了11家“独角兽”，即估值超过10亿美元，其中不乏明星公司。271个学生出了11家独角兽，这个比例真是高得吓人。如果这些学生没有退学，还能取得这样的成绩吗？

无论从哪个角度看，这个实验结果都证明了他的主张：对于那些有创业想法和能力的学生，大学的作用不大，早点去追求自己的梦想，可能是更好的选择。

除了拖延创业时机、分散精力，大学还可能消磨一个人的创业意愿。等到读完大学、拿到学位，大多数人的心态是，已经投入了多年的时间和金钱成本，不再愿意冒险，宁愿选择收入稳定、风险小的职业，所以金融、咨询、公务员才这么吃香。

	公司	创始人	估值（亿美元）
1	以太坊	Vitalik Buterin	2000
2	Luminar	Austin Russel	85
3	Upstart	Paul Gu	48
4	Figma	Dylan Field	200
5	Scale AI	Lucy Guo	75
6	Loom	Shahed Khan	10

彼得·蒂尔的实验，并不是否定大学教育，提倡退学创业，而是向社会表明，应该建立一种机制，识别有企业家潜质的学生，帮助他们追求自己的梦想，创造一些世界上还不存在的东西，而不是要求他们读完大学。

至于那些不具备企业家潜质，不知道自己想干什么的大多数学生，读完大学才是更好的选择。只是不要局限在大学为你安排的教育，而要找到自己想为社会创造的东西。

八、闲情逸趣

[Ilya Sutskever: 一个改变AI世界的天才科学家](#)

因为 OpenAI 眼花缭乱的政权争夺大戏，OpenAI 的首席科学家 Ilya Sutskever 的名字才逐渐被世人所知。Ilya Sutskever 在科学和工程实现上为 ChatGPT 的诞生做出了巨大贡献，可谓是 ChatGPT 的发明人。然而，也许是因为 Sam Altman 是负责搞融资和商业化的 CEO，也许是因为 Altman 的名字总是被国内媒体翻译为“奥特曼”，他的激进的作风更能引起国内社交媒体的关注，这使得 Altman 被冠上了“ChatGPT 之父”的称号。那么究竟谁才是真正的 ChatGPT 之父呢？知乎上有网友专门为此展开了争论。

这里暂且放下争论，其实少了 Ilya Sutskever 和 Sam Altman 任何一人，都没有现如今的 ChatGPT 和 OpenAI，但如果没有 Ilya Sutskever 在 AI 领域中的诸多突破与成就，一定不会有现在 ChatGPT 的出现。这里，我们先不谈“Altman”，而是着重回顾 Ilya Sutskever 的故事，看看他是如何凭借着他的聪明才华一步步从普通人成为 AI 界的一代传奇人物。



1986—2002年：早年生活和教育

Ilya 于 1986 年出生在苏联，5 岁时搬到耶路撒冷生活，自幼对计算机产生浓厚兴趣，从 7 岁时就开始自学编程。之后，在以色列开放大学学习了一段时间后，16 岁那年再次搬家去了加拿大，转而在多伦多大学读书，据说他在新家的第一件事就是去多伦多公共图书馆找一本关于机器学习的书。从那时起，Ilya 对人工智能的热情日益加深，他决定将构建通用人工智能（AGI）设定为自己的主要目标。

2003年：拜入深度学习之父Hinton门下

在学术生涯中，对 Ilya 影响最大的人莫过于他的老师：Geoffrey Hinton——被誉为深度学习之父的那个男人。2003年，还在就读本科的 Ilya 每天都会敲 Hinton 的门，希望能被接纳进实验室。在读了几篇论文后，他在深度学习领域的思考与 Hinton 教授的思想发生了碰撞，引起了 Hinton 教授对 Ilya 独特之处的认识：他有能力发现那些其他人可能需要数年才能发现的东西。尽管当时正值人工智能领域的寒冬，研究者们对人工智能的兴趣逐渐下降，

