



GOTC 2023

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

OPEN SOURCE, INTO THE FUTURE

元宇宙未来展望与开源机会分析

王闻宇 2023年05月28日

王闻宇

现任 PPIO边缘云 联合创始人 CTO

2004 开始做音视频

在学生宿舍写出PPTV第一版，后来主导了PPTV的音视流，P2P/CDN，点播，直播，编解码等，担任PPTV首席架构师

2014 联合创办极豆车联网，服务BBA车厂

2018 联合创办 PPIO边缘云，发展靠近用户侧的边缘节点，提供边缘的算力，传输，存储业务，主要服务音视频行业。



全球开源技术峰会

GLOBAL OPEN SOURCE TECHNOLOGY CONFERENCE

科幻中的 Metaverse

斯蒂芬森将这个平行于物理世界的数字世界命名为“Metaverse”

世界科幻大奖雨果奖桂冠作家力作
尼尔·斯蒂芬森

SNOW CRASH



雪崩
精装版

赛博朋克、宗教哲学、苏美尔神话、语言学

一本信息量相当“秘密”的
超元科幻小说

小说《雪崩》



电影《黑客帝国》



电影《头号玩家》



游戏《赛博朋克2077》

科幻中的元宇宙 — 给人以“类现实”音视频体验

Bysum

全球开源技术峰会

THE GLOBAL OPEN SOURCE TECHNOLOGY CONFERENCE

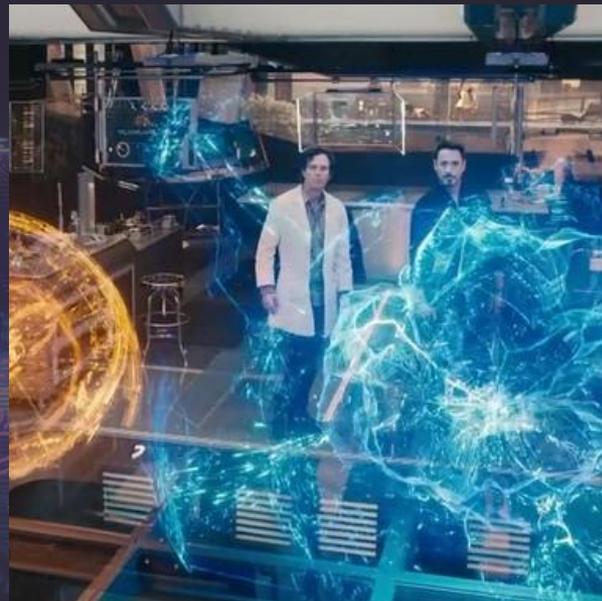
元宇宙的未来展望



视网膜级的视觉体验



精致虚拟世界

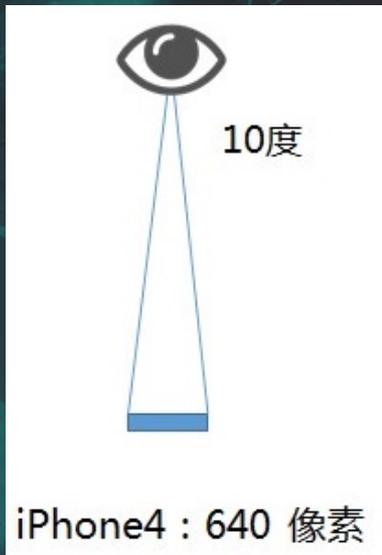
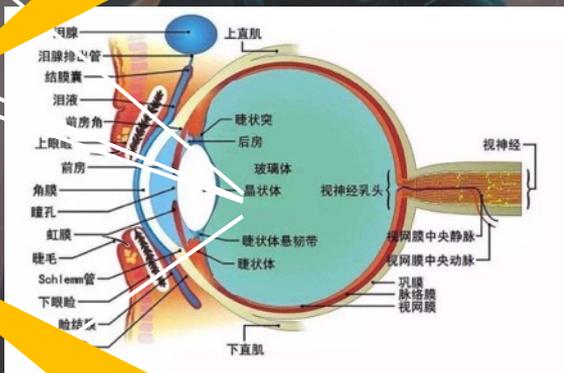


类人智慧的NPC

视网膜级体验的分辨率

PPD=Pixel Per Degree

PPI=Pixel Per Inch



苹果对视网膜屏的定义
大于 300PPI
假定:
屏幕距离眼镜40cm
Iphone4: 640
10度 \leftrightarrow 640像素

FoV = Field of View
视域/视角

相当于:
PPD = 64

根据多项实验证明 视网膜分辨率 \sim 60PPD (每度60像素)

