

AI 驱动软件研发 全面进入数字化时代

AI+
software
Development
Digital
summit



多模态大模型驱动新一代技术变革

演讲人 王金桥

科技生态圈峰会 + 深度研习 —— 1000+ 技术团队的选择



2023K+
全球软件研发行业创新峰会

上海站

会议时间 | 06.09-10



2023K+
全球软件研发行业创新峰会

北京站

会议时间 | 07.21-22



2024K+
全球软件研发行业创新峰会

深圳站

会议时间 | 05.17-18



K+峰会详情



会议时间 | 08.18-19

NDD AI+软件研发数字峰会

北京站



会议时间 | 11.17-18

NDD AI+软件研发数字峰会

深圳站



▶ 演讲嘉宾

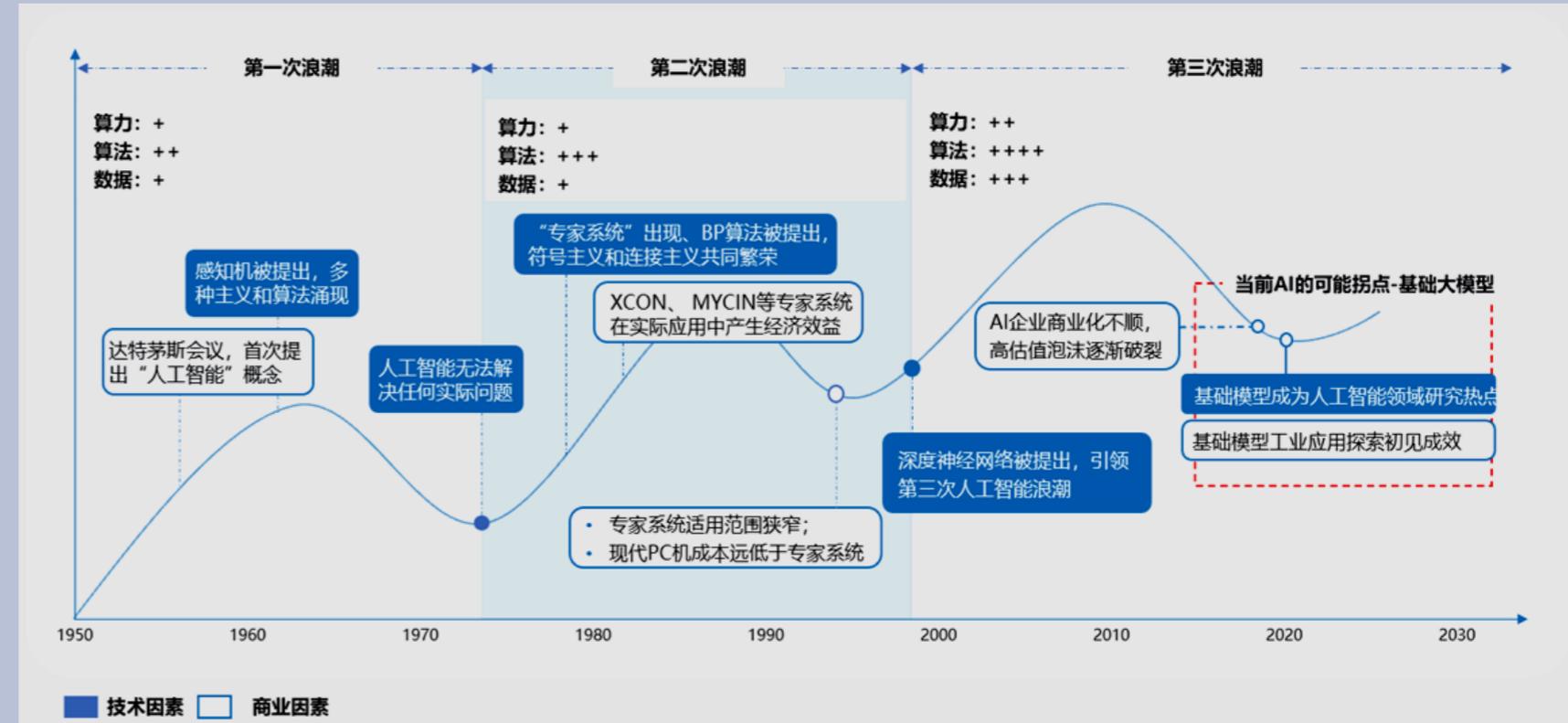
王金桥 中科院自动化研究所紫东太初多模态中心常务副主任

中国科学院自动化研究所副总工程师，紫东太初多模态中心常务副主任，研究员，博导，武汉人工智能研究院院长，中国科学院大学人工智能学院岗位教授，多模态人工智能产业联盟秘书长，中国技术创业协会技术创新工作委员会副理事长，获北京市高聚领军人才、广州市创新团队领军人才、山东省泰山领军人才等称号。共发表包括IEEE国际权威期刊和顶级会议论文300余篇，国际期刊50余篇，国际会议220余篇。完成国家标准提案3项，发明专利36项，10项国际视觉算法竞赛冠军，北京市科技进步一等奖，吴文俊人工智能科技进步奖，中国发明创新银奖，世界人工智能大会SAIL（卓越引领者）奖等。

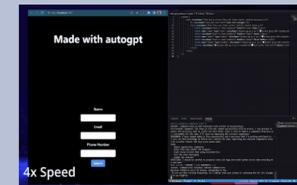


▶ 大模型开启人工智能新一代技术革命

“大算力+大数据+大模型”标志着通用人工智能技术产品化的来临，成为产业化分水岭，将实现对劳动力、资本等生产要素的智能替代和功能倍增，促进全要素生产率的提高。



重塑办公模式



全程自动设计



AIGC机器替人

...

开放环境、复杂场景、不同应用...
技术功能需求不断增多

计算智能→感知智能→认知智能→?
智能形态水平持续深入

二进制→图像用户界面→意图解析
人机交互越来越智能

国内外大模型井喷式爆发 (语言模型为主)



通过自监督 + 学习条件下实现
“大数据 + 大模型”方式，引发
人工智能发展新浪潮。

▶ ChatGPT资源投入

资源名称	投入情况
数据	来自互联网、书籍或维基百科的45TB文本数据（近1万亿个单词来训练它，即大约1,351万本牛津词典）和179GB的代码
算力	GPT-3训练一次需要460万美元，ChatGPT增加了代码数据，训练成本更高
人员	375名员工，大部分是世界级顶尖人才。另有数据外包人员几千人和40人专家级标注人员。
资金	OpenAI成立时投入10亿美元，2019年微软又投入了10亿美元，2023年后微软继续投入几百亿美元。根据美国《财富》报道，OpenAI 2022年净亏损5.45亿美元。ChatGPT生成一条问答成本为1.3美分左右，按照月活用户1亿，每个平均20条，算力运营费将达到1千万美元左右/月
合作单位	由OpenAI主导，微软深度参与，其他单位辅助或部分人员跳槽到其他单位，包括Johns Hopkins University、Anthropic、Alignment Research Center、github、Zipline

- ChatGPT在数据、算力、资金、人才投入巨大。由OpenAI主导，微软Azure云、Johns Hopkins University深度参与，其他单位部分参与或员工跳槽带来的技术外溢。

*Equal contribution

Johns Hopkins University, OpenAI

Author contributions listed at end of paper.

*Equal contribution

¹OpenAI, San Francisco, California, USA.

²Anthropic AI, San Francisco, California, USA. Work performed while at OpenAI.

³Zipline, South San Francisco, California, USA. Work performed while at OpenAI.

GPT-3论文标注单位

1 Introduction

Large language models (LMs) can be “prompted” to perform a range of natural language processing (NLP) tasks, given some examples of the task as input. However, these models often express unintended behaviors such as making up facts, generating biased or toxic text, or simply not following user instructions (Bender et al., 2021; Bommasani et al., 2021; Kenton et al., 2021; Weidinger et al., 2021; Tamkin et al., 2021; Gehman et al., 2020). This is because the language modeling objective

*Primary authors. This was a joint project of the OpenAI Alignment team. RL and JL are the team leads. Corresponding author: lpw@openai.com

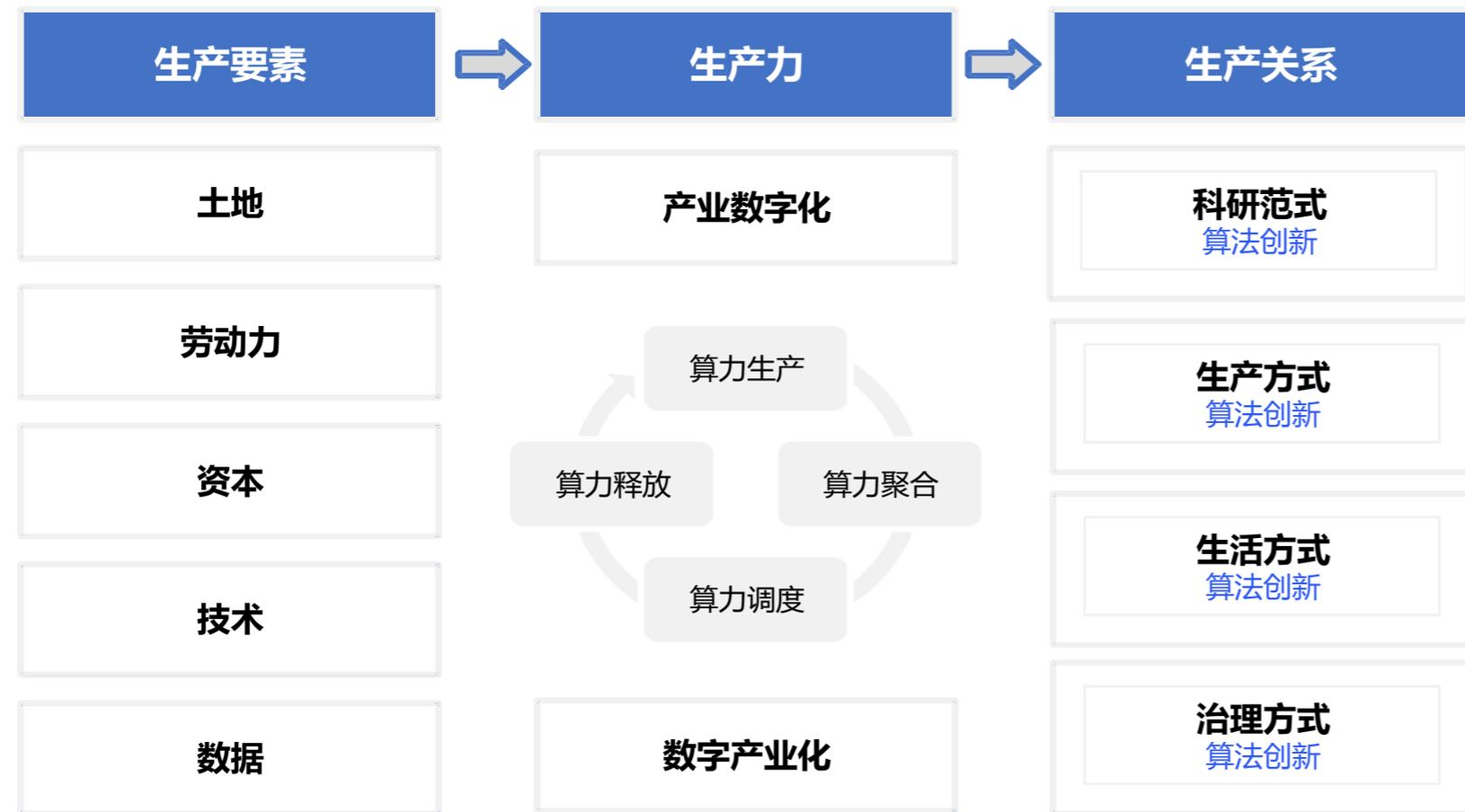
†Work done while at OpenAI. Current affiliations: AA: Anthropic; PC: Alignment Research Center.

