

2023 AI 大事件



ZOMI



01

百模大战



百模大战

- ChatGPT的引领：
 - 2022年11月，OpenAI发布ChatGPT，迅速引爆全球对大模型的关注。2023年，ChatGPT 月活用户突破1亿，成为历史上用户增长最快的消费级应用。
- 国内大模型崛起：
 - 反观国内截至 2023 年 10 月，国内已发布的大模型数量超过130个，大模型数量迅速增长，涵盖通用大模型、行业大模型和应用服务型大模型，也就是从 L0、L1、L2 全覆盖。
- 科技巨头：
 - 百度、阿里、腾讯、华为、字节等互联网大厂纷纷入局大模型。23 年3 月份百度率先发布“文心一言”，阿里推出“通义千问”，华为发布“盘古大模型 3.0”。



百模大战：初创企业

- 初创企业方面，出现了AI大模型6小虎，智谱AI、月之暗面、MiniMax等。
 - 百川智能：2023年3月成立，由前搜狗CEO王小川创立。以开源大模型为特色，迅速推出 Baichuan 7B/13B 系列模型，并在医疗领域展开布局。
 - 零一万物：2023年5月成立，由李开复团队孵化，推出了 Yi-34B 模型和 AI 工作平台万知，不过因为开源代码被发现套用 LLAMA，所以在开发者中积累了不好的口碑（不过在 25 年年初宣布被阿里并购）。
 - 月之暗面：成立于2023年，创始人为杨植麟，推出了Kimi智能助手，凭借超长上下文输入一鸣惊人，在C端市场迅速走红。
 - Minimax：闫俊杰带领商汤原班人马出走，早期推出了文本到视觉、文本到语音等多模态模型，并通过C端产品专注于海外市场。



百模大战：初创企业

- Minimax：闫俊杰带领商汤原班人马出走，早期推出了文本到视觉、文本到语音等多模态模型，并通过C端产品专注于海外市场。
- 阶跃星辰：2023年6月成立，创始团队包括微软前全球副总裁姜大昕等人创立，主要归属上海的国家队，2024年6月完成一轮20亿美元的融资，推出了多个多模态大模型。阶跃星辰：成立于2023年。
- 智谱：成立于2019年，是国内最早探索大模型的公司之一。从成立之初就专注于大模型技术。2023年，智谱发布了 GLM 系列模型。在 2023 年完成多轮融资，估值超过 200 亿元。
- 科研机构：
 - 北京智源研究院、上海人工智能研究院、清华大学、中科院等高校和科研机构，也在大模型领域取得重要进展，推动了技术的开源和普及。



百模大战：商业化探索

- 应用落地：
 - 同年，大模型开始从实验室走向实际应用。百度“文心一言”在教育、金融等领域应用（不过没有动其搜索），阿里“通义千问”在电商和营销场景的落地（不过落地情况不是很好，24 年通义千问实验室拆分到不同业务，云保留通义千问实验室）。
- 价格战与成本压力：
 - 2023年下半年，大模型企业开始通过降价策略争夺市场。字节的“豆包大模型”将每千 tokens 输入价格降至 0.003 元，比行业价格低 85%。
- 算力与成本挑战：
 - 大模型的训练和推理成本高昂，算力资源成为制约发展的关键因素。GPT-4 训练需要 1 万 NV A100，成本高达 1 亿美元。



02

多模态与生成式AI



多模态与生成式AI：GPT-4 发布

- GPT-4 发布：
 - OpenAI 在 2023 年 3 月发布 GPT-4 多模态大模型，能够处理文本和图像，并在复杂任务中表现出色。
- 意义：
 - GPT-4 发布进一步推动了生成式 AI 普及和成熟，展示了在自然语言处理、代码生成等领域的强大能力。
- 多模态大模型 OpenAI GPT-4、DALL·E 2 成为生成式 AI 技术的主流路线，推动了文本、图像、视频等多模态内容的生成与应用。
- 生成式图像领域迎来了 Stable Diffusion 等开源工具的繁荣，而生成式视频领域在23 年下半年迎来了Runway、Pika 等应用。



03

AI 在跨学科领域



医疗和生命科学中的应用

- AlphaFold:
 - DeepMind 的 AlphaFold3 成功预测了 2.14 亿个蛋白质结构，几乎涵盖了整个蛋白质宇宙。为生物学研究和药物研发提供了重要工具。
- 意义:
 - AlphaFold 展示了 AI 在生命科学领域巨大潜力，推动了蛋白质结构预测的普及和应用，也为 2024 年诺贝尔奖颁给计算机学家埋下伏笔。

04

AI 自动驾驶应用



AI 自动驾驶应用

- 事件：
 - 百度发布了第六代量产无人车 Apollo RT6（萝卜快跑使用的技术），成本降低至 25 万元。
 - tesla 2023 AI day 没有来，2023 年 8 月 26 日，马斯克通过直播展示了特斯拉 FSD V12 的端到端自动驾驶能力。
 - 2023 年 10 月 24 日，小鹏汽车发布 XNet 2.0 感知架构，首次将大模型技术应用于自动驾驶感知系统，提升了车辆对复杂交通环境的理解能力。
- 意义：
 - 自动驾驶技术的商业化标志着 AI 在交通领域的实际应用迈出了重要一步，而特斯拉则展示了大模型对传统自动驾驶算法的冲击（2024 年车厂的 AI 技术革命埋下悬念）。

05

AI 开源生态



AI 开源生态的繁荣

- Meta 发布 LLaMA：
 - Meta发布了 LLaMA 模型并被开源，引领了开源热潮。随后，23 年 7 月份 Meta 又发布了免费可商用的 Llama 2，成为许多模型选择的基座大模型。
 - Meta发布的 LLaMA 引发了开源大模型的热潮，推动了生成式 AI 的生态发展开源与生态建设。开源文化的盛行推动了大模型的技术快速共享与进步，降低了大模型技术门槛。



06

AI 芯片与硬件



AI芯片与硬件的突破

- H200:
 - 英伟达在 2023 年 11 月发布了 H200，与 H100 相比，H200 的内存更大（141GB），带宽更高（4.8 TB/s），分别约为 H100 的 1.8 倍和 1.4 倍。
- 意义：
 - H200的发布为 AI 计算提供了更强大的硬件支持，进一步巩固 NV 在 AI 硬件领域的领先地位。
- 企业自研芯片：
 - 包括微软、Meta、亚马逊、阿里、腾讯、字节等多家企业开始自研 AI 芯片，以满足日益增长的 AI 算力需求。



总结



总结

- 2023 年是 AI 领域取得重大突破的一年，引起了国内外的百模大战。OpenAI 发布 GPT-4 和 ChatGPT 进一步推动了生成式 AI 普及，展示了多模态和自然语言处理的强大能力。DeepMind 的 AlphaFold 成功预测了 2.14 亿个蛋白质结构，为生命科学领域带来革命性突破。在医疗、自动驾驶和艺术创作等领域得到广泛应用。
- 同时，AI 芯片（英伟达 H200）和开源模型（Meta Llama 2）快速发展为 AI 提供了强大发展动力。
- 最后，AI 伦理与监管成为焦点，欧盟推出《人工智能法案》，美国在拜登的领导下发布 AI 行政命令，强调安全与责任，也进一步加码对中国禁售最先进的英伟达 AI 芯片。



Thank you

把AI系统带入每个开发者、每个家庭、
每个组织，构建万物互联的智能世界

Bring AI System to every person, home and
organization for a fully connected,
intelligent world.

Copyright © 2024 XXX Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. XXX may change the information at any time without notice.



GitHub <https://github.com/chenzomi12/AIFoundation>