视网膜级体验的分辨率

		角度	Pixel	分辨率	像素	对标
	注意力视域	25度	1500	1500*1500	225w	1080P+
窄域	单眼舒适视域	60度	3600	3600*3600	1300w	4K+
宽域	单眼通常视域	124度	7740	7740*7740	6000w	8K+
	单眼最大视域	156度	9360	9360*9360	8800w	16K-

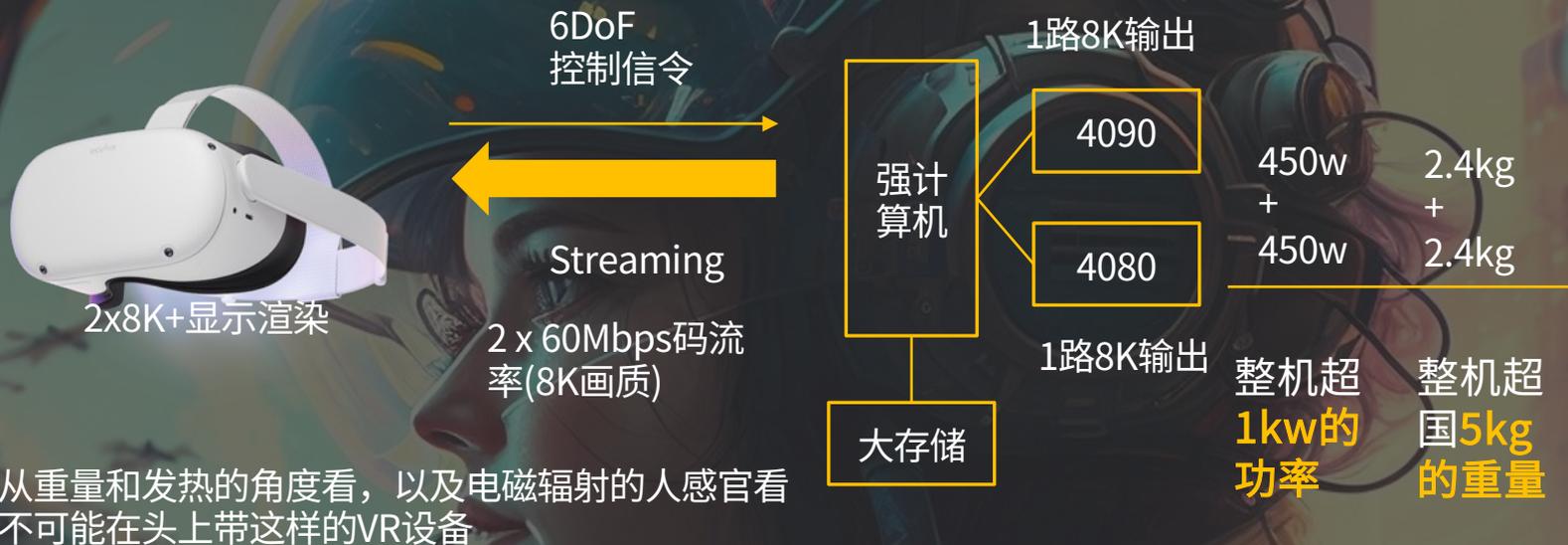
眼动追踪技术
只对注意力集中视域做高清渲染

视网膜分辨率 60PPD
(每度60 Pixel)

视网膜级视觉体验
硬件需要覆盖单眼通常视域
8K+级的分辨率

	分辨率	像素
1080P	1920 x 1080	207万
4K	3840 x 2160	829万
8K	7680 x 4320	3300万
16K	15360 x 8640	1.3亿