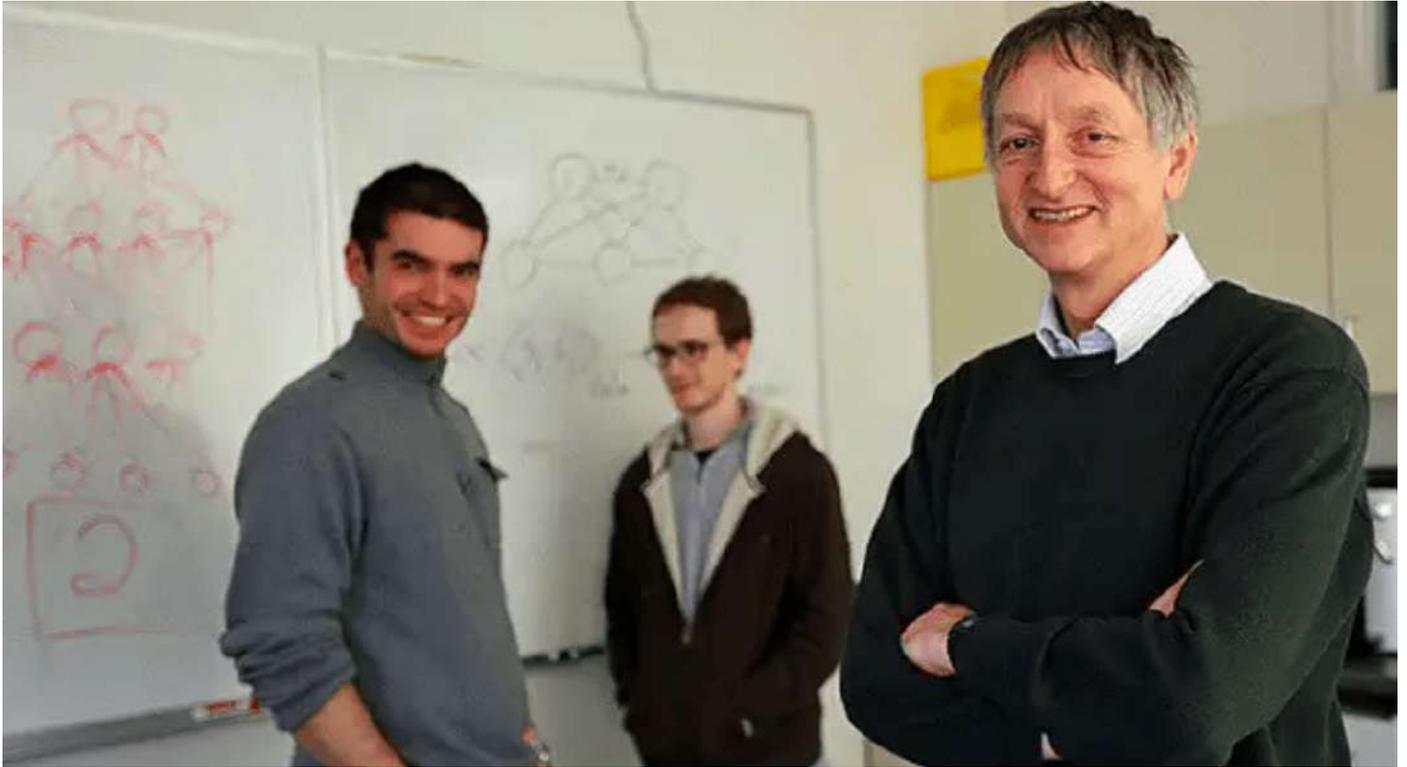
但 Hinton 教授还是决定接纳 Ilya 进入他的实验室。2005年，Ilya 在多伦多大学获得数学学士学位，2007年获得计算机科学硕士学位，2013年获得计算机科学博士学位。



2012年：图像识别领域的革命—— AlexNet

当时正值人工智能的低谷期，Hinton 的神经网络理念并没有被同期的 AI 研究者们广泛接受，这让他们面临了巨大的压力。然而他们并没有因此而退缩，反而更加坚定了他们的信念。2012年，ImageNet 大赛为他们提供了证明自己的机会——看谁能设计出最优秀的图像识别算法，成功识别出更多的图像。Hinton 带着他的两位学生 Ilya Sutskever 和 Alex Krizhevsky 在这次大赛中发明了 AlexNet，他们摒弃了传统的手工设计解决方案，而是使用了深度神经网络并在 GPU 上训练它们。