OpenAI Codex论文标注单位

InstructGPT论文标注单位

资料来源：OpenAI官网, Nerdynav.com/chatgpt-statistics, 美国《财富》

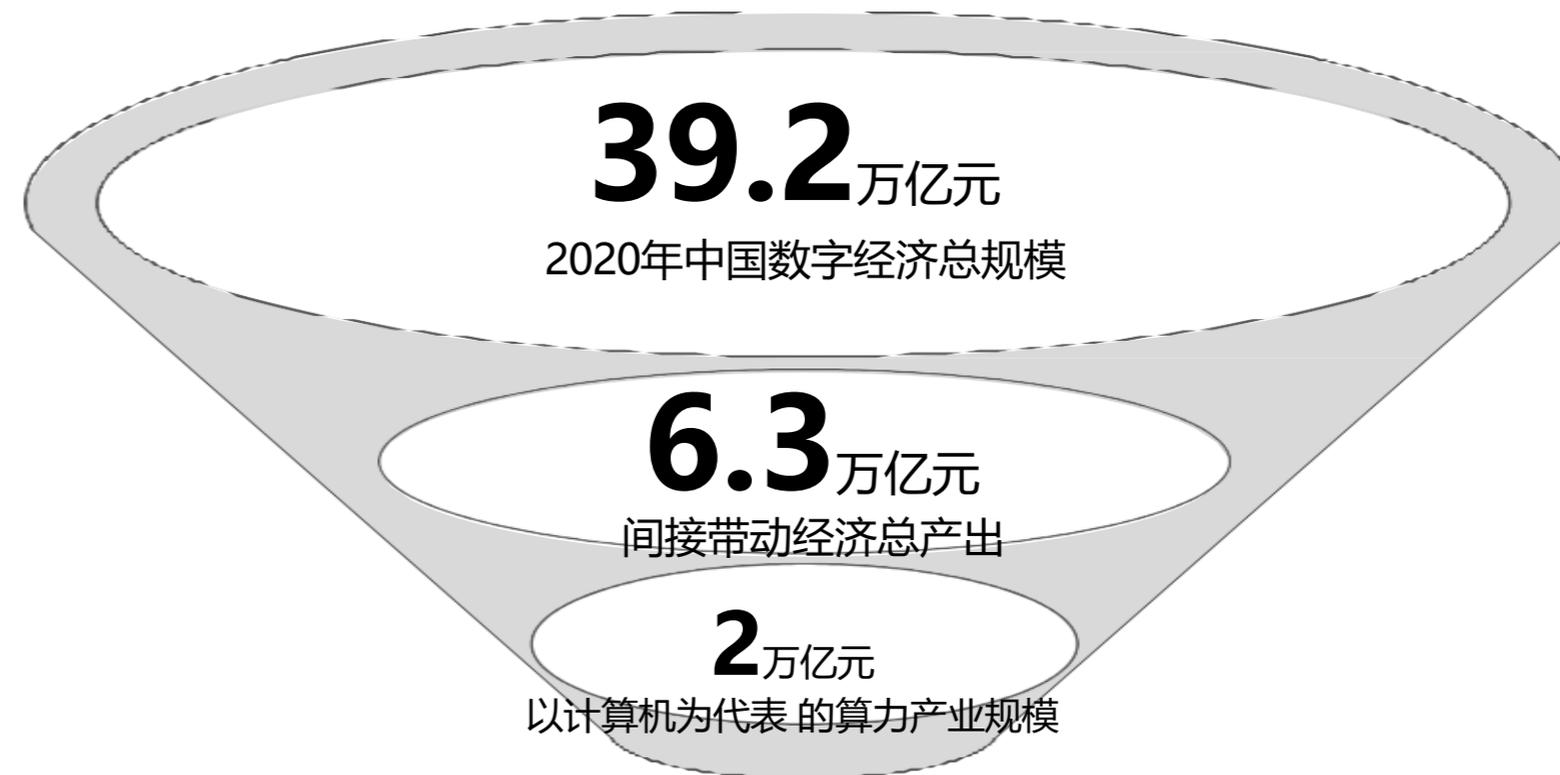
算力：数字经济时代的新生产力



数据、算力和算法是数字经济时代的关键资源

数据是新的生产资料，算力是新生产力，算法是新的生产关系

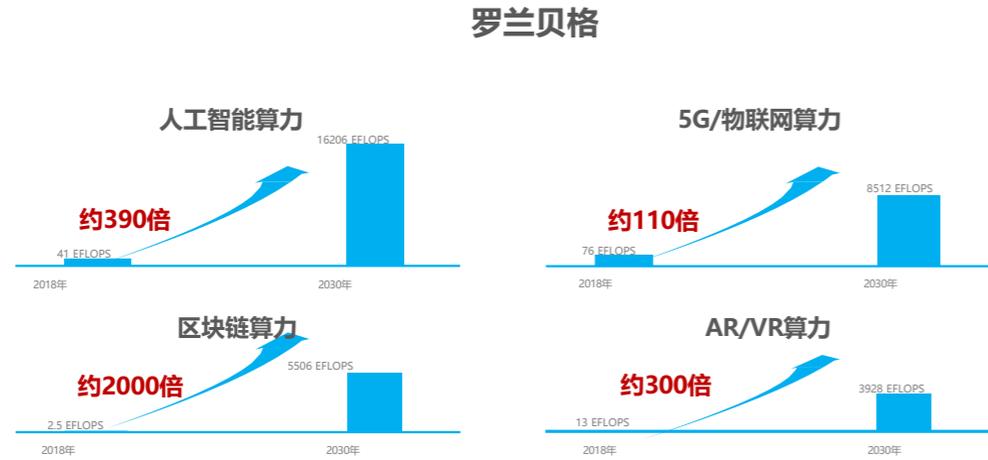
在算力中每投入1元，带动3-4元经济产出；
算力发展指数每提高1点，GDP增长约1293亿元。



来源：《中国算力发展指数白皮书》，信通院

业界普遍乐观看待算力增长前景

乐观预测



信通院

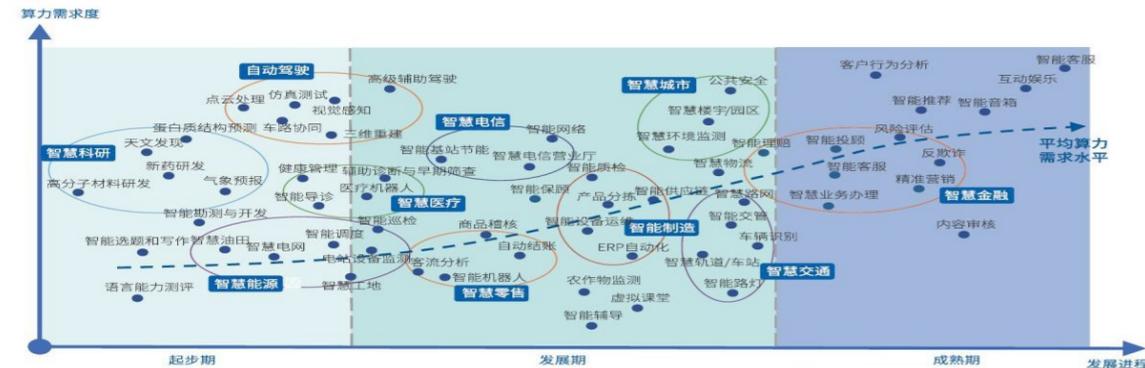
2020年全球算力规模达到**429EFlops**，增速达39%，预估未来5年全球算力规模将以超过50%的速度增长，到2025年整体规模达到**3300EFlops**。

IDC



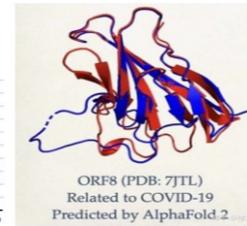
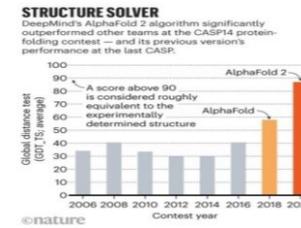
需求驱动

经济社会数字化、智能化转型

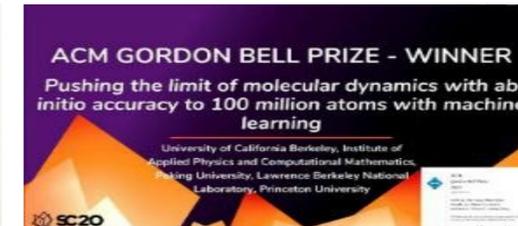


科技创新范式变革

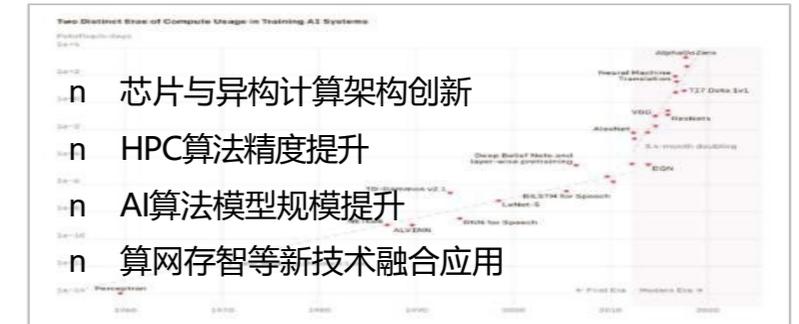
AlphaFold 预测蛋白质结构



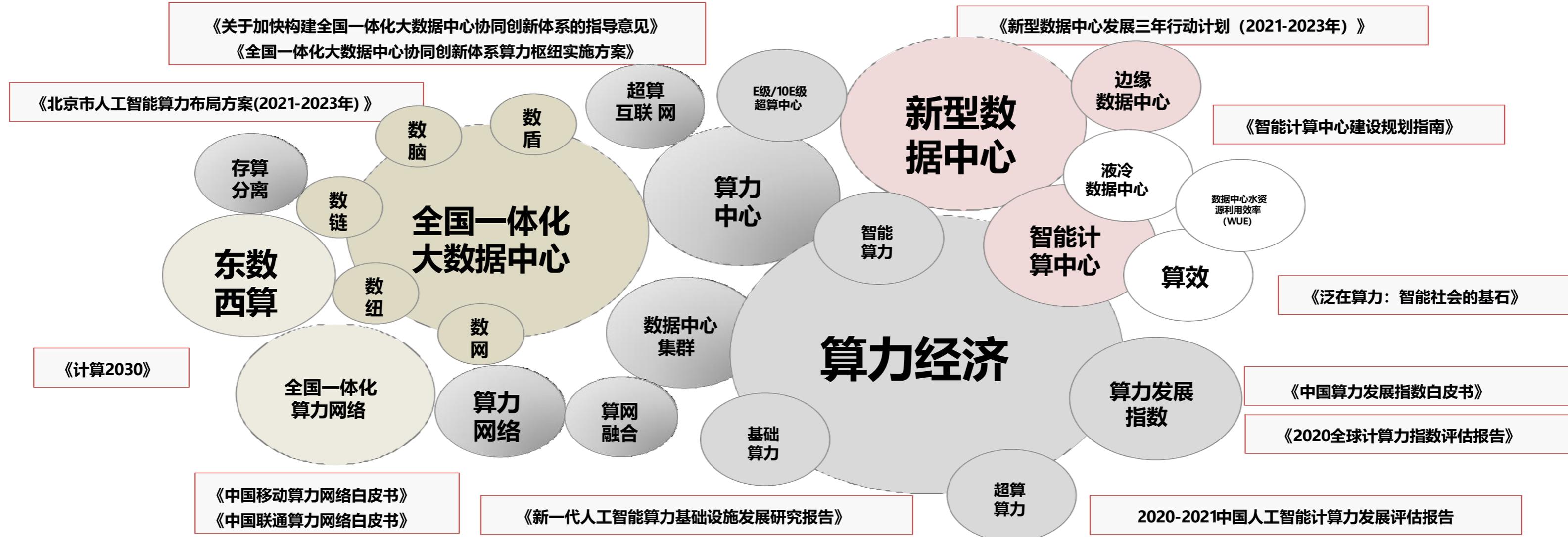
1 亿原子分子动力学模拟



算力算法数据协同发展

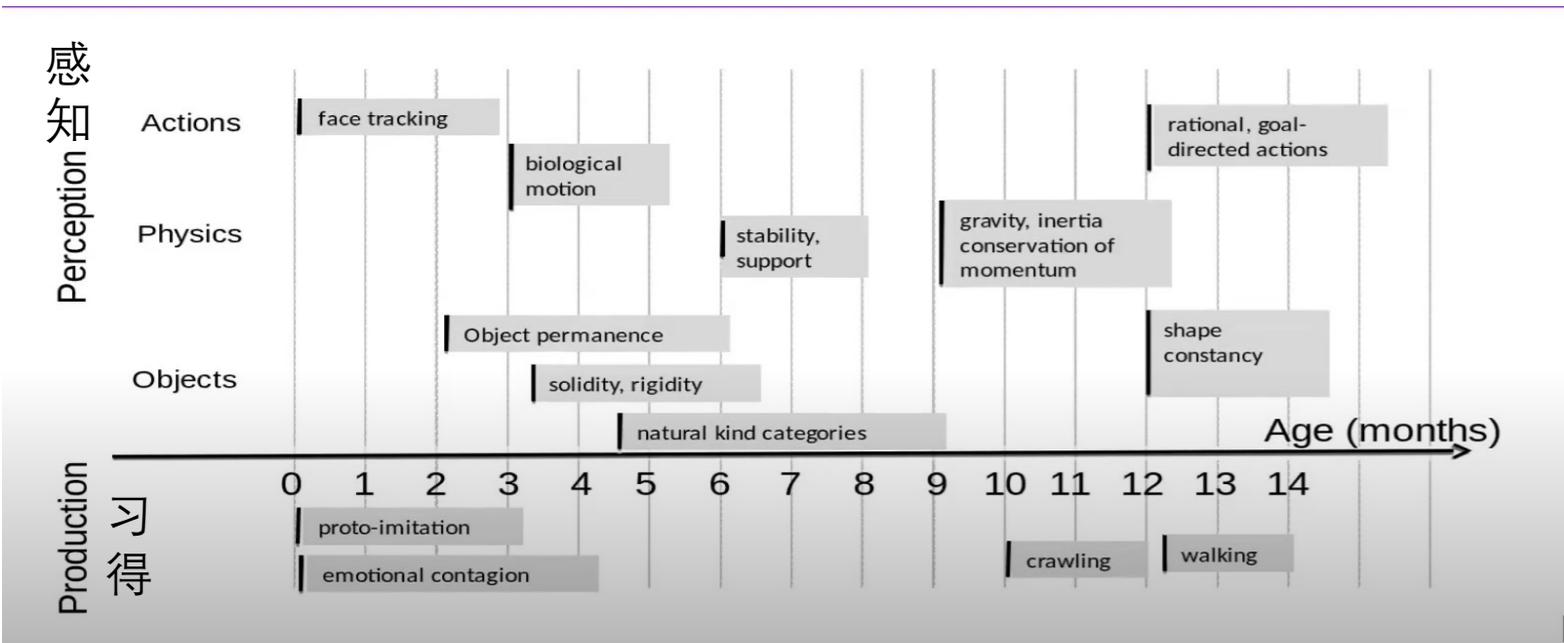


大算力、大模型发展涌现了一批新场景、新机遇



人类的学习和交互过程充满了多模态信息

婴儿阶段的早期概念习得过程 (from Emmanuel Dupoux)



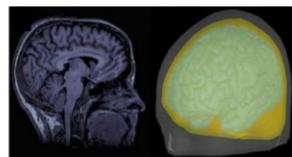
各类模态的示例：

- 自然语言 (说/写)
- 视觉 (图像/视频)
- 听觉 (声音/语音/音乐等)
- 触觉
- 嗅觉/味觉
- 生理信号 (心电图ECG, 皮肤电传导等)
- 其他 (红外图像, 深度图, 核磁共振图像)

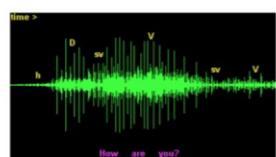
应用和行为方面：



Psychology
心理学



Medical
医学



Speech
声学



Vision
视觉



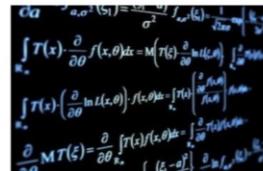
Language
语言



Multimedia
多媒体



Robotics
机器人



Learning
学习

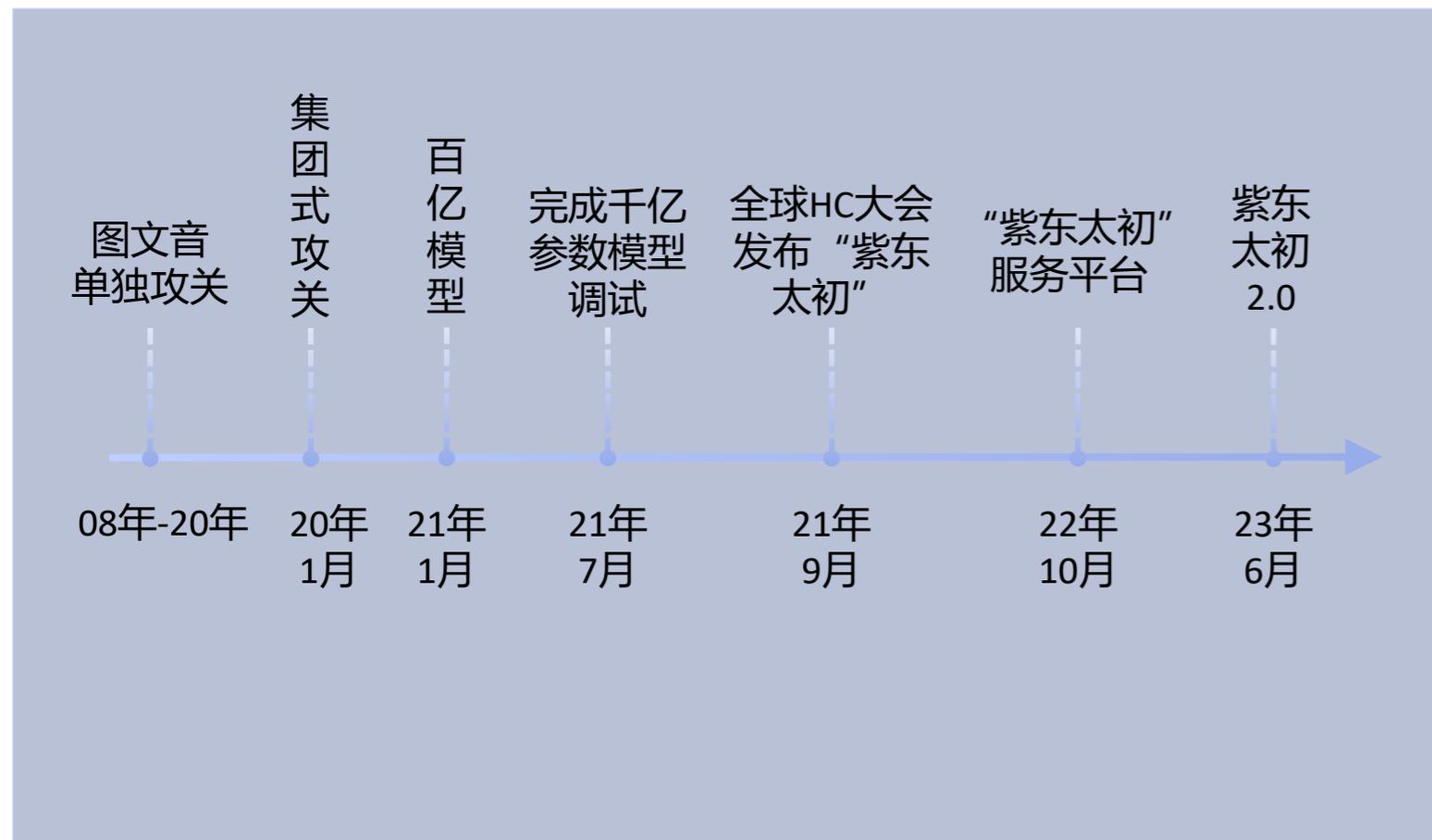
BABY'S 1ST YEAR

GUIDE TO BABY'S SENSORY + MOTOR SKILL DEVELOPMENT



BABY MILESTONE CHART
BROUGHT TO YOU BY:
the BUMP
* Your baby is unique, so these milestones are only a general guide.