云端渲染是 元宇宙的未来设施



元宇宙设备，实时互动的视网膜级视觉体验，**必须串流方案**

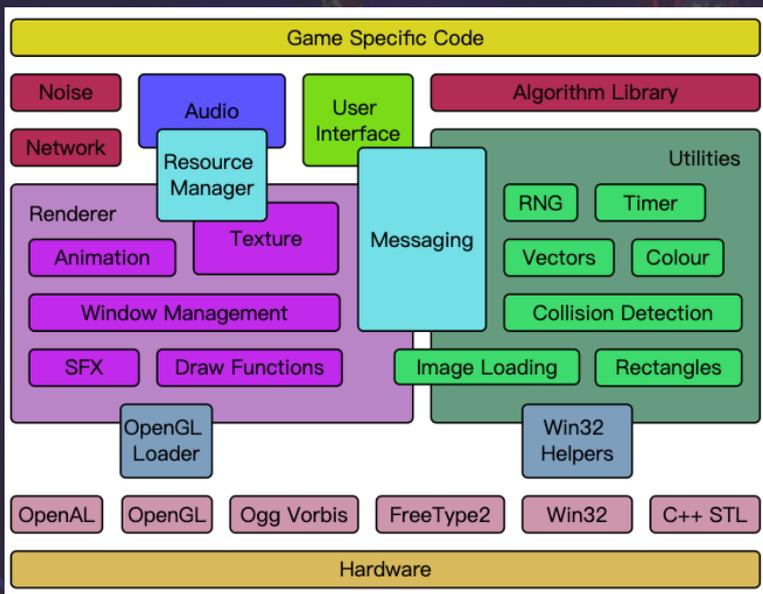
元宇宙精致虚拟世界的基础 – 游戏引擎

自由和灵活

社区支持和合作

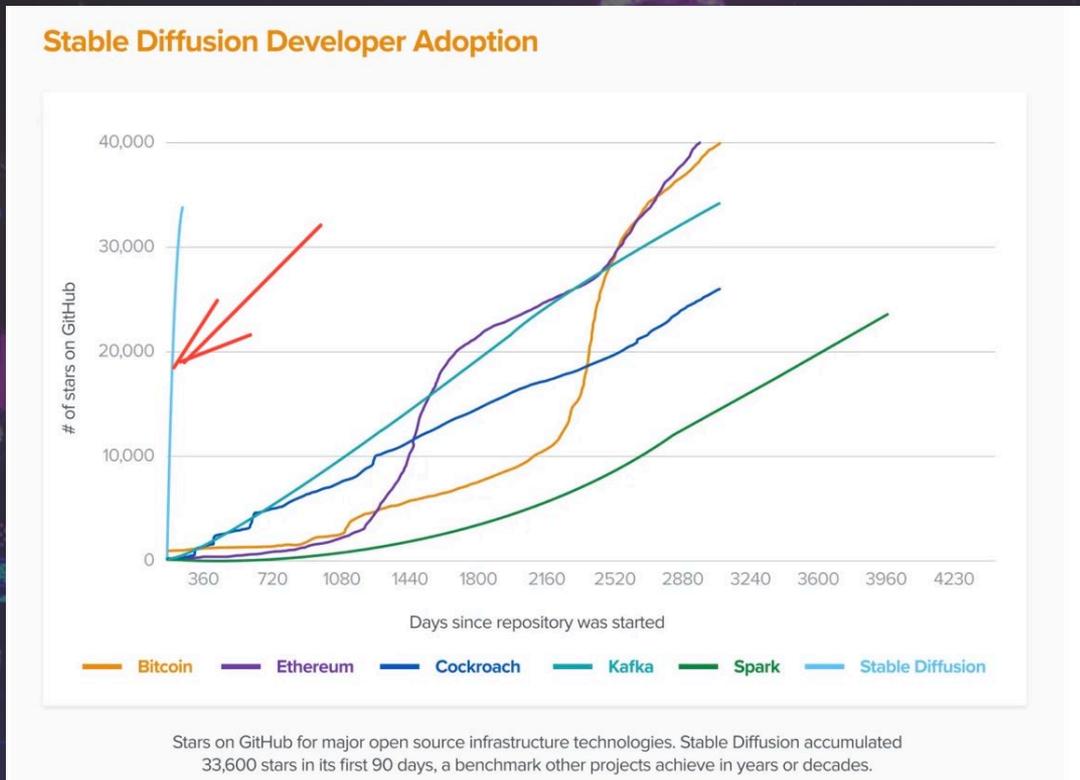
低成本可持续性

学习和教育



项目名称	项目介绍	开源协议	关注数
Godot Engine	一个免费的全能跨平台游戏引擎，支持创建2D和3D游戏 ¹	MIT License	61.4k
cocos2d-x	一个套件，包含了多个开源的跨平台2D和3D游戏引擎，被数百万开发者使用 ²	MIT License	17.1k
libgdx	一个用于桌面、移动和网页的跨平台Java游戏开发框架 ³	Apache-2.0 License ³	21.5k
jMonkeyEngine	一个完整的3D游戏开发套件，用Java编写	BSD-3-Clause License-	3.4k
Open 3D Engine	一个开源的高性能3D游戏引擎支持创建AAA级游戏和电影-	Apache-2.0 License-	1.8k