AlexNet 突破了75% 的准确度，远远超越了所有竞争对手，向大家展示了深度学习的巨大潜力，也标志着深度学习革命的开端。由他们三人共同撰写的 AlexNet 论文至今引用量已超过 6 万次，成为计算机科学领域引用最多的论文之一，而当时的 Ilya 才只有 26 岁。



2013年：从 DNNResearch 到加入谷歌

在这场竞赛后，三人共同创办了一家名为 DNNResearch 的公司，然而，该公司并没有推出任何产品，也没有明确的发展计划。随后，包括百度、谷歌、微软和 DeepMind 在内的四家公司参与了竞标收购。仅仅过去了4个月，谷歌在 2013 年 3 月以 4400 万美元的金额收购了 DNNResearch，并聘请 Ilya 担任 Google Brain 的研究科学家。

在 Google Brain 工作期间，Ilya 参与了许多后来惊艳世人的工作，比如与 Demis Hassabis 和 David Silver 等人共事，参与 AlphaGo 的核心工作。此外，他还积极参与了 TensorFlow 的开发，造福了众多深度学习研究人员。在这个阶段，Ilya 的愿景变得更加宏大——开始坚信通用人工智能（AGI）近在咫尺。

Article | [Published: 27 January 2016](#)

Mastering the game of Go with deep neural networks and tree search

[David Silver](#) , [Aja Huang](#), [Chris J. Maddison](#), [Arthur Guez](#), [Laurent Sifre](#), [George van den Driessche](#), [Julian Schrittwieser](#), [Ioannis Antonoglou](#), [Veda Panneershelvam](#), [Marc Lanctot](#), [Sander Dieleman](#), [Dominik Grewe](#), [John Nham](#), [Nal Kalchbrenner](#), [Ilya Sutskever](#), [Timothy Lillicrap](#), [Madeleine Leach](#), [Koray Kavukcuoglu](#), [Thore Graepel](#) & [Demis Hassabis](#) 

[Nature](#) **529**, 484–489 (2016) | [Cite this article](#)

463k Accesses | **7908** Citations | **3082** Altmetric | [Metrics](#)

2014年：机器翻译的革命——Sequence-to-Sequence

在谷歌工作期间，Ilya 发明了一种用于将英语翻译成法语的变体神经网络。他提出的序列到序列学习方法捕捉输入的时序结构（如英语句子）并将其映射到具有时序结构的输出（如法语句子）。

当时的研究者们并不相信神经网络能够进行翻译工作，而他的发明击败了表现最佳的翻译器，并促使谷歌翻译重大升级，让机器翻译领域从此焕然一新，也自此引起了自然语言领域的研究新热潮，这对后来 Transformer 的诞生也做出了重要贡献。

Sequence to Sequence Learning with Neural Networks

Ilya Sutskever
Google
ilyasu@google.com

Oriol Vinyals
Google
vinyals@google.com

Quoc V. Le
Google
qvl@google.com

2015年：联合创建OpenAI，开启新篇章

谷歌在 AI 领域一直处于领先地位，而 Ilya 作为备受重视的研究人员，有天收到了一封来自 Sam Altman 的邀约。他赴约去与 Sam Altman、Greg Brockman 和 Elon Musk 等人共同讨论人工智能的未来，畅想未来可能发生的事情，以及他们是否能够采取一些积极措施来影响人工智能的发展。

就在这场聚会中，OpenAI 的创始理念首次被明确提出——“利用人工智能造福全人类，并以负责任的方式推动人工智能的发展”，这激发了所有参与者的期待和激情。然而，当时 Ilya 还在谷歌工作，对于是否加入 OpenAI 的问题犹豫不决。最终，在多次思考和马斯克等人的劝说下，Ilya 决定放弃在谷歌数百万美元的工作机会，成为非营利组织 OpenAI 的联合创始人，并担任研究总监。