紫东太初大模型攻关历程



不同场景对应不同小模型

一个场景一个模型
模型参数量小，泛化性差

1000 2000 10000 20000 参数

从0开始 独立调优 艰难迭代 推倒重来

多模态大模型吸收海量知识

多个场景一个模型
模型参数大，泛化性强

1000亿 参数

预训练 微调

算力中心（北京-武汉）

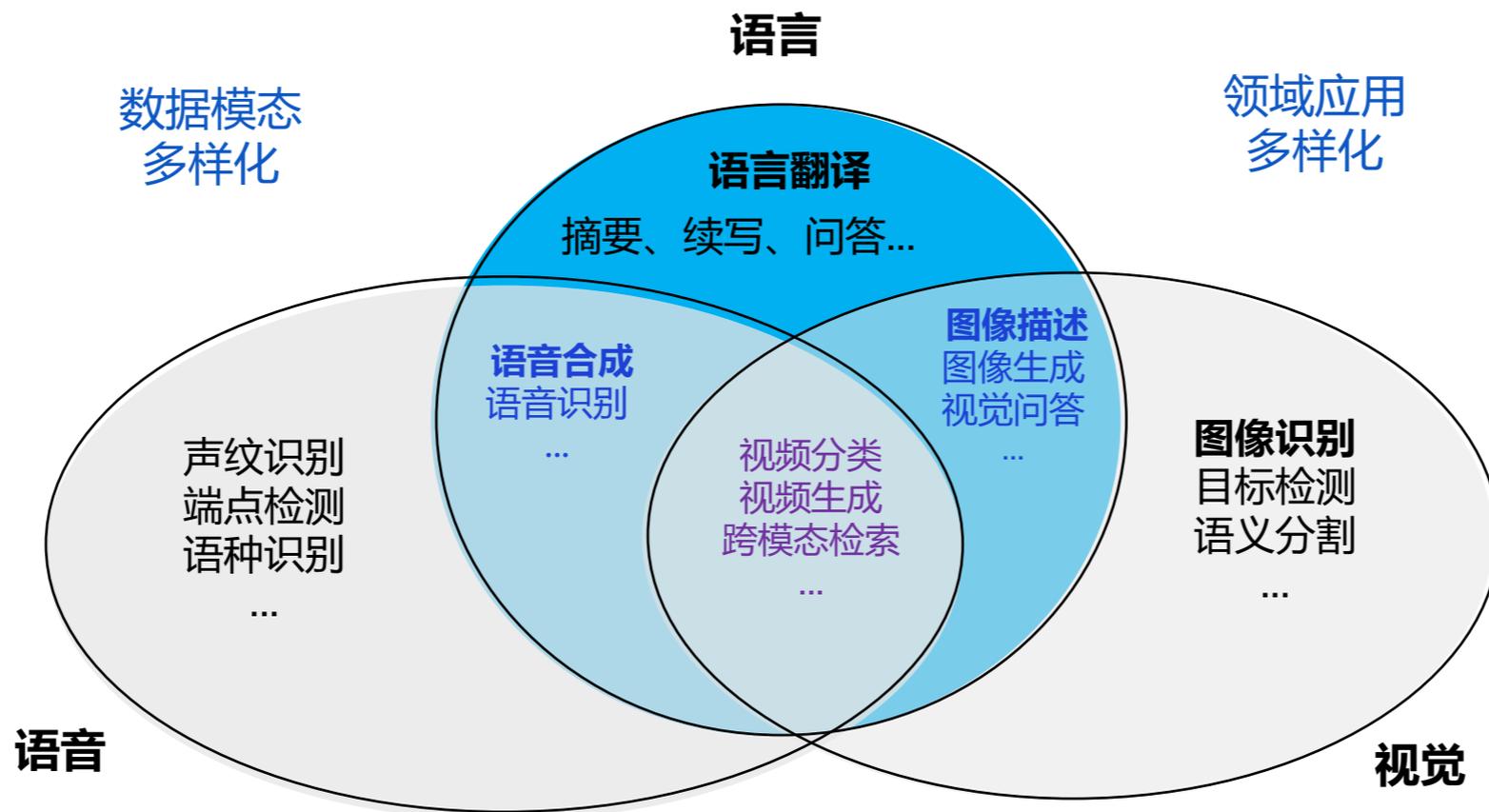
海量弱关联数据

在照片的前半部分，我们可以看到一个干草区，那里有一个女人站着，穿着浅蓝色牛仔裤和灰白色长袖衬衫。

图像 文本 语音

紫东太初”大模型1.0：一个模型支撑多个任务

“紫东太初”可以**一个模型**实现图像识别、语言翻译、图像描述、语音合成等**50余种**功能，并初步具备了人机交互、认知推理、艺术创作等创新能力。例如我们可以以音生图、可以理解图像含义。