元宇宙精致的虚拟世界 — AIGC将大大降低虚拟世界设计成本



Image

IMAGE GENERATION

- MitJourney
- craiyon
- wombo.ai
- ROSEBUD.FI
- Lexica
- mage.space
- KREA
- OpenArt
- PLAYGROUND
- PhotoRoom
- alpaca
- Nyx + gallery
- artbreeder

CONSUMER/SOCIAL

- MitJourney

MEDIA/ADVERTISING

- SALT THE CULTURE DAO

DESIGN

- Diagram
- VIZCOM
- Poly
- INTERIOR AI
- uizard
- Aragon
- market
- CALA

MODELS: OPENAI DALL-E 2, STABLE DIFFUSION, CRAIYON

Video

EDITING/GENERATION

- runway
- Fliki
- Düiverse
- Opus

PERSONALIZED VIDEOS

- tavus
- synthesia
- Hour One
- Rephrase.ai
- Colossyan
- Morio

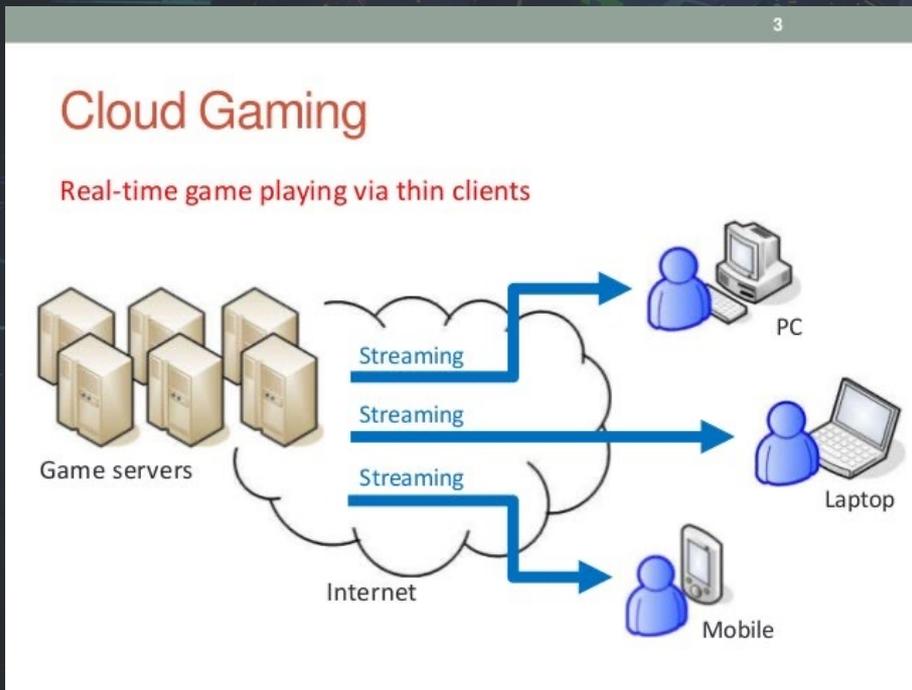
MODELS: MICROSOFT X-CLIP, META MAKE-A-VIDEO

3D

3D MODELS/SCENES

- mirage
- CSM

MODELS: DREAMFUSION, NVIDIA GET3D, MDH, APPLE ML



项目名称	项目介绍	主要原理	开源协议	关注数
Moonlight	一个实现了NVIDIA的GameStream协议的开源项目，支持多种客户端设备	基于NVIDIA的GameStream协议的云游戏传输方法，利用NVIDIA显卡的硬件编码功能提高传输效率和质量	GPL-3.0 License	8.8k
Sunshine	一个自托管的游戏流媒体主机，支持Moonlight客户端	基于Moonlight客户端的云游戏传输方法，支持AMD、Intel和Nvidia GPU的硬件编码和软件编码	GPL-3.0 License	4.8k
GamingAnywhere	一个开源的云游戏平台，支持Windows和Linux操作系统	基于RTSP/RTP/RTCP协议的云游戏传输方法，利用可配置的模块化架构提高可扩展性和可移植性	GPL-2.0 License	1.6k