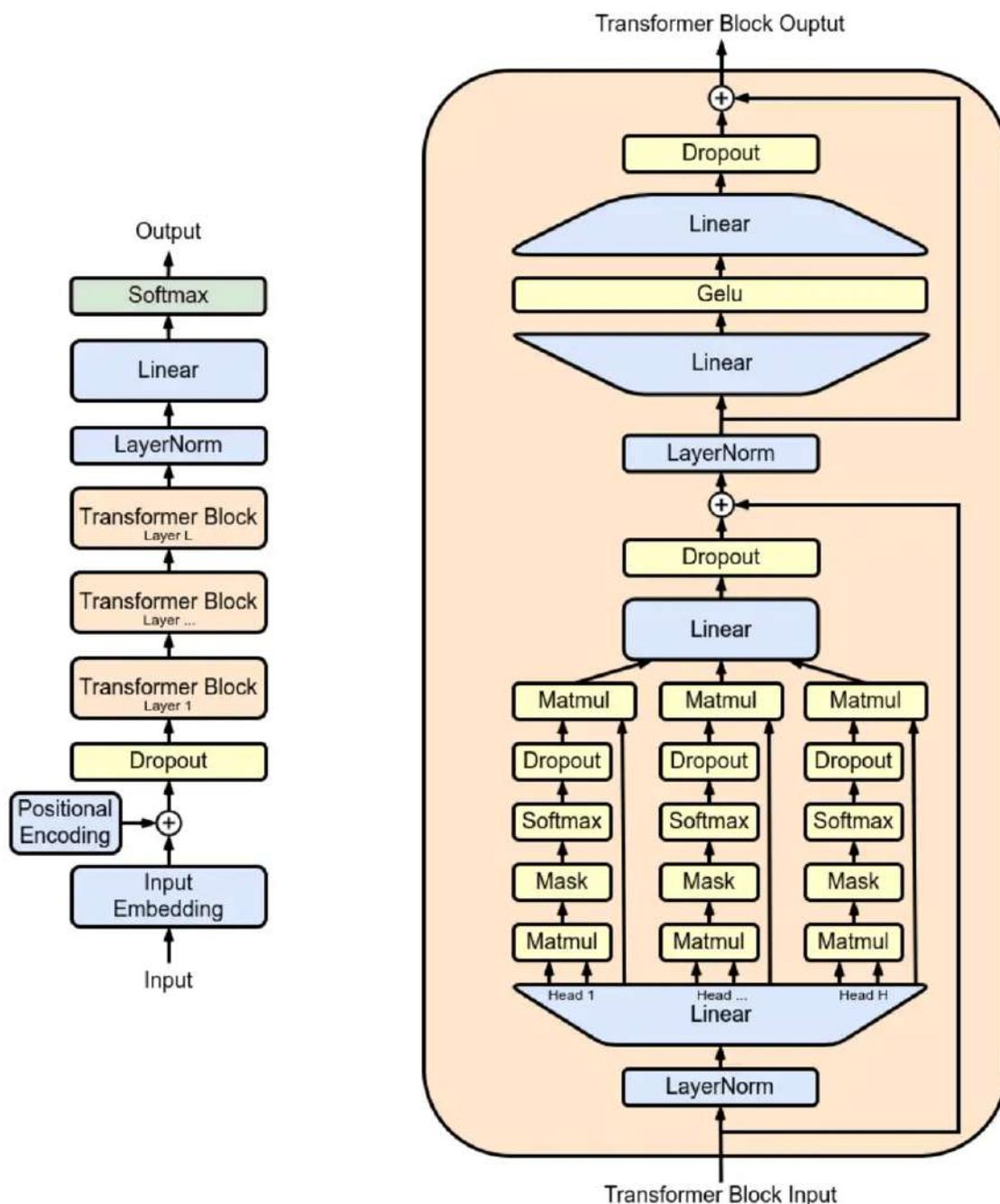
OpenAI 的早期发展并不顺利，作为非营利组织，寻找投资人成为一项艰巨的任务。然而，马斯克等人决定向该项目投入 10 亿美元，这使得 OpenAI 能够聘用领域内的优秀人才，并取得了一些重要突破。但是这些发展也付出了昂贵代价——每个月需要在云计算上花费数百万美元，而且还需要从其最大的竞争对手谷歌那里租用计算能力。