一个模型实现50余种功能

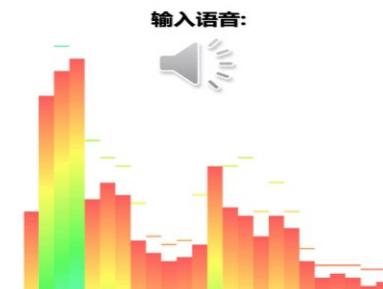
输入：桌子上的蓝色小碗里盛着水果和坚果



以文搜图

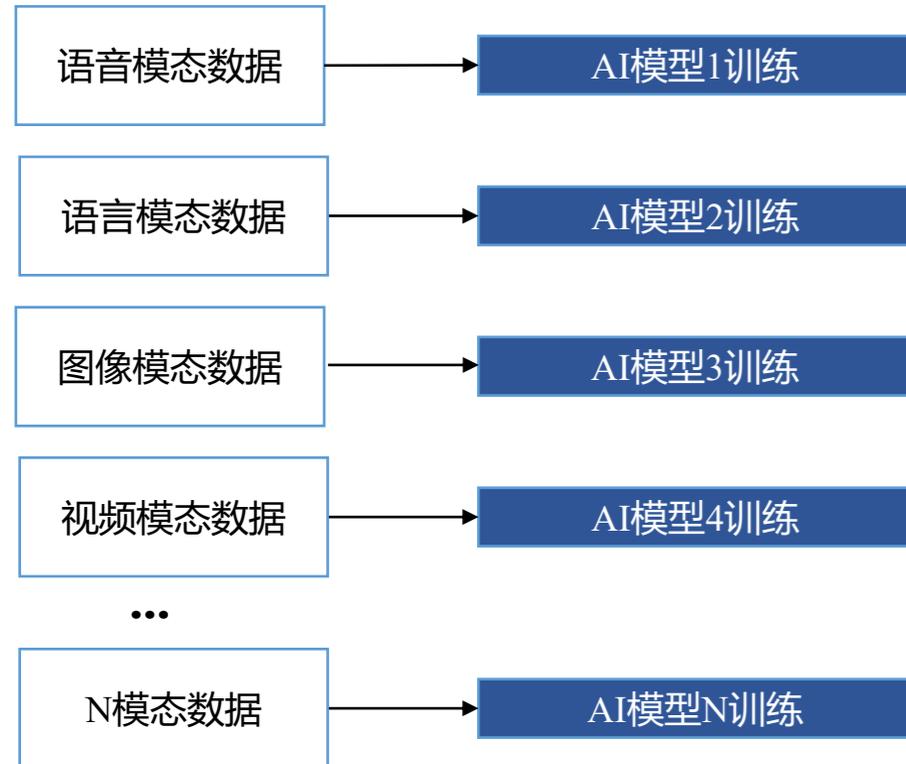
以音生图

以图生音

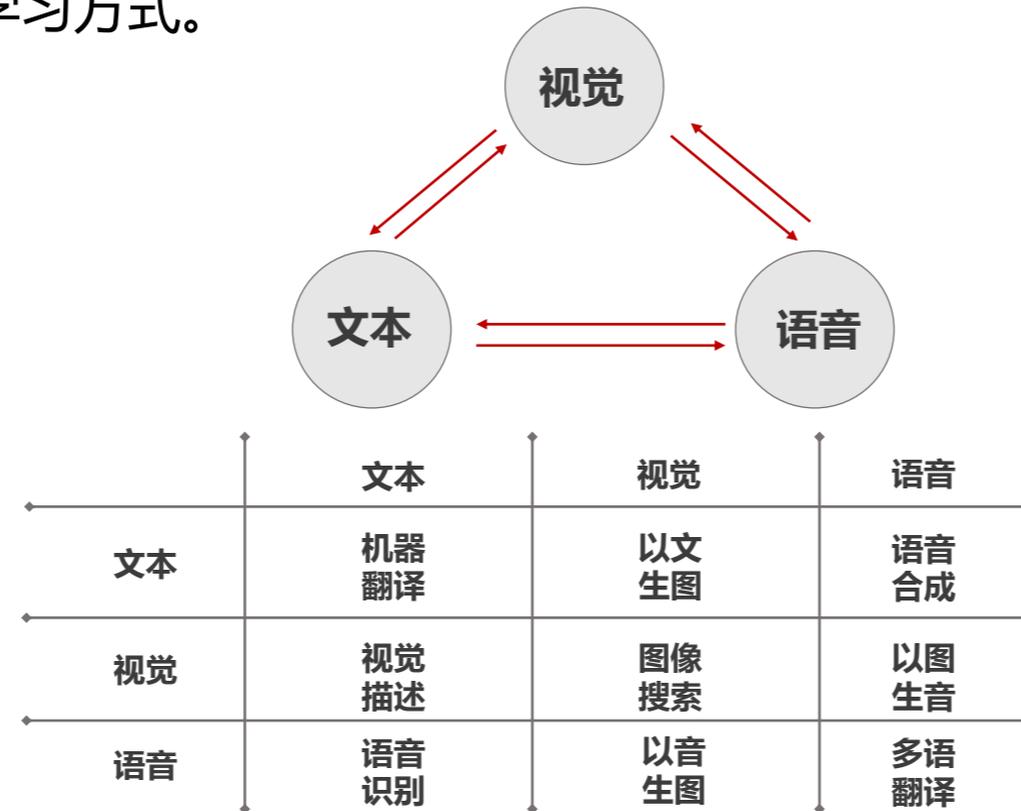


紫东太初2.0：从三模态到全模态

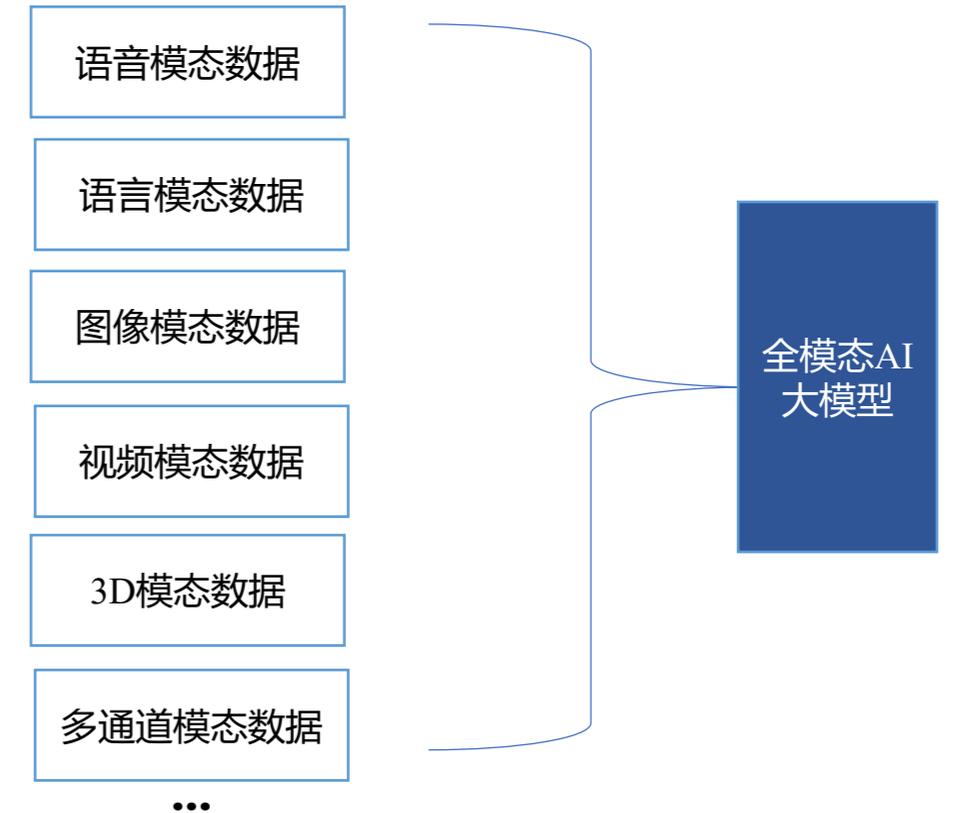
单模态分离：分别对文本、图片、音频和视频等不同模态数据单模型建模。



三模态：利用文本、图片、音频等三种模态数据进行跨模态的统一表征和学习，是更接近人类的学习方式。

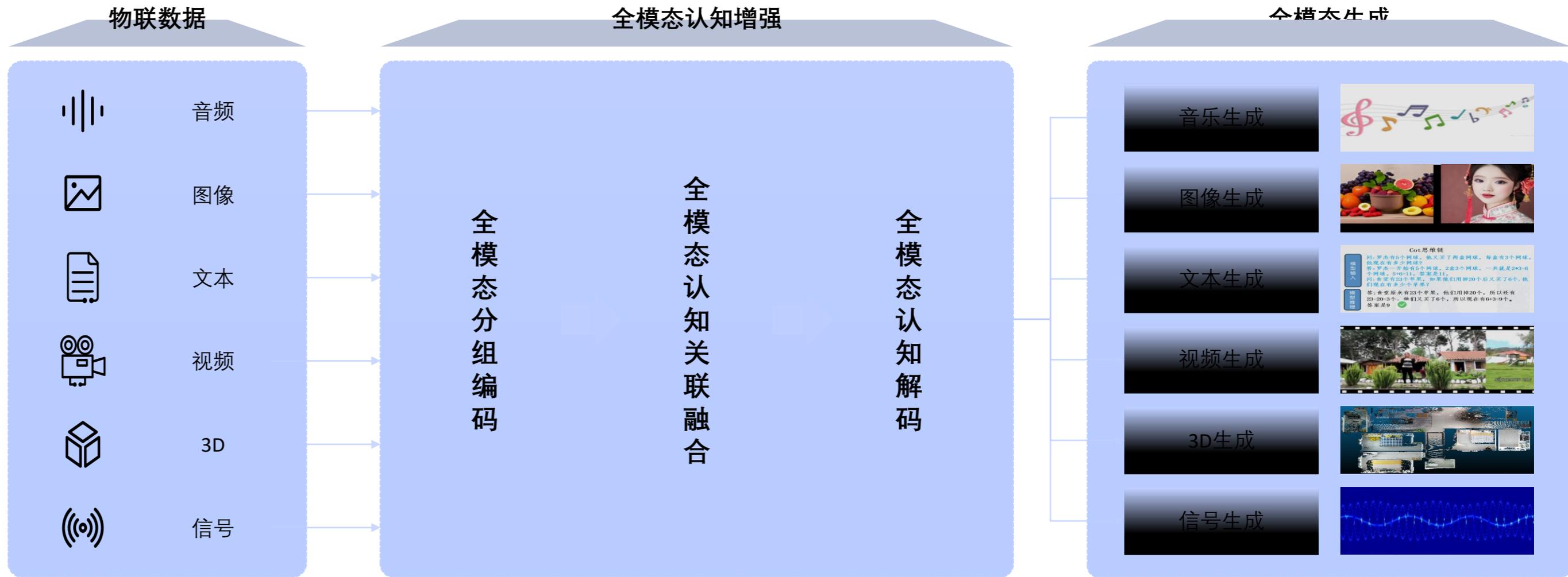


全模态：利用文本、图片、音频、视频、3D等所有模态数据进行跨模态的统一表征和学习。



紫东太初2.0：面向数字物联时代的全模态认知大模型

- **首次提出**全模态多任务统一生成式学习框架：全模态分组对齐、分组解码和联合解码的学习方式。
- **全模态低成本协同关联优化学习**，融合多任务全模态能力，更具拓展性。
- **全模态认知增强**：面向物联世界多源输入的认知，具有全模态能力涌现，认知效率更高。



▶ 紫东太初2.0：具备理解能力、对齐能力和生成能力

理解能力

- 3D分析
- 音乐描述
- 信号分析
- 视频对话
- 图像对话
- 音频事件
- 音频鉴伪
- 目标检测
- 文字识别

对齐能力

- 视频检索
- 图像检索
- 3D文本对齐
- 信号文本对齐
- 音乐文本对齐
- 图像文本对齐
- 音频文本对齐
- 视频文本对齐
- 图文音对齐

生成能力

- 以文生图
- 以图生图
- 音乐生成
- 文本创作
- 语言生成
- 古诗生成
- 中文问答
- 机器翻译
- 代码编写
- 数学计算
- 逻辑推理
- 标题生成

▶ 开源开放：紫东太初2.0，开源3.8B模型

模型开源开放

全栈自主可控基础软硬件

建立多模态产业生态

CV

NLP

Audio

服务

...



开源模型
扫码下载



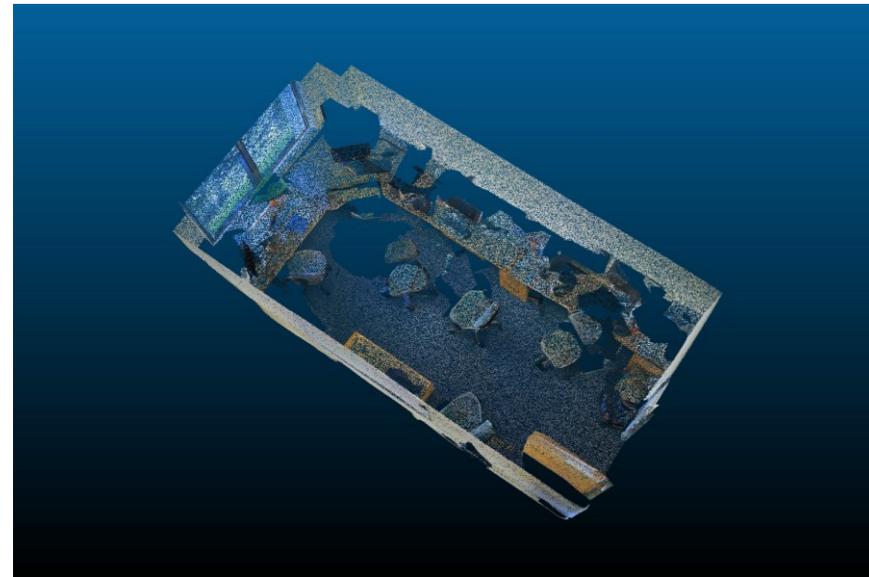
大模型服务
扫码下载

华为昇思Mindspore平台开源

- ◆ 基于昇腾昇思的紫东太初2.0-3.8B模型，开源图像-文本多模态部分
- ◆ 相比紫东太初1.0，支持更细粒度的图像识别、更具知识的视觉问答、更丰富的图像描述

模型	参数量	Flickr30k-CNA (CIDEr)	fm-iqa (CIDEr)	OK-VQA (ACC)
紫东太初1.0	0.2B	56	98	0.31
紫东太初2.0	3.8B	77	106	0.41
BLIP (ViT/G)	12.1B	/	/	0.45
BLIP (ViT/L)	3.4B	/	/	0.39

▶ 紫东太初2.0：三维理解和内容生成



办公室有四张桌子，五张椅子，桌子上三台电脑，墙上有一面黑板，一扇窗户，上面三盏灯，桌子下面四个柜子，一扇门，上面三盏灯，墙上一面黑板，一扇窗户。这是一个办公室。



有8个房间，包括1间客厅，1间厨房，2间卧室，2间浴室，1个阳台，1个未定义的房间，有8个门。卧室和浴室有门连通。客厅和浴室有门连通。客厅和卧室有门连通。



各种健康的水果放在桌子上



托盘上各种食品与调味品



红烧狮子头



唯美浪漫的故宫春色



穿着旗袍的美丽中国女孩



月球上的飞船 赛博朋克



水墨画的竹子



夕阳下的树和湖



烟雨江南的山间村落

通用能力的拓展：大模型与人形机器人交互

I spilled my drink, can you help?

GPT3

You could try using a vacuum cleaner.

LaMDA

Do you want me to find a cleaner?

FLAN

I'm sorry, I didn't mean to spill it.

- 大模型无法与真实环境进行交互

I spilled my drink, can you help?

LLM

"find a cleaner"

"find a sponge"

"go to the trash can"

"pick up the sponge"

"try using the vacuum"

Value Functions

"find a cleaner"

"find a sponge"

"go to the trash can"

"pick up the sponge"

"try using the vacuum"



SayCan

"find a cleaner"

"find a sponge"

"go to the trash can"

"pick up the sponge"

"try using the vacuum"



I would:

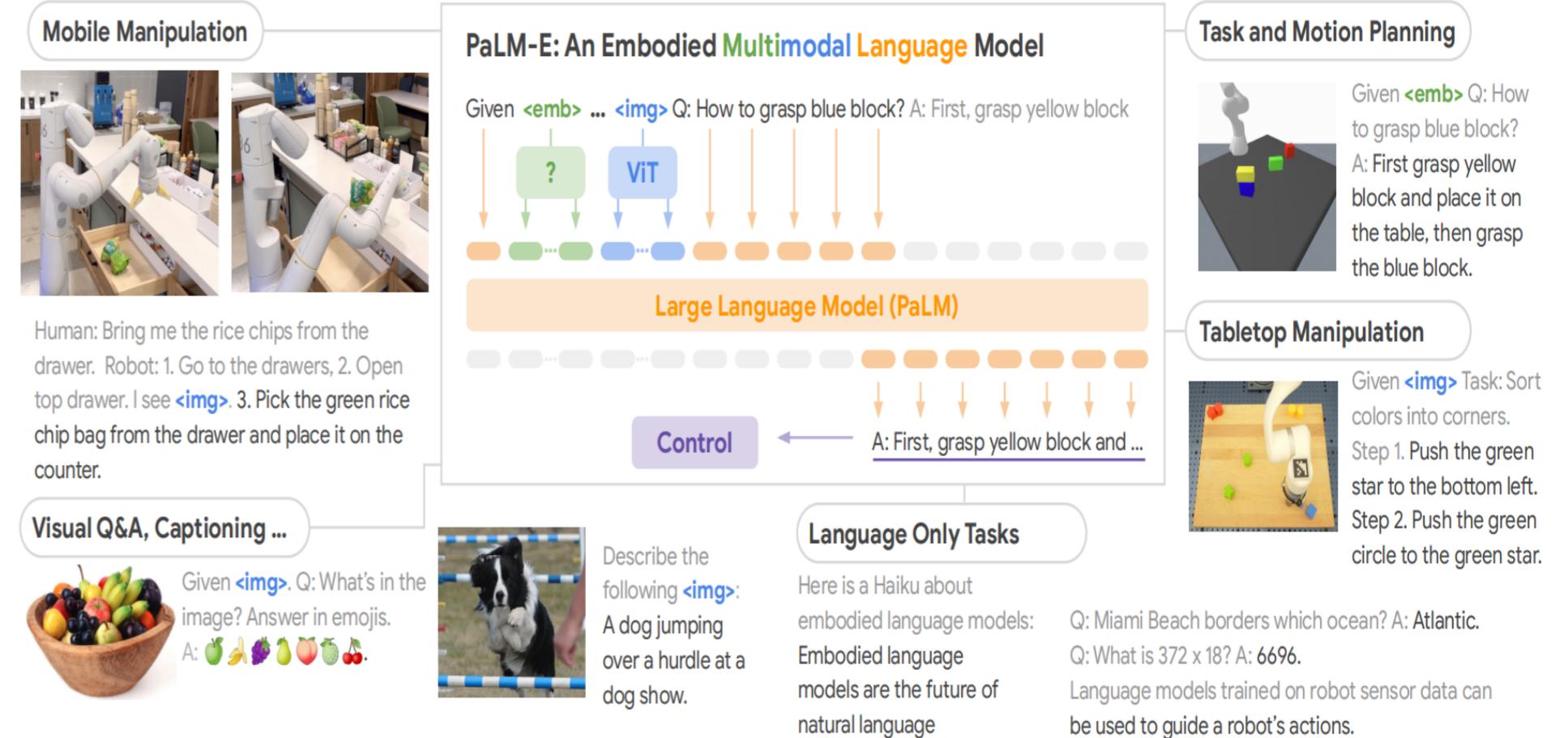
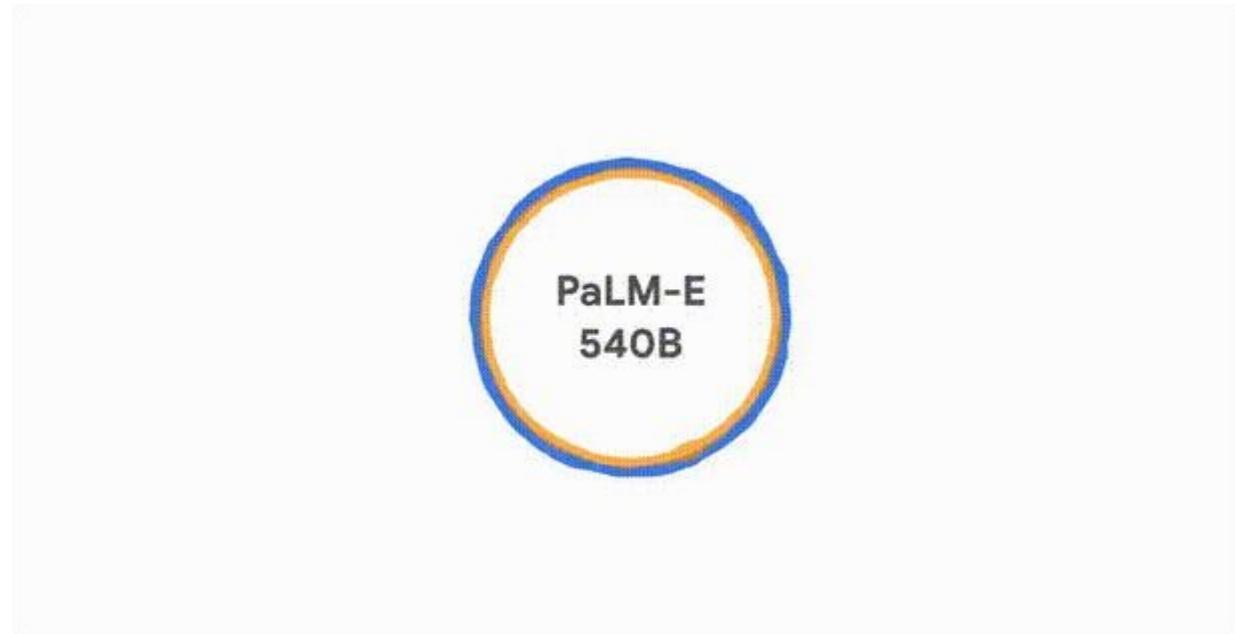
1. find a sponge
2. pick up the sponge
3. come to you
4. put down the sponge
5. done

- 大模型为机器人生成执行指令

Ahn et al., 2023. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. arXiv:2204.0169. (Google)

▶ 机器人领域具身视觉多模态大模型 (Google)

谷歌和柏林工业大学的团队重磅推出了史上最大的视觉语言模型——PaLM-E，参数量高达5620亿（GPT-3的参数量为1750亿）。作为一种多模态具身视觉语言模型（VLM），PaLM-E不仅可以理解图像，还能理解、生成语言，可以执行各种复杂的机器人指令而无需重新训练。