SD

640 x 480 px

Full HD

1920 x 1080 px

4K

3840 x 2160 px

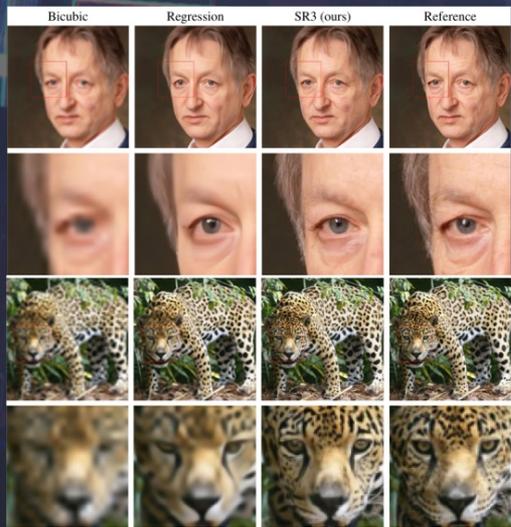
8K

7680 x 4320 px



面对庞大的云端算力，和超大的带宽，其成本非常高
如何找到高性价比方案？会成为一个重要的话题

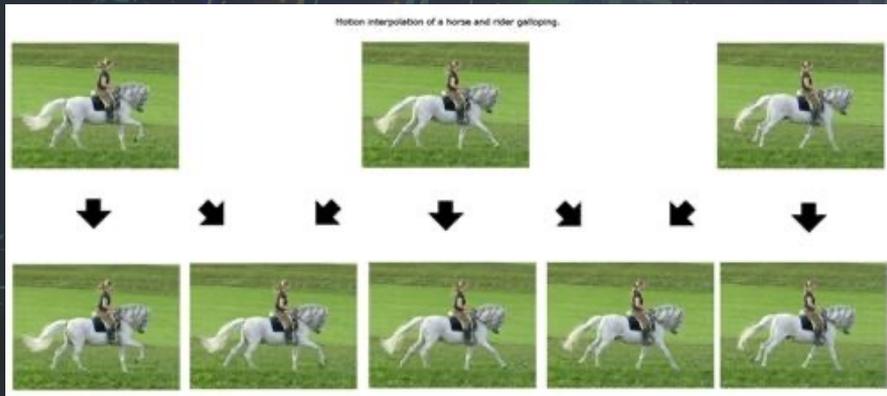
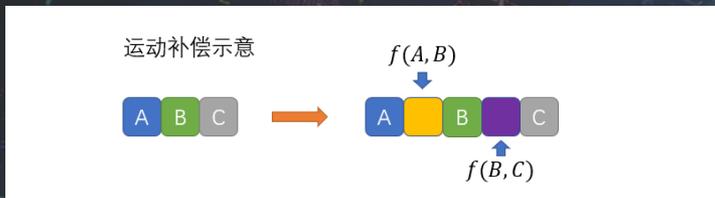
超分辨率



音视频的超分辨率技术，将低分辨率的图片或视频转换为高分辨率。

项目名称	项目介绍	主要原理	开源协议	关注数
Real-ESRGAN	一种基于生成对抗网络（GAN）的超分辨率算法。它使用带有重建和感知损失的对抗网络来提高图像的质量和细节。	生成对抗网络（GAN）、重建损失、感知损失	CC BY-NC-SA 4.0	20.7k
Waifu2x-Extension-GUI	视频、图像和GIF的放大（超分辨率）和视频帧插值工具	基于深度学习的图像重建和插值方法利用多种预训练模型提高图像质量和分辨率	GPL-3.0 License	9.9k
mmagic	OpenMMLab多模态先进、生成和智能创作工具箱	基于深度学习的生成对抗网络（GAN）和扩散模型，实现图像/视频修复/增强等任务	Apache-2.0 License	4.8k