马斯克曾是 OpenAI 的董事会成员，在 2018 年因与特斯拉发生利益冲突而选择离开，这一度使得 OpenAI 面临严重的资金问题。在如此困境下，Sam 试图利用他的人脉寻找新投资者，但由于 OpenAI 是个非营利组织，需要的资金过多，几番尝试并未成功。他们必须迅速找到解决方案，否则所有努力都将付诸东流。在这个关键时刻，微软的 CEO 同意与他们会面，会议结果改变了 OpenAI 的窘境，Sam 成功说服了微软投资 10 亿美元，并获得了使用 Azure 云计算平台的权限。这样 OpenAI 就不再需要向谷歌支付费用，而微软转而成为他们的重要合作伙伴。

2018-2020：GPT-1~3 的发展

GPT 模型的每一次迭代都代表了自然语言处理领域的重大进展：

- GPT-1 (2018)：这是该系列的首个模型。其关键创新之一是采用了无监督的预训练方法，经过对大规模互联网文本数据集的训练，通过学习根据前面的单词上下文来预测句子中的单词，使模型能够深入理解语言结构并生成类似于人类的文本。
- GPT-2 (2019)：在 GPT-1 的成功基础上构建，采用更大的数据集训练，生成更为强大的模型。GPT-2 的一项重大进展是其能够在各种主题上生成连贯而流畅的段落文本，使其成为无监督语言理解和生成任务中的关键参与者。
- GPT-3 (2020)：在规模和性能上都取得了显著突破。GPT-3 拥有当时震惊世人的 1750 亿参数，在众多语言任务上取得了最先进的性能，在问答、机器翻译和摘要生成等能力上可以媲美人类水平。此外，它还展示了执行简单编码任务、撰写连贯新闻文章甚至是诗歌的能力。



在微软的支持和 Ilya 等人的领导下，OpenAI 的研究人员在 2018 年创建了原始的 GPT 模型。一年后，他们又发布了 GPT-2，这个新模型展示了大型语言模型的巨大潜力，但同时也引发了 AI 社区的不满，因为 OpenAI 不再是一个非营利实体，而且开源作品越来越少。然而，尽管面临争议，OpenAI 仍取得了显著的研究成果。2020 年，他们发布了 GPT-3，使得 OpenAI 到达一个重要转折点。尽管 GPT-3 的性能出色，但它并没有引起太多关注。为了让更多的人能够使用 GPT-3，Sam 说服了另外两位创始人，提出了用户友好界面的想法。