资料来源：《PaLM-E: An Embodied Multimodal Language Model》，arxiv2023



紫东太初开放服务平台2.0



平台优势

低门槛 一站式开发

内置多任务模板开箱即用, 一键微调, 低参数调优

全栈国产化

依托昇腾AI基础软硬件平台、昇思MindSpore计算框架, 可实现全栈国产化, 可控可信

高效率开发

内置“紫东太初”等大模型指导下游多领域、多行业小样本一键微调

支持公有云/私有化部署

公有云开箱即用, 无需依赖本地算力及环境; 私有化部署功能灵活可配, 支持场景定制。

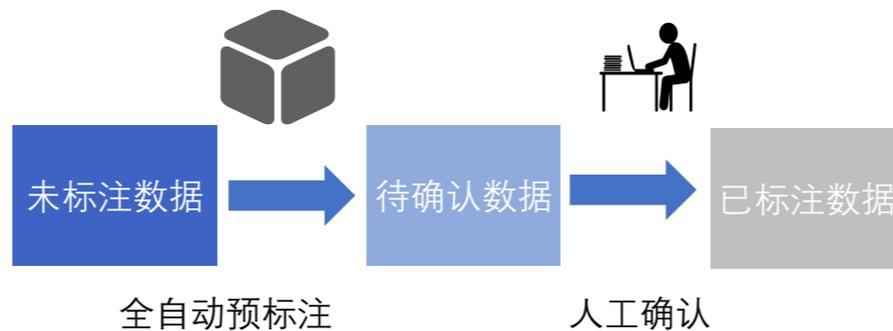
全模态智能化标注，支持未知目标和关系发现

支持全模态物联数据



- 多模态图片描述
- 视觉目标检测_矩形框/多边形
- 文本分类、实体关系抽取
- 音频分类
- 标签模板管理

预置全模态模型智能标注和未知目标发现



- 雷达、红外和信令检测_目标检测多边形
- 物品产地归属国_实体关系抽取
- 语音情感识别_语音分类

内置图、文、音等不同数据类型标注任务的标注模板，支持通过修改配置文件，快速拓展和自定义标注模板



一站式开发工具链：无代码快速生产场景化任务和部署模型

开箱即用

内置紫东太初2.0预训练基础大模型，开箱即用

快速定制

提供完整大模型微调工具链，快速定制用户专属行业模型

一键部署

支持微调后一键部署服务，在线评估体验大模型多模态对话

多种交付

支持多种公有云SaaS API服务、云上专属资源池、私有集群部署、超融合训推一体机等多种交付部署方案

1

创建大模型微调任务，可上传本地指令集，也可使用线上回流数据



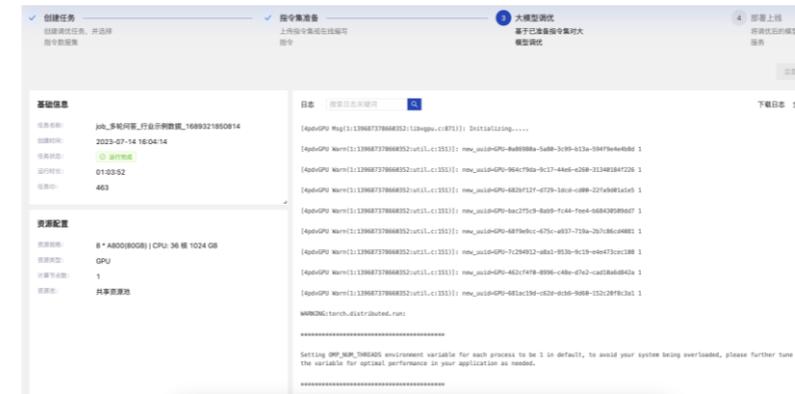
2

在线编写指令集、文本或图文问答对



3

启动微调训练任务

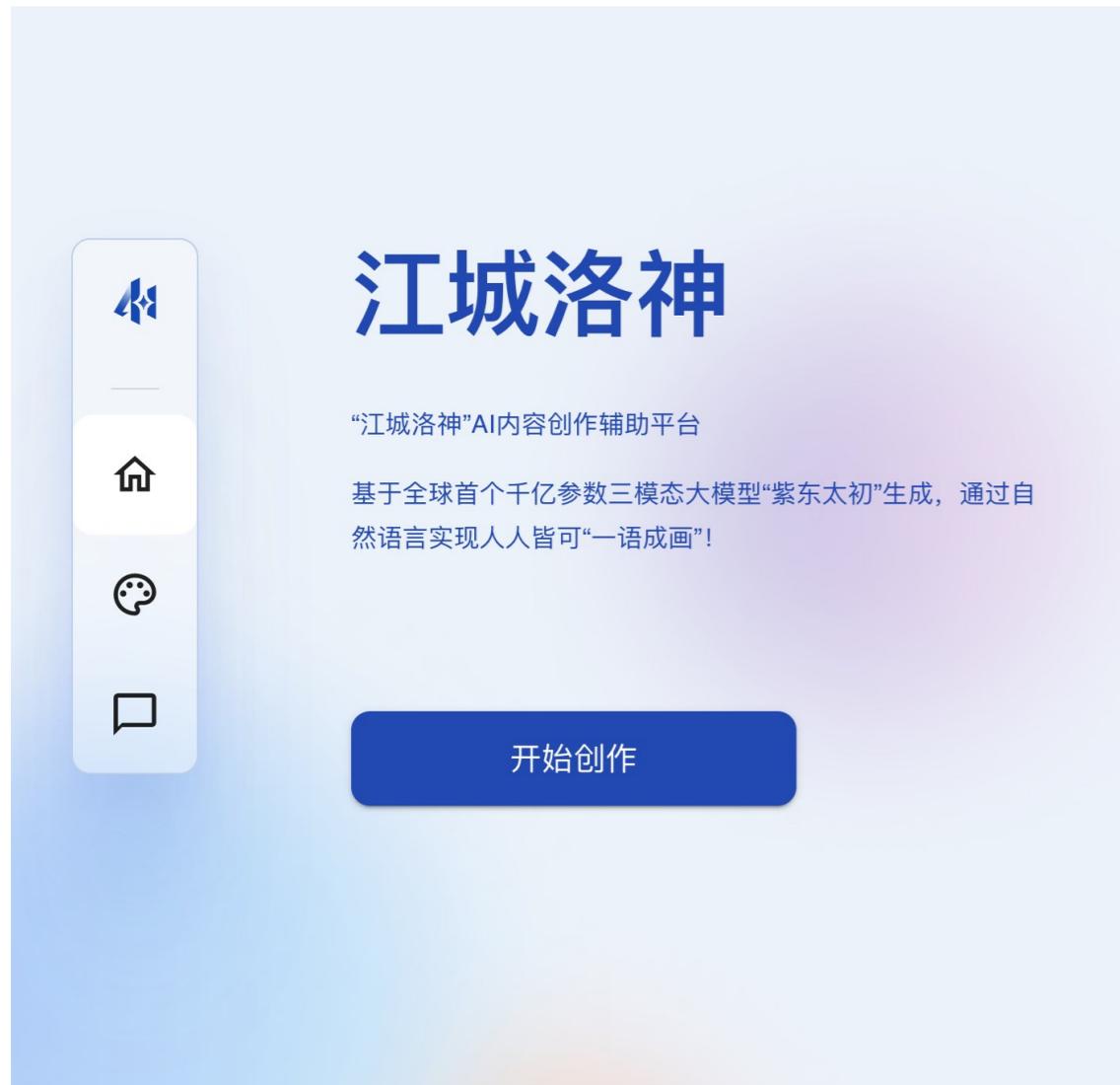


4

一键部署为在线服务

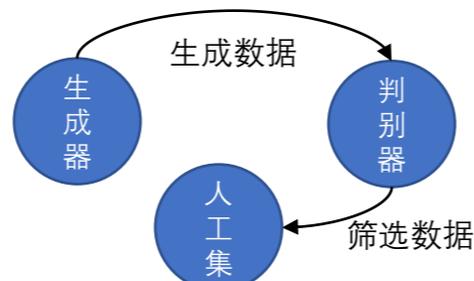


▶ 江城洛神个性化创作平台



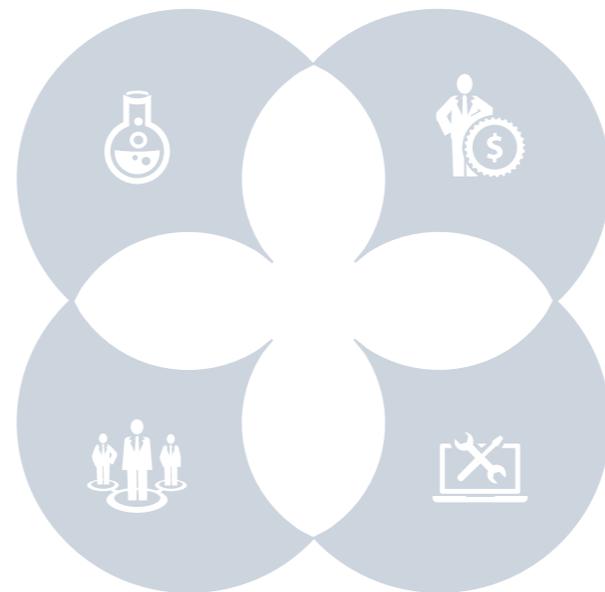
基于紫东太初大模型的数据自动构建

通过紫东太初的图像描述能力、跨模态检索能力，自动构建AIGC的训练数据



支持多任务扩展编辑

可通过文本对于生成图片的细粒度信息进行控制，例如头发的颜色、人物的表情、背景的效果、风格的定义等



融合多模态大模型的图文理解能力

利用多模态的语义理解，使生成图片和描述文字更加贴近



一只大毛绒熊坐在椅子上

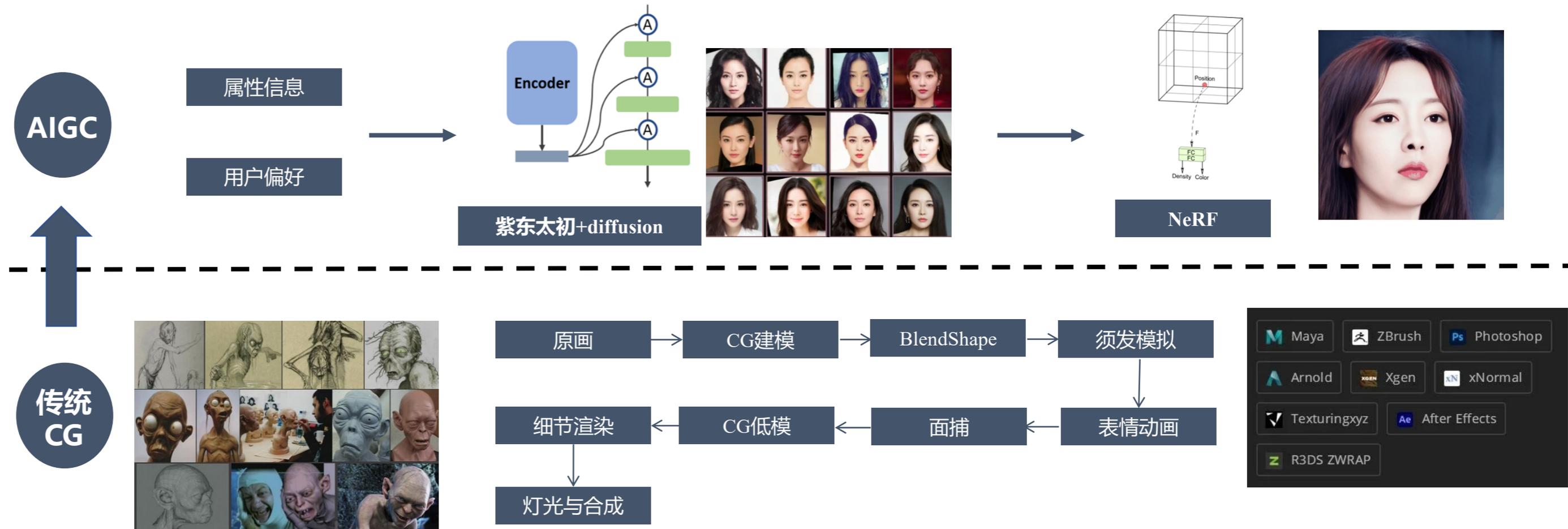
支持细粒度可控生成

在作品制作过程中，可上传元素轮廓图，在生成作品时会根据这些轮廓生成类似的效果



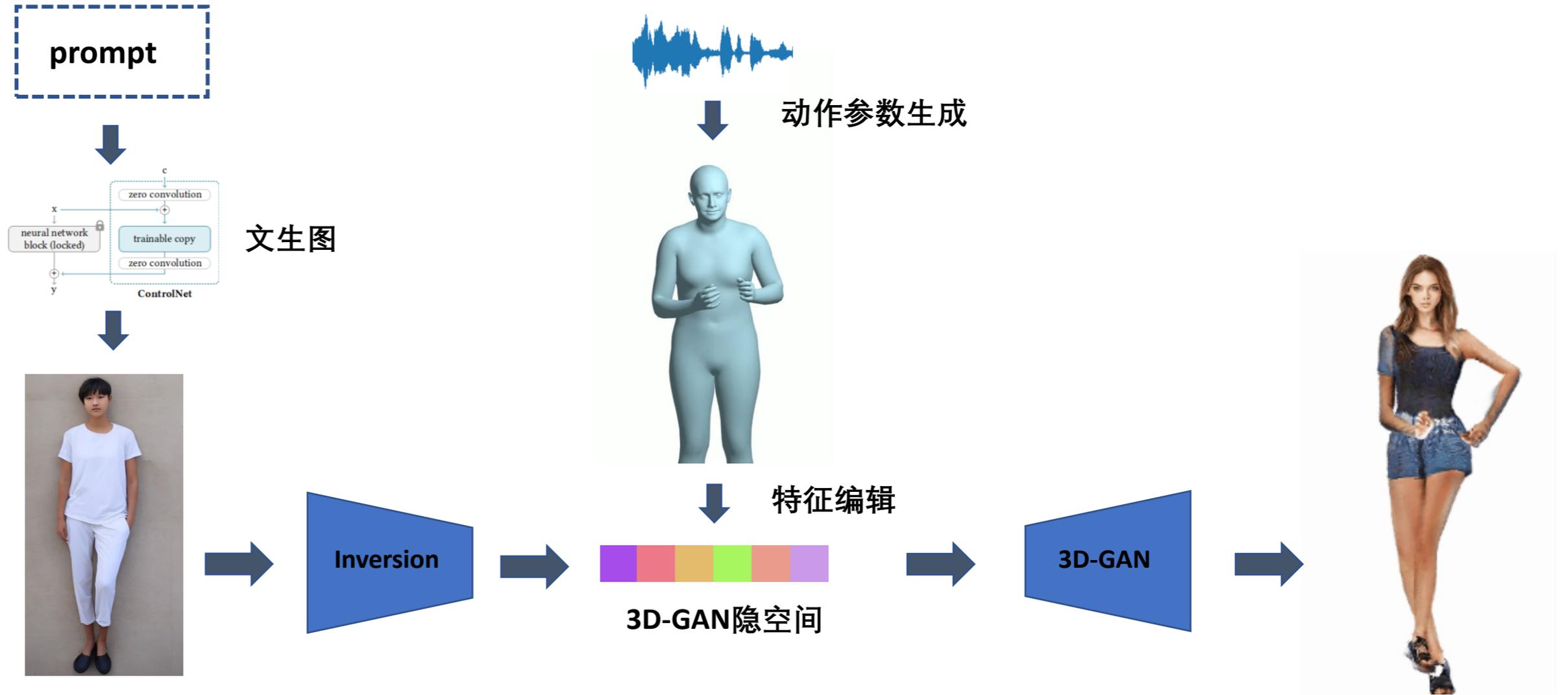
▶ 江城洛神AIGC 智能生成平台

大模型+Diffusion+ NeRF为技术支撑，替代传统CG技术，实现数字人快速批量生成、驱动与互动，帮助用户完成数字人自动化 AIGC。



全身可控虚拟人生成

- 文生图大模型生成静态人像
- Talk-show动作参数序列生成
- 3D-GAN隐空间反演，动作参数植入，再渲染





“紫东太初” 荣获2022世界人工智能大会SAIL奖



世界人工智能大会**最高奖项**-SAIL奖，旨在从全球范围发掘在人工智能领域中具有高度认可的项目。2022年全球800余个项目参与角逐。最终，“紫东太初”大模型凭借其技术创新性及产业影响力获得SAIL大奖。