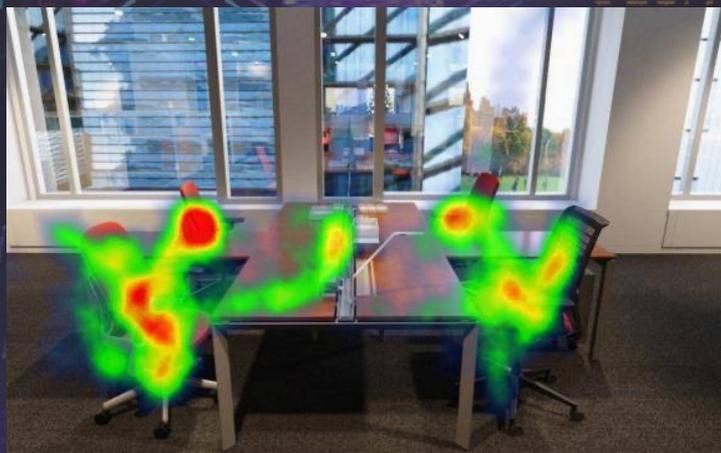
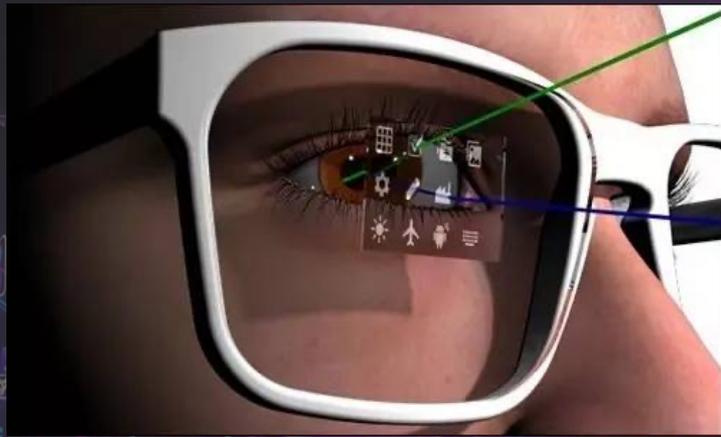
插帧技术



插帧技术：在原始视频的每两帧画面中增加一帧或多帧，缩短帧与帧之间的显示时间，从而提升视频的流畅度和清晰度的技术

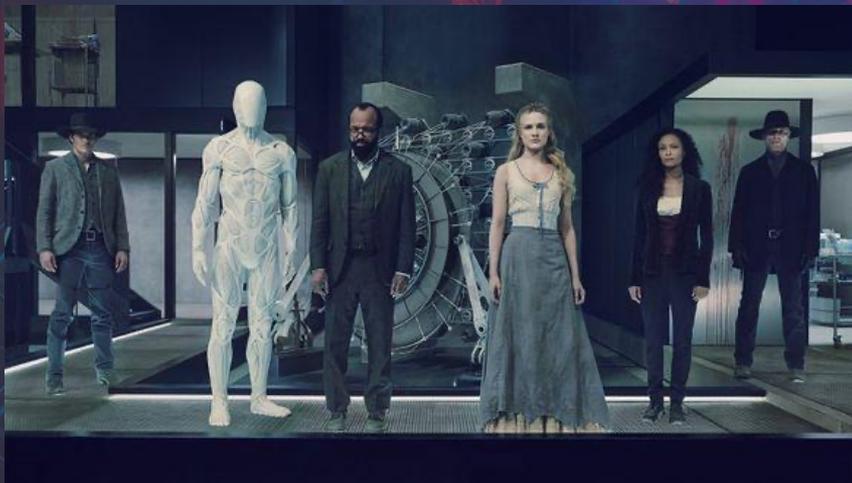
项目	项目介绍	主要原理	开源协议	关注度
DAIN	深度感知视频帧插值方法，可以根据输入帧的深度信息生成中间帧，适用于动态场景和大运动场景。	基于深度感知的上下文特征提取和融合的视频插帧算法包括一个深度估计网络和一个插值网络。	BSD-3-Clause	7.8k
RIFE	旷视科技&北大提出的一种实时插帧方案相比已有基于光流的方案，具有好的性能与推理速度，针对720p视频可以实时进行2x插帧	基于实时中间流估计的视频插帧算法包括一个端到端的高效的中间流估计网络和一个融合网络	MIT	3.3k
Super-SloMo	高质量的慢动作视频生成方法，可以将任意帧率的视频转换为任意慢动作倍数的视频。	基于双向光流估计和双向光流补偿的视频插帧算法，包括一个光流估计网络和一个光流补偿网络。	MIT	2.9k

眼动追踪开源项目



项目名称	项目介绍	主要原理	开源协议	关注数
pupil	一个开源的眼球跟踪软件平台，支持VR和AR设备	基于摄像头和红外光源的眼球跟踪方法，利用机器学习算法提高精度和鲁棒性	GPL-3.0 License	1.3k
gaze-detection	一个用JavaScript实现的眼动检测库，可以用于构建基于凝视控制的体验 ³	基于TensorFlow.js的深度学习模型，利用Webcam捕捉眼睛的位置和方向	MIT License	0.5k
EyeTrack VR	一个开源和低成本的VR眼动追踪平台支持多种社交VR游戏	基于摄像头和红外光源的眼球跟踪方法，利用OSC和UDP协议传输数据	MIT License	0.3k
ImageProcessing Projects	一个用Python和OpenCV实现的图像处理项目集合，包括眼球跟踪	基于OpenCV的图像处理方法，利用Haar特征检测人脸和眼睛，利用Hough变换检测瞳孔中心	MIT License	0.2k