2021年: 对 DALL-E 1 的研究

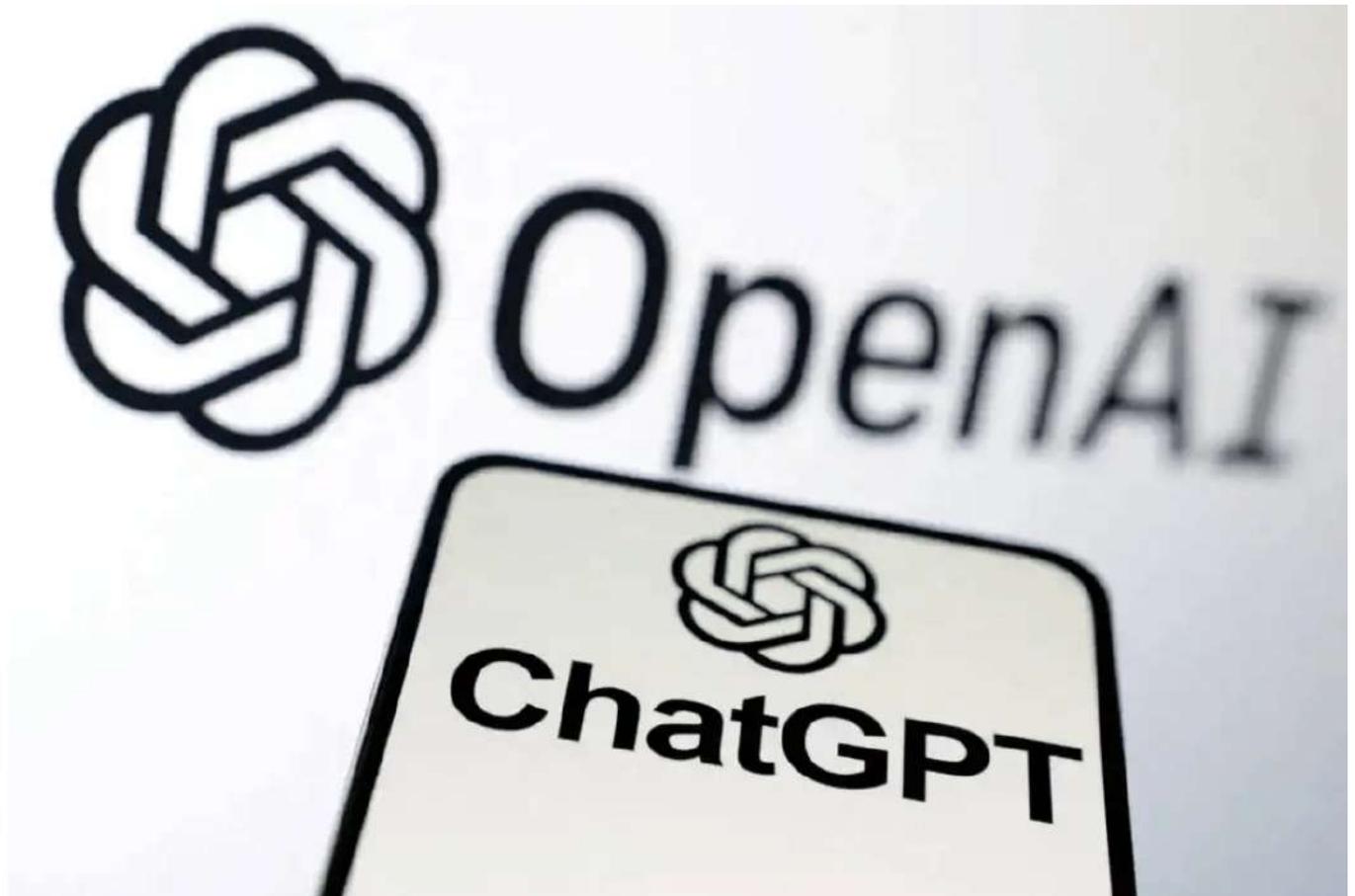
Ilya 总是走在创新前沿，2020 年 6 月，OpenAI 提出了新概念——Image GPT，旨在利用神经网络依据用户的文本指令生成新的高质量图片，而 DALL-E 便是实现了该想法的杰作。在 Ilya 的领导下，OpenAI 团队创造出了由 AI 驱动的图像生成模型 DALL-E，采用与 GPT 模型类似的架构和训练过程，可以根据用户的文本输入生成各种风格的图像。DALL-E 的发展与设计过程对后来的模型（如 DALL-E 2 和 Midjourney）都具有指导作用，它的成功为推动多模态领域的研究和创新奠定了基础。

DALL·E: Creating images from text

2022年：ChatGPT 的诞生与发展

随后，OpenAI 的研究团队对 GPT-3 模型进一步改进，并将新模型命名为 GPT-3.5，采用了最简单的用户界面。由此，语言模型的不断发展促使了 ChatGPT 在 2022 年 11 月 30 日诞生。如今，又衍生出升级版本 GPT-4，再次将生成式人工智能的成就推向一个新高度。

正如 Sam 的先前所说的那样，这一切引发了世界范围内的 AI 变革。在短短的五天内，ChatGPT 吸引了逾百万用户注册使用，创下了产品史上用户增长最快的记录。包括 ChatGPT 的创造者在内，所有人都对这一产品的成功感到震惊。