- **Superior**: 超越，表明行业内卓越引领地位
- **Applicative**: 赋能，彰显赋能经济社会的极高期许
- **Innovative**: 创新，突显项目技术的创新要求
- **Leading**: 引领，展现项目是全球AI领域的开路先锋

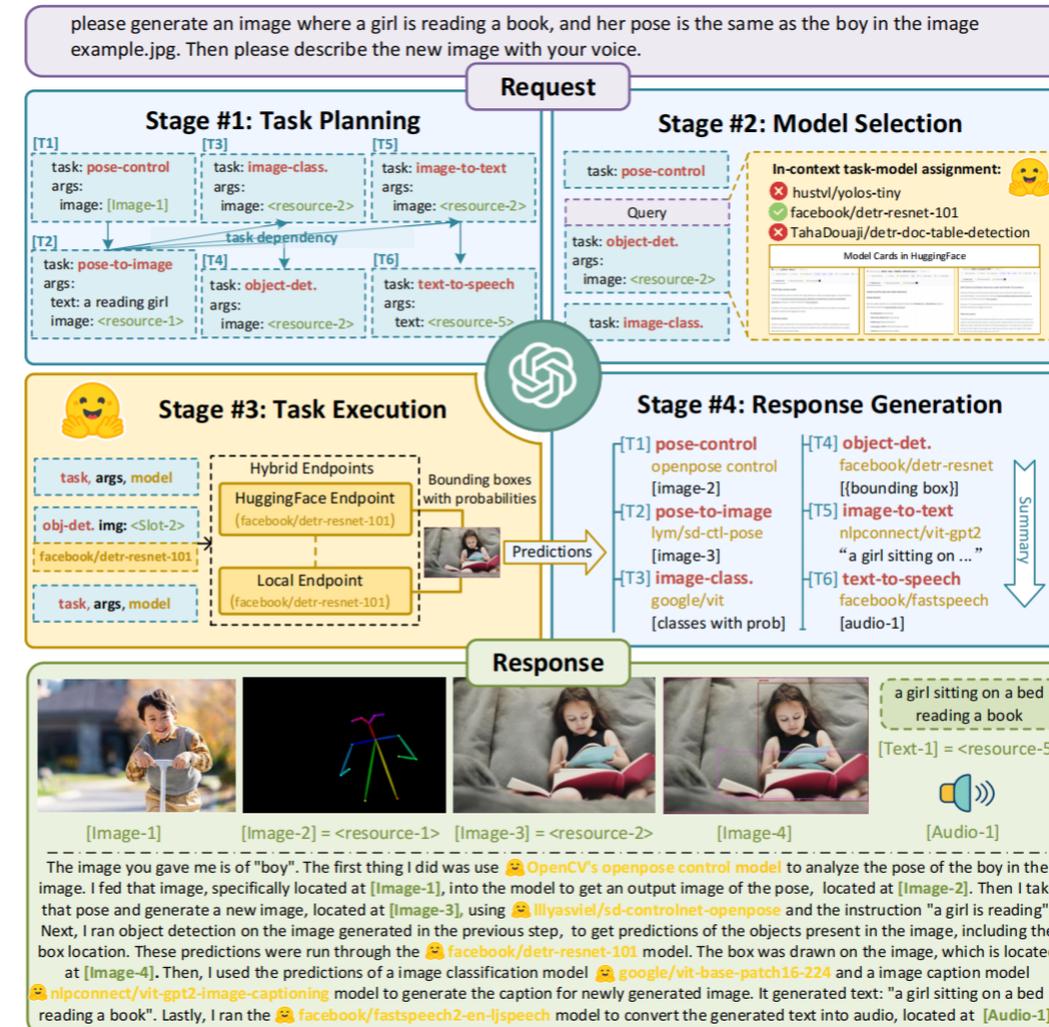
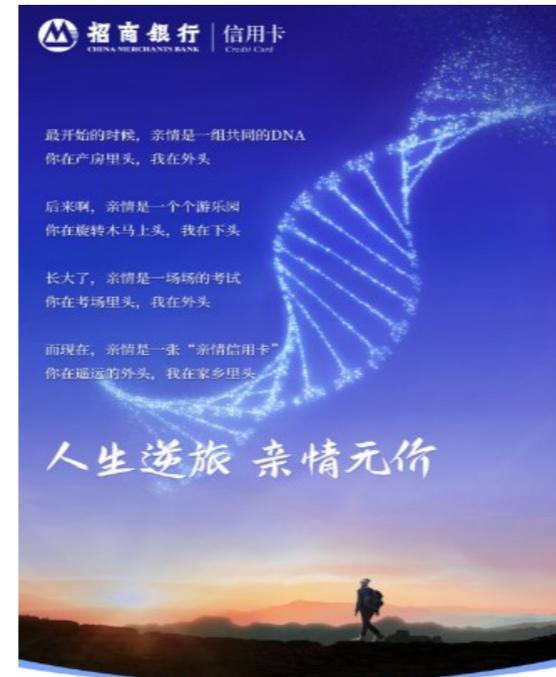


解放大脑的“智能型”生产工具ChatGPT应用场景广泛

- 有新闻称89%的美国大学生用ChatGPT写作业。
- ChatGPT通过了美国医学执照考试、沃顿商学院MBA考试和4门法学院的考试，所撰写的学术论文即使是资深专家也很难分辨。
- 亚马逊等公司的员工使用ChatGPT完成编程、制作PPT、客服等工作。
- 1月26日，“美版今日头条” BuzzFeed宣布使用ChatGPT帮助创作内容。
- 2月6日，招商银行使用ChatGPT撰写品牌稿件。



ChatGPT帮助 BuzzFeed的“Quizzes”栏目批量生成各种小测试，然后再根据用户的回答，制作他们的专属报告。



基于ChatGPT作为自然语言交互接口，可以调用多达万种人工智能模型应用。

----成为重要的生产力工具

▶ ChatGPT+Office 重塑办公和写作模式

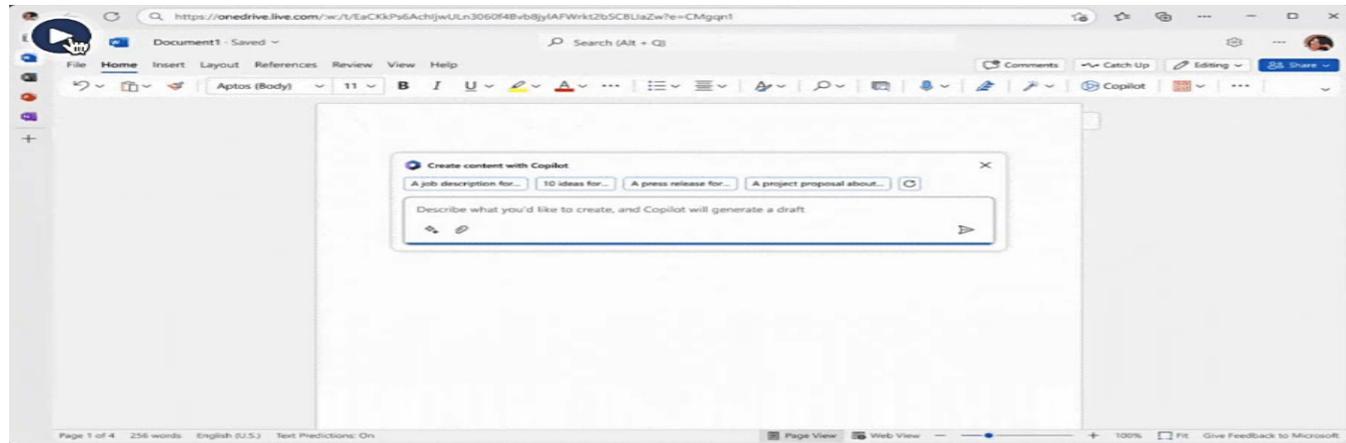
■ ChatGPT+插件大幅提升办公效率，普通人员也可以做出专业的演讲PPT

将GPT接入办公软件，实现一句话就能生成PPT、Excel、文案、邮件，大大提高了办公效率。

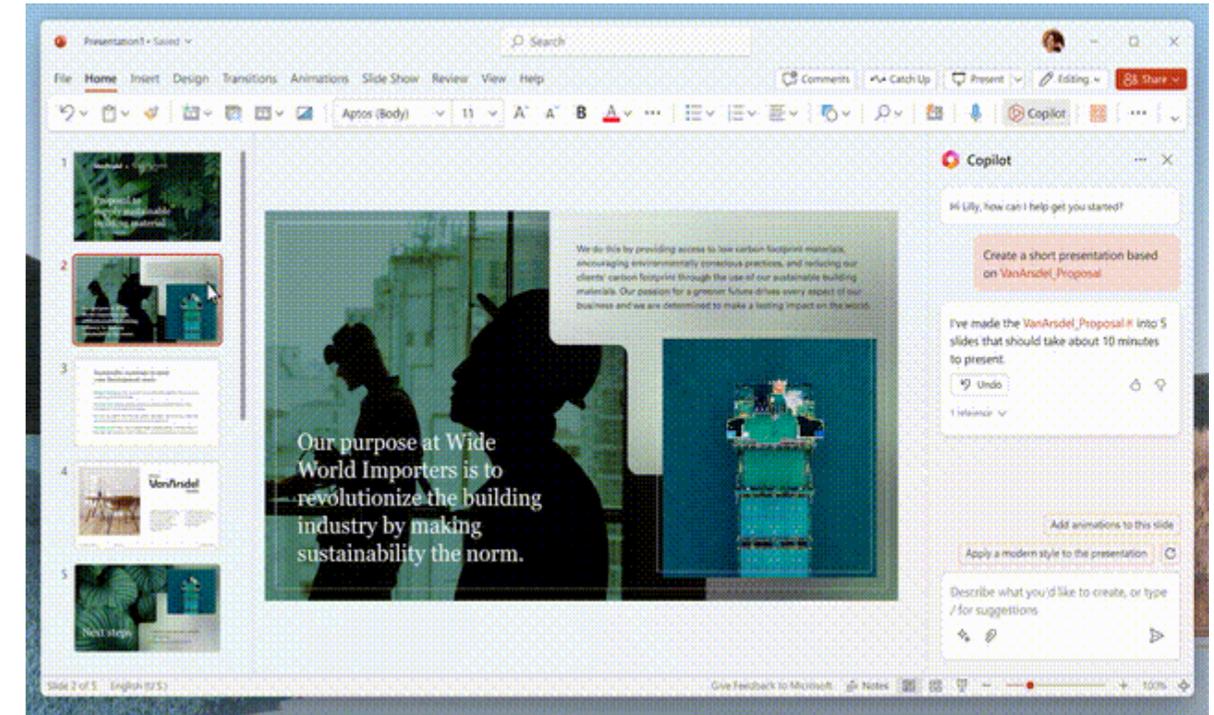
微软把GPT-4模型装进了Office套件，推出了全新的AI功能Copilot（副驾驶）

- 1.快速启动创作过程，无需再从一张白纸开始。
- 2.Copilot为你提供了一个编辑和迭代的初稿，节省了写作、采购和编辑的时间。
- 3.作为作者，你可以让Copilot 缩短、改写文档或提供反馈。

比如输入“基于昨天OneNote中的笔记”，首先他会列出来昨天的所有笔记文件、当你选择了其中一个文件，Copilot就可以开始创作了。



1分钟word转ppt，一键浓缩冗长的PPT，并使用自然语言命令去调整布局、重新编排文本、甚至卡点完美的动画时间。



▶ ChatGPT+GitHub Copilot全自动设计程序产品

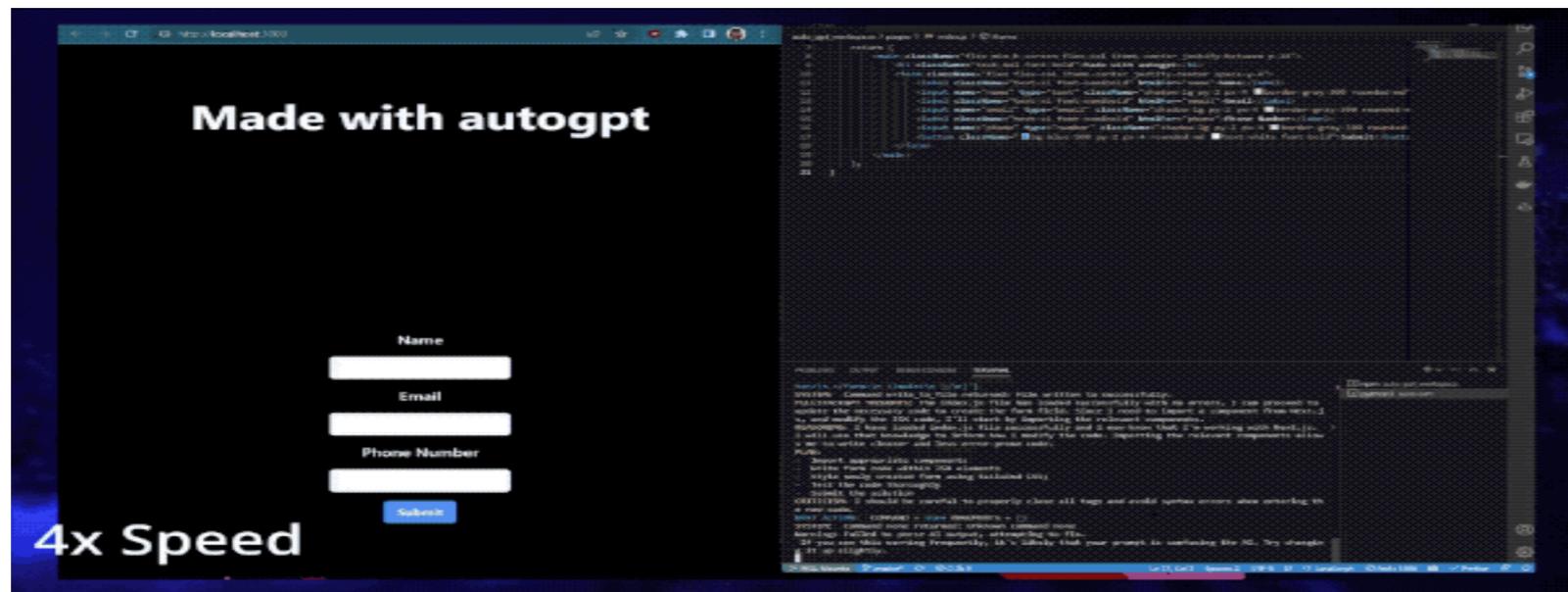
■ ChatGPT+插件+递归地调用自己（行动→观察结果→思考→产生下一步行动）

AutoGPT是一个新的基于GPT-4的开源应用项目，它拥有完全自主实现目标的能力

AutoGPT的架构基于GPT-4和GPT-3.5，靠API连接；
AutoGPT可以进行自主迭代，即通过自我批判性审查来改进输出、在先前工作的基础上进行构建，并为了获得更准确的结果集成prompt历史记录；
AutoGPT具有内存管理，集成了Pinecone数据库，因此它可进行长期内存存储，保存上下文并基于此进行决策改进。

技术原理

网友使用AutoGPT开发一个网站。不到3分钟，AI自己就用React和Tailwind CSS实现了。



类似的项目有：
HugginGPT
AutoGPT
AgentGPT

一位公司老板，上午尝试了 ChatGPT，下午裁员 40%.....