元宇宙中的《西部世界》，智能涌现的NPC



在最近爆一篇论文中，研究者们成功地构建了一个「虚拟小镇」，25 个 AI 智能体在小镇上生存，它们不仅能够从事复杂的行为，而且这些行为比人类角色的扮演更加真实。

Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

Meredith Ringel Morris
Google Research
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu



Figure 1: Generative agents create believable simulacra of human behavior for interactive applications. In this work, we demonstrate generative agents by populating a sandbox environment, reminiscent of The Sims, with twenty-five agents. Users can observe and intervene as agents they plan their days, share news, form relationships, and coordinate group activities.

AI的发展历程

基于模板和规则的前深度学习阶段

根据一定范围的数据进行参数分类

开始模仿人脑进行大量数据的标记和训练

对人脑学习过程进行重点关注

基于规则的少量数据处理

Machine Learning

神经网络

CNN

RNN

VAE

GAN

自注意力机制

Transformer

Diffusion

走向了大语言模型

走向了AIGC绘画

1950年开始

1980年开始

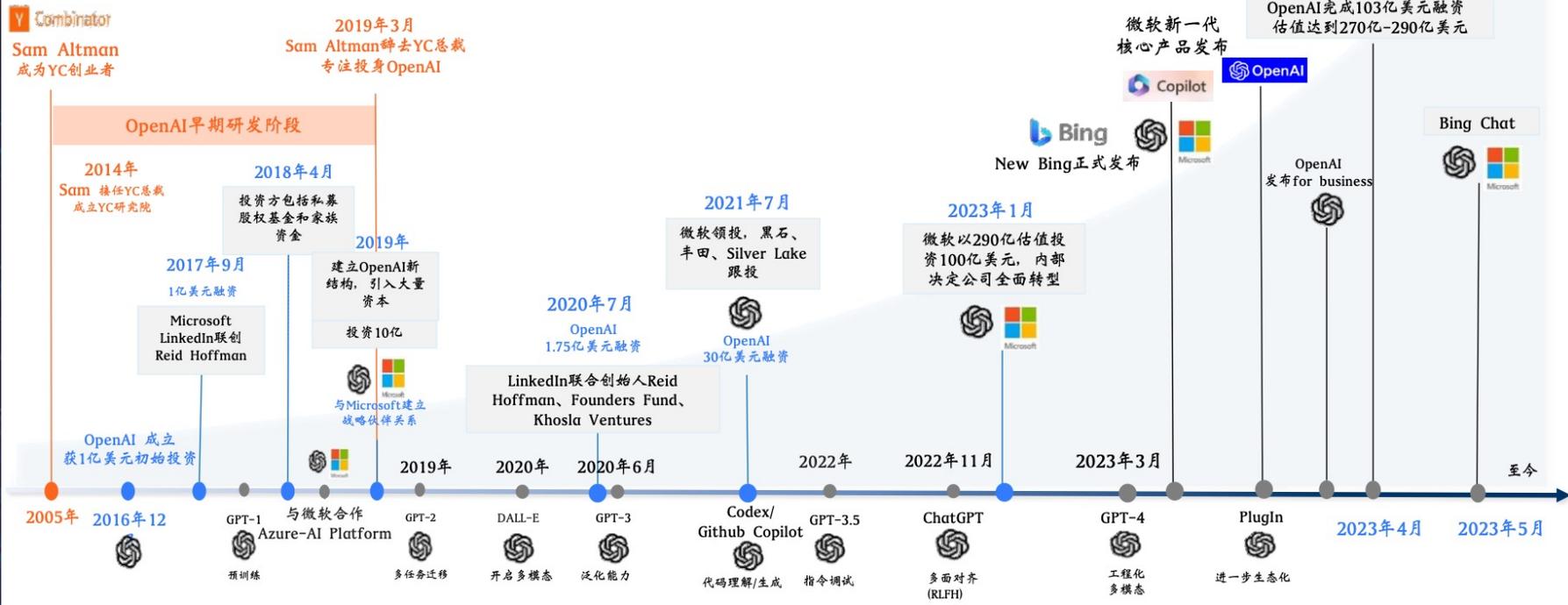
1990年

2014年

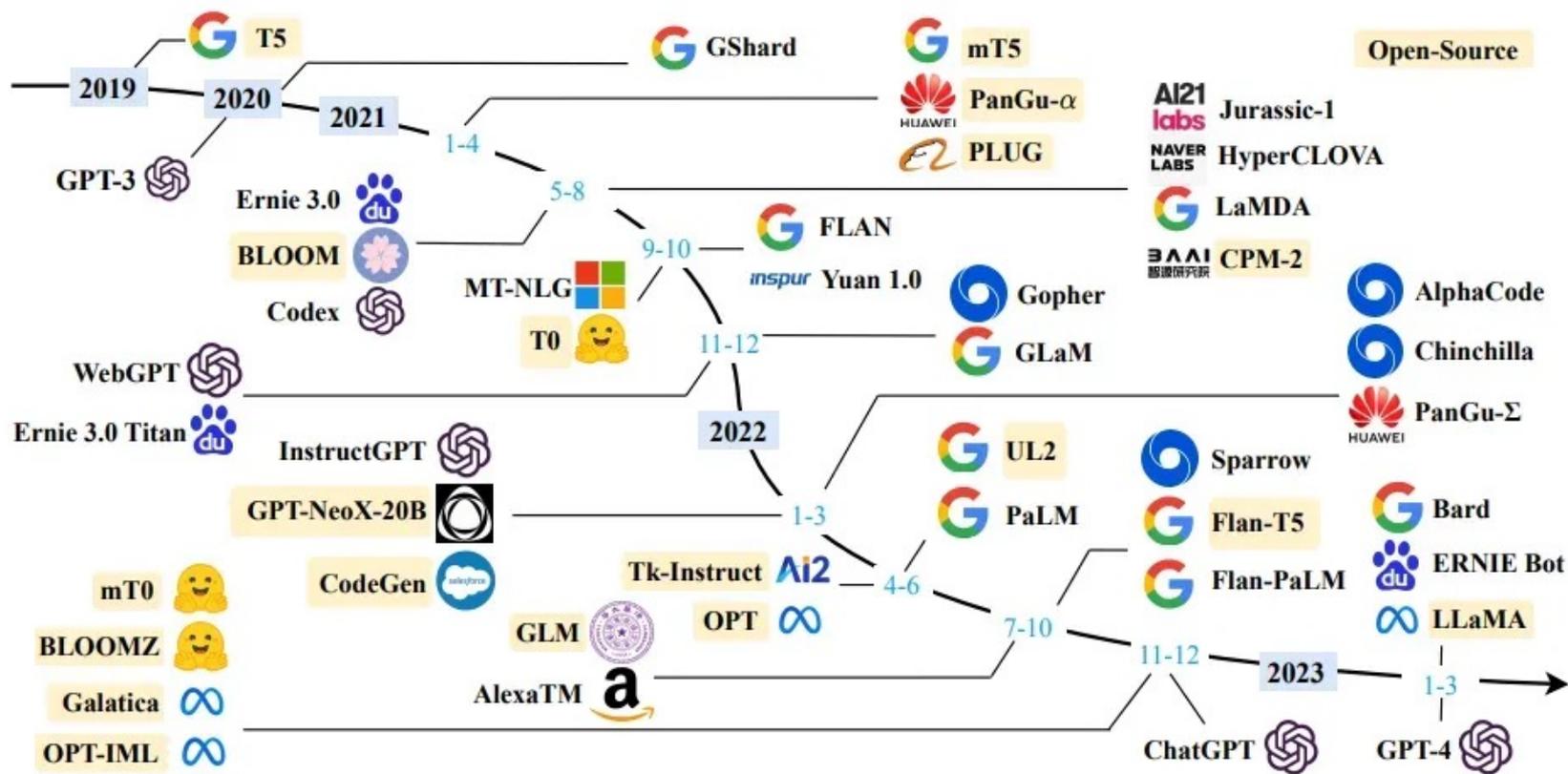
2017年

2020年

大语言本省的发展历程



大语言开源模型其实也没有停滞



开源大语言模型的思考



移动互联网时代



公开数据集
越来越通用
模型越来越大



群雄逐鹿
最终1-2家胜出

私有数据集
行业应用
推理模型变小
普惠算力

PPIO边缘云 – 打造元宇宙时代的基础设施



PPIO边缘云使命：汇聚全球计算资源，并为全人类服务





THANKS