在 OpenAI 的发展过程中，Ilya 不仅关心技术创新，而且认真对待潜在的风险，他格外关注人工智能的安全性，并在公司内积极为人工智能系统的安全性分配更多资源。他领导着公司的 Superalignment 团队，专注于管理人工智能带来的潜在风险。为实现这一目标，将公司计算能力 20% 的资源用于解决与人工智能安全性相关的问题。

这又将我们带回了他与 Sam 的冲突上，Ilya 的谨慎立场显然与 Sam 更为激进的做法产生了分歧，后者倾向于更快地推进开发强大的人工智能，事态随后发展到 OpenAI 董事会成员策划撤职 Sam Altman，用似乎更为稳重的 Emmett Shear 替代他。但随之另一位联合创始人 Greg 也宣布辞职，同时 700 多名员工联名声讨董事会，仅仅几天后，Sam 又回到了 OpenAI，重新担任 CEO。

总结

在过去的十多年里，Ilya 立足于人工智能变革的前沿，其每个阶段的成果都可能是我们一生也无法企及的高度。Ilya 参与或主导的工作一度改变了计算机视觉领域、机器翻译甚至自然语言处理领域的研究进程，其研究工作被引用超过 42.7 万次，成为引用量最多的计算机科学家之一。在 2015 年，MIT 还将他列为全球 35 位最重要的创新者之一，时代杂志也将他评为最具影响力的人物之一。或许有朝一日，当 AGI 真正改变世界的时候，Ilya 将会被誉为历史上最重要的科学家之一，在人工智能领域的贡献也将会被所有人熟知。

九、数字与言论

1、程序员的任务不仅仅是解决今天的问题，还要确保你的代码不会成为明天的问题。——[《好的代码就像写给接手者的一封信》](#)

2、苏共拥有 20 万党员时夺取了政权，拥有 200 万党员时打败了希特勒，而拥有近 2000 万党员时却失去了政权。——[新华网](#)

3、电影公司为了票房，让电影从独立的故事，转变为互相交叉和联系的系列电影，这样做的风险不断增加，因为故事变得如此复杂，以至于看电影从一种娱乐变成了需要事前做一些家庭作业。——[《超级英雄故事的生命周期》](#)

- 4、按照目前的学术界气氛，我肯定会因为生产力不足，而被大学解雇，幸好后来得到了诺贝尔奖提名，大学才让我留下来。——[彼得·希格斯](#)，“希格斯玻色子”的发现者。从1964年至今，他发表的论文不超过10篇。
- 5、作为一名创始人意味着，要长年累月保持执行力，这是无法外包的。——[Sam Altman](#)，OpenAI 总裁
- 6、Debian 官方下载站，20多年来坚持不使用加密协议 HTTPS，只使用签名文件确保文件可靠。这是为了让那些不支持 HTTPS 的老电脑也能下载到更新包。——[《旧电脑的可用性》](#)
- 7、当你成为某个领域的专家，遇到的问题就会非常具体，很可能只有少数人遇到过。再想进一步前进，互联网也不是很有用，只能靠自己摸索、应用、实验和协作。这些都相当慢，这就是为什么专家再要提升自己，往往要花费数年时间。——[《像旅行者一样学习》](#)
- 8、我拆解了大部分从中国买来的电子产品，里面有很多工程创新，但这些创新并不是针对消费者的，而是针对如何以更少的成本制造这些电子产品。如果有一天，这些工程师转移注意力，关注如何让这些产品有更好的性能、更多的功能，消费者就会感知到他们的创新。——[Hacker News](#)
- 9、世界上最富有的10位富豪，离婚过13次。——[Collab Fund](#)
- 10、具有讽刺意味的是，忠诚的员工更可能被选为企业的剥削目标。企业会假设，忠诚的员工更愿意为企业做出个人牺牲。——[杜克大学的一项研究](#)
- 11、创业公司几乎就是一个“生活在未来”的地方。在创业的早期，你就像活在未来，因为你只关注一件事，就是如何做出一个未来大受欢迎的新产品。——[《你的问题不是你独有的》](#)
- 12、量子纠缠是一种奇怪的量子力学现象，处于纠缠态的两个量子不论相距多远都存在一种关联，其中一个量子状态发生改变，另一个的状态会瞬时发生相应改变。——[微博](#)