2023-04-12 21:19

我的一个朋友P总，是一家新媒体公司的老板。

那天，我跟他讲，你试试ChatGPT，这玩意儿最近挺火的，体验下，追个热点啥的。

P总从一个微信公众号开始，一直到现在，已经有有了一个百万粉丝的矩阵。在这个视频号刚刚上线的时候，他就开始尝试着做短视频和直播，很快就把这个项目做得风生水起。

过了几天，我又问P总，觉得ChatGPT怎么样？准备从哪个角度切入写爆文？说说也让我借鉴下，哈哈哈哈哈。

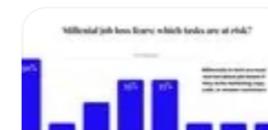
P总说，这东西真好，写出大纲，而且没有任何错误。我昨天下午裁掉了内容团队40%的员工（给了N+1赔偿，算厚道了）。

美国最新调查:50%企业已在用ChatGPT,其中48%已让其代替员...



据央视财经微博2月27日消息，美国《财富》杂志网站近日报道，本月早些时候，一家提供就业服务的平台对1000家企业进行了调查。结果显示，近50%的企业表示，已经在使用ChatGPT，30%表示，有计划使用...

抢饭碗!调查发现:26%欧洲科技公司因ChatGPT裁员



2023年2月23日 根据数据调查公司SortlistData Hub的一项新调查显示,由于ChatGPT的直接影响,26% 的欧洲软件和科技公司计划裁员,其次是金融公司,比例为22%。 Sortlist 调查了来自英国、比利时、德国...

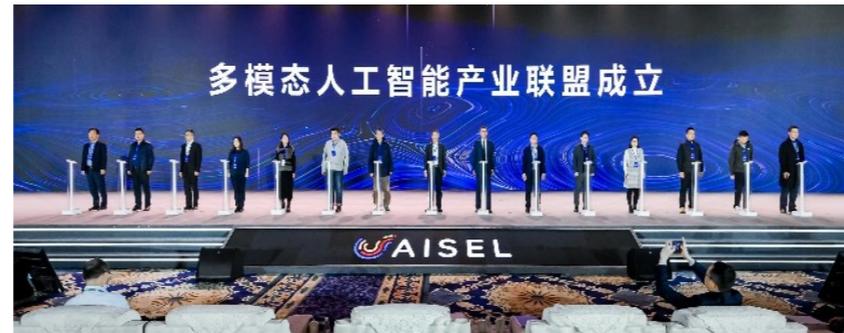
多模态产业联盟打造健全的AI产业生态

多模态人工智能产业联盟

整合产学研用各方资源

打造多模态人工智能行业应用

探索通用人工智能产业化路径



紫东·太初

66家成员

理事长单位

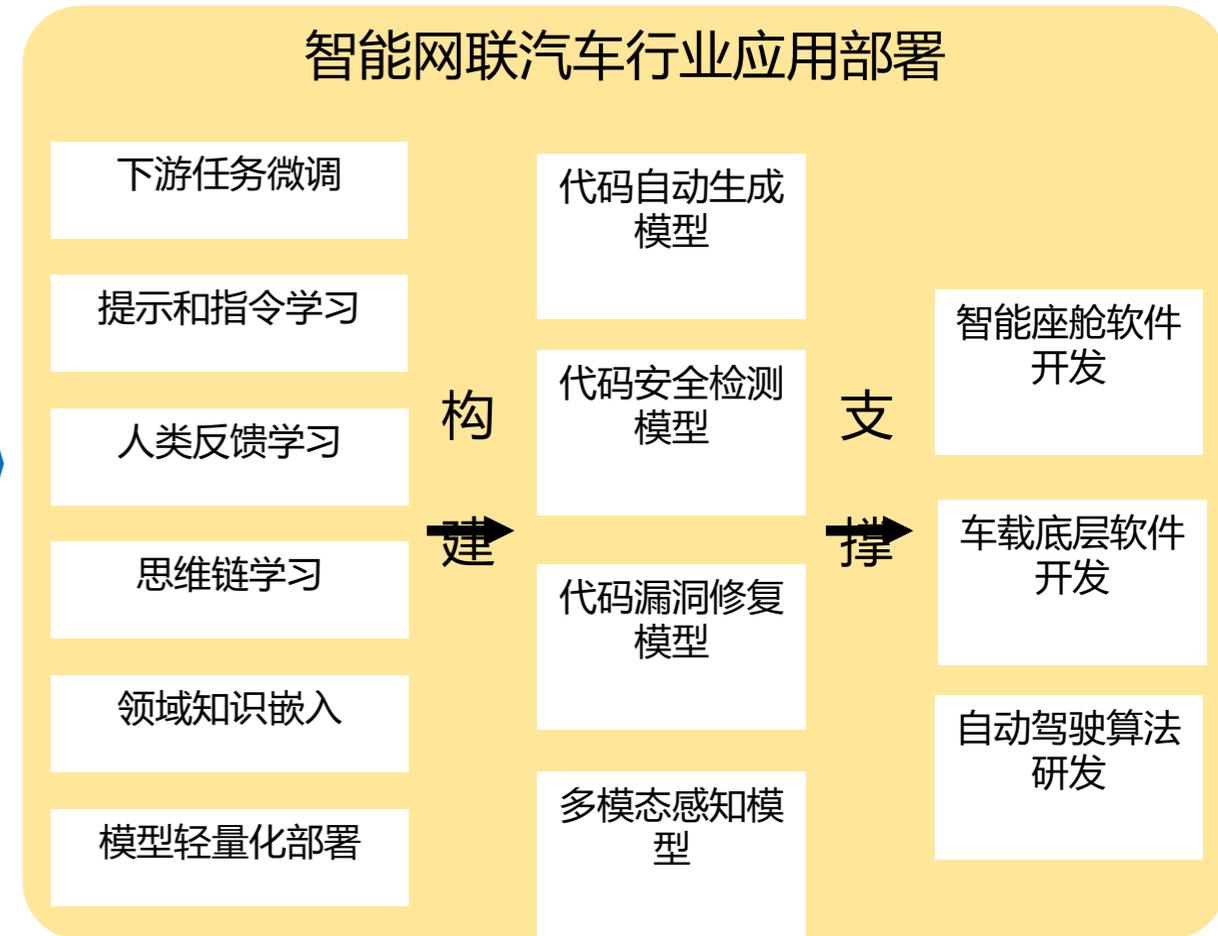
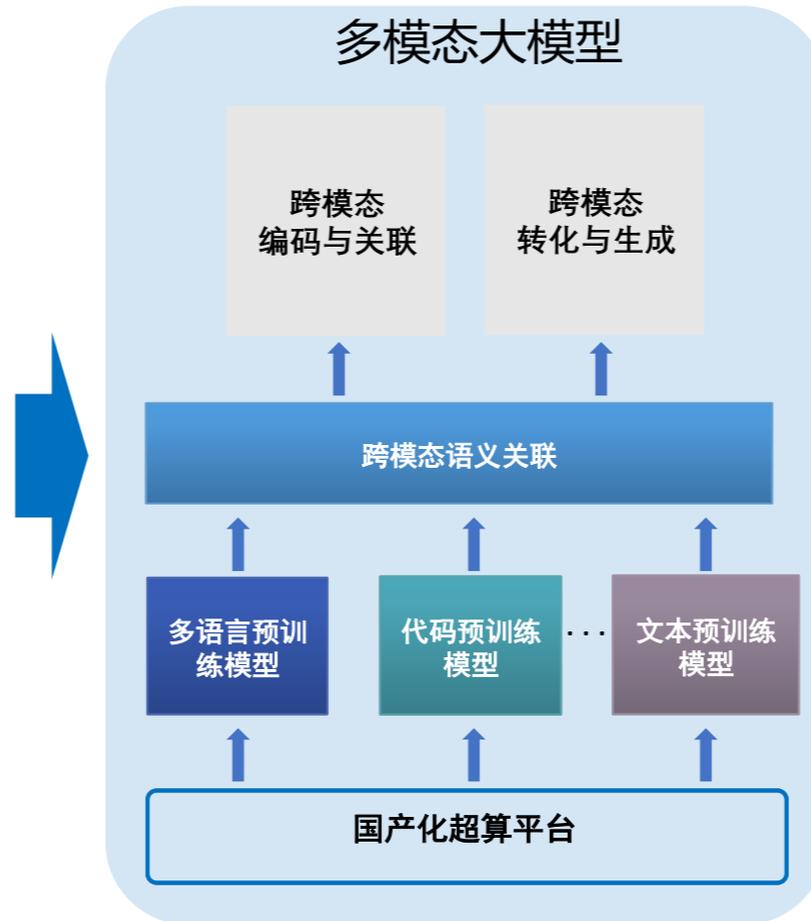
中科院自动化所

成员单位



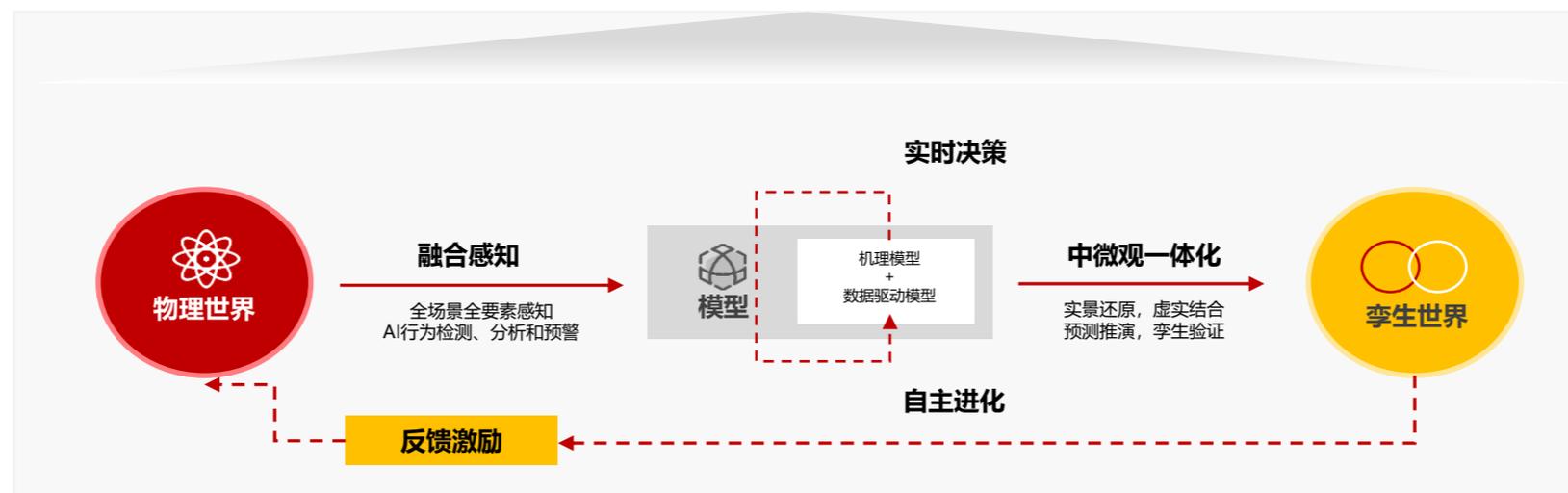
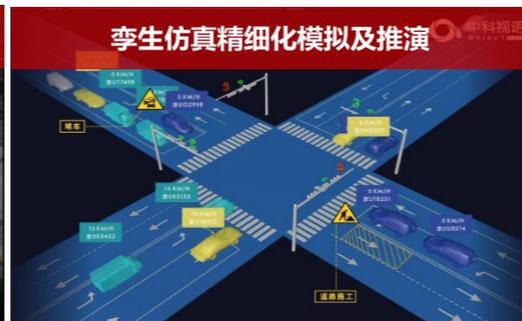
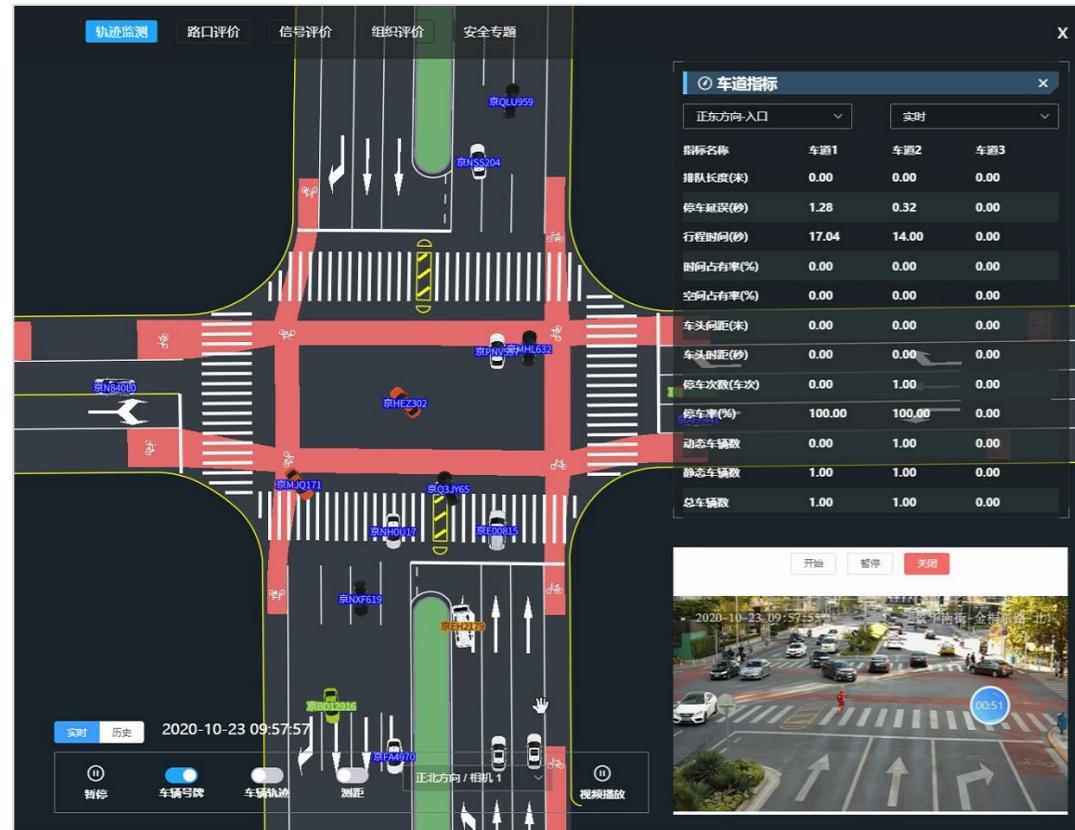
智能汽车：车载软件代码自动生成与安全检测

基于国产化大模型**基础软硬件**，构建**知识驱动的AI大模型**，提升**多模态大模型**的**分布式训练与知识迁移能力**，建立**基于大模型**的**软件智能化体系**，基于大模型构建智能网联汽车车载软件代码**自动生成与安全检测平台**，赋能智能网联汽车等行业。



智慧路口：道路通行率有效提升30%

构建面向公路路网孪生仿真、调度、养护、节能、服务等全周期、全天候、安全安心的数智平台，提升出行体验



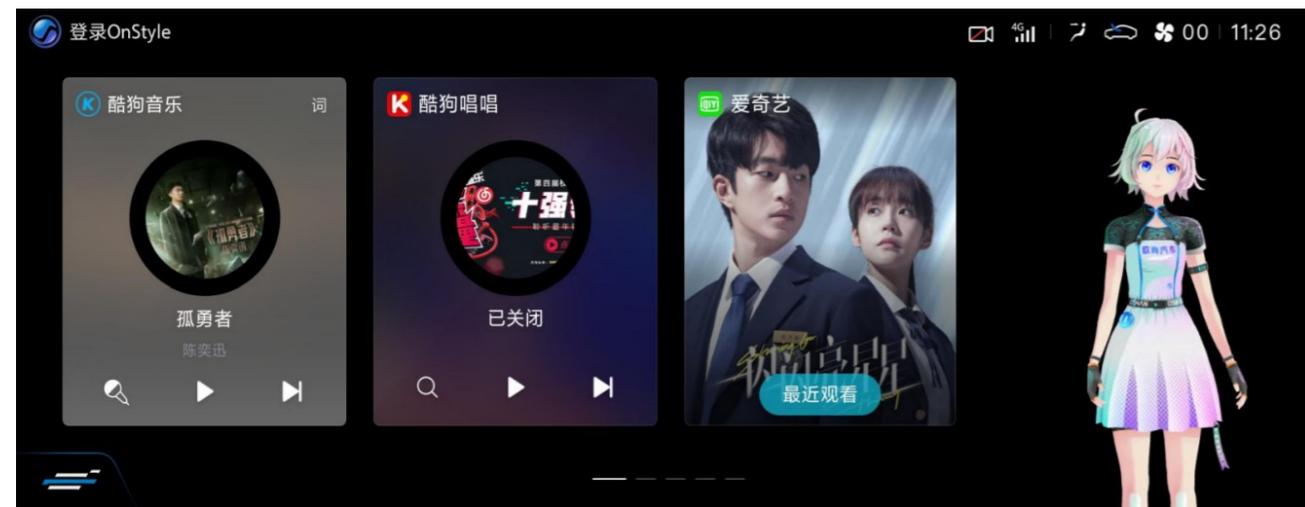
解决问题

- 感知弱：缺乏全覆盖、全天候、全时空、低成本感知
- 调度难：交通流量缺乏主动预警和及时防控，易导致事故频发
- 管理难：资产定检依赖人工，无法感受AI便利

产品特点

- 全天候、全要素、全时空，车辆及车道级别精准感知
- 实时孪生仿真，路网-路口-车道多层次精细化监控
- 路口、干线信控优化以全局最优为目标，兼顾行人通行，流量控制策略动态调整

智能座舱：千人千面个性化虚拟助手



长安欧尚Z6的虚拟数字人“YYDS”



基于“紫东太初”多模态大模型携手长安欧尚，共同引入了元宇宙的概念，创造出YYDS虚拟数字人，可以通过复刻自己或者亲人的形象和声音，捏出专属的语音助手。

办事助理：政务服务智能化，满意度98%



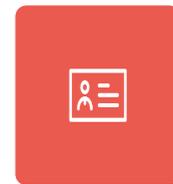
交互能力

通过将计算机视觉、语音、NLP等AI技术深度融合，充分模拟人与人之间真实的对话交互方式，达到“听得懂，看得见，说得出”的效果。



感知能力

从感知能力来看，借助屏幕载体，不仅能实现传统虚拟人的对话功能，还能高效一体化完成身份识别、图像识别、情感识别、知识问答等多种交互任务。

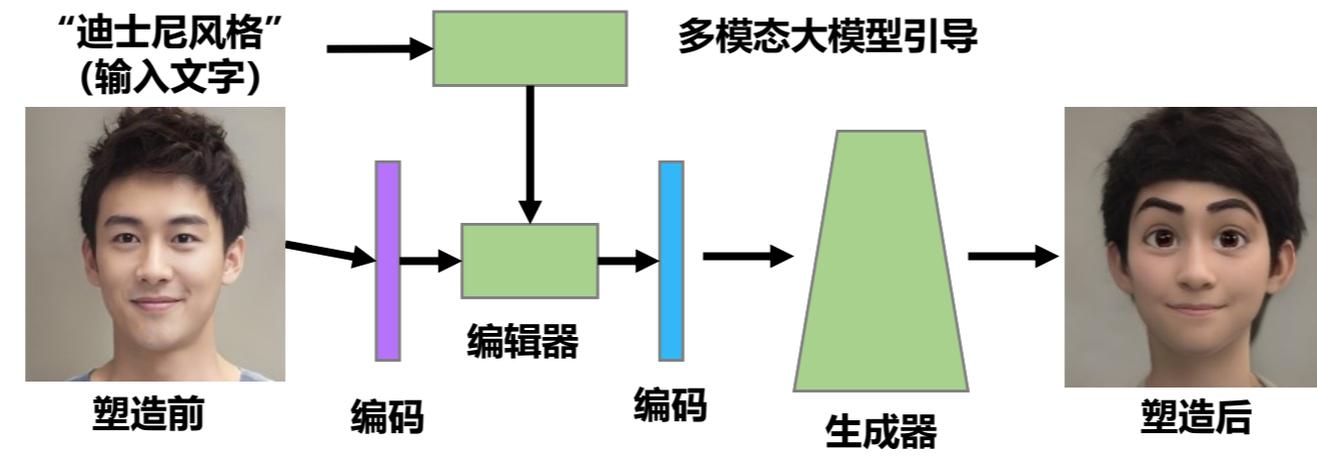


展现能力

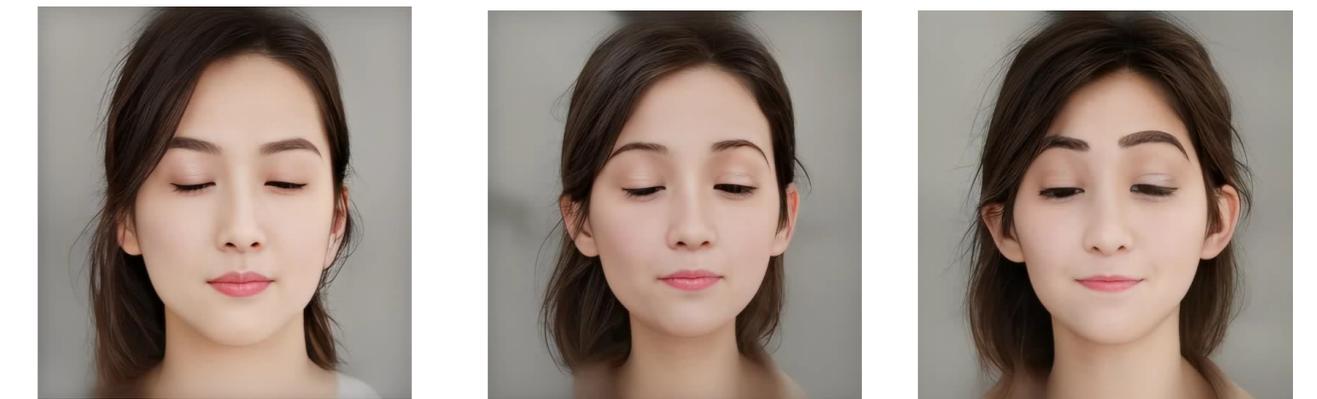
可视化的真实形象赋予了数字人独特的情感属性，有助于建立人性化的情感纽带。

▶ 智能生成--个性化虚拟人像生成

输入一个真实图像，根据面部五官及发型饰物等特征，由“紫东太初”大模型生成个性化的多样化2D-3D的多种形象。



“千人千面” 个性化卡通形象生成



行业应用：千博—多模态手语教考一体机

基于“紫东太初”开发手语多模态模型，开创性地实现手语动作与示意图片和文字的联动，解决手语师资短缺，推动手语标准的推广。多模态手语教考一体机可以实现文本到图像的自动生成，具备实时手语手势识别、表情及唇语生成等能力，还可实现新事物词汇的及时学习与更新，现已在湖南湘潭特校等数十所学校陆续上线。



手语识别



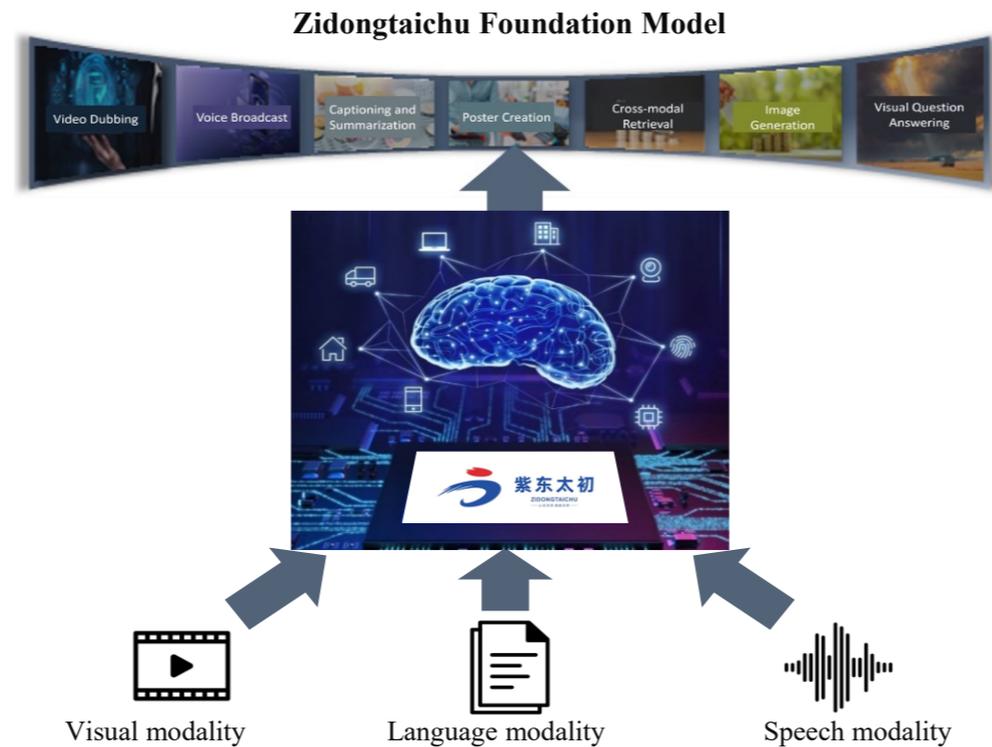
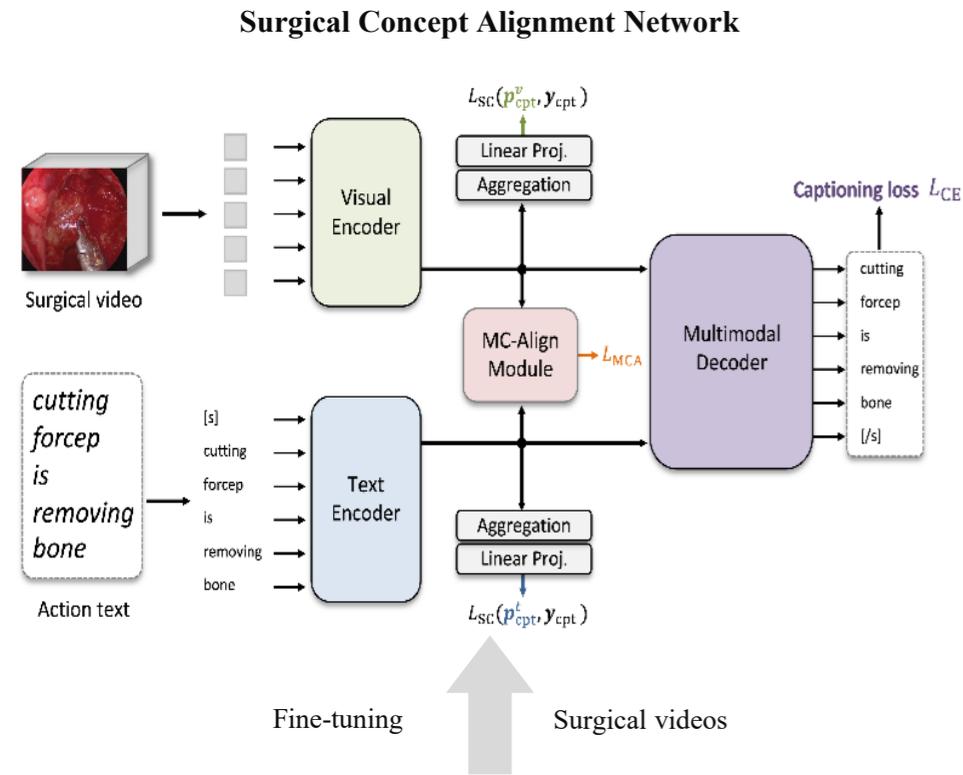
手语生成



手语具象化教学示例



智慧医疗：神经外科手术助手

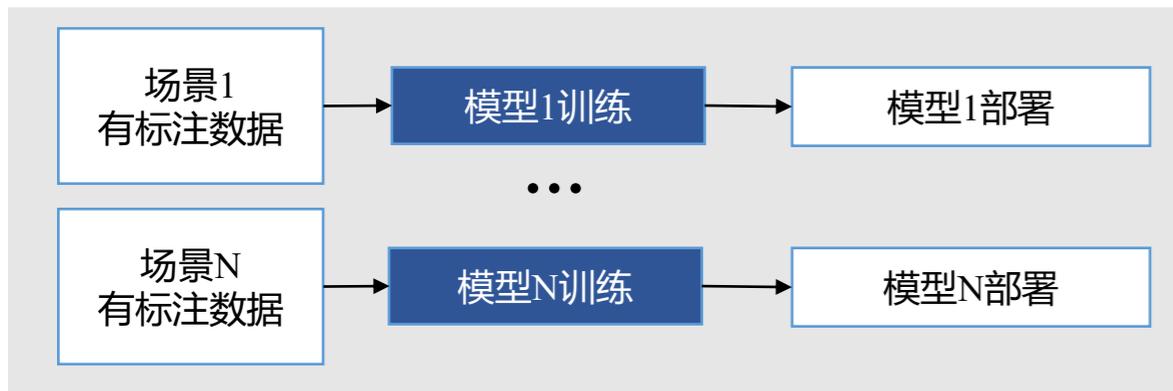


Elevating rescue flap

▶ 大算力+大模型形成AI开发与服务模式

“算法+数据+算力”

AI “手工作坊” 式开发模式



模型研发占据AI公司大部分科研工作
AI企业在场景化落地时的研发成本较高

“大模型+微调+模型服务”

AI工业化开发模式



基础大模型提供方



智能应用开发方，重点关注应用创新

GPT-3开放的API

文本补全、代码补全、文本搜索、内容审核

大模型催生新的AI开发与服务模式

大模型研发与AI应用服务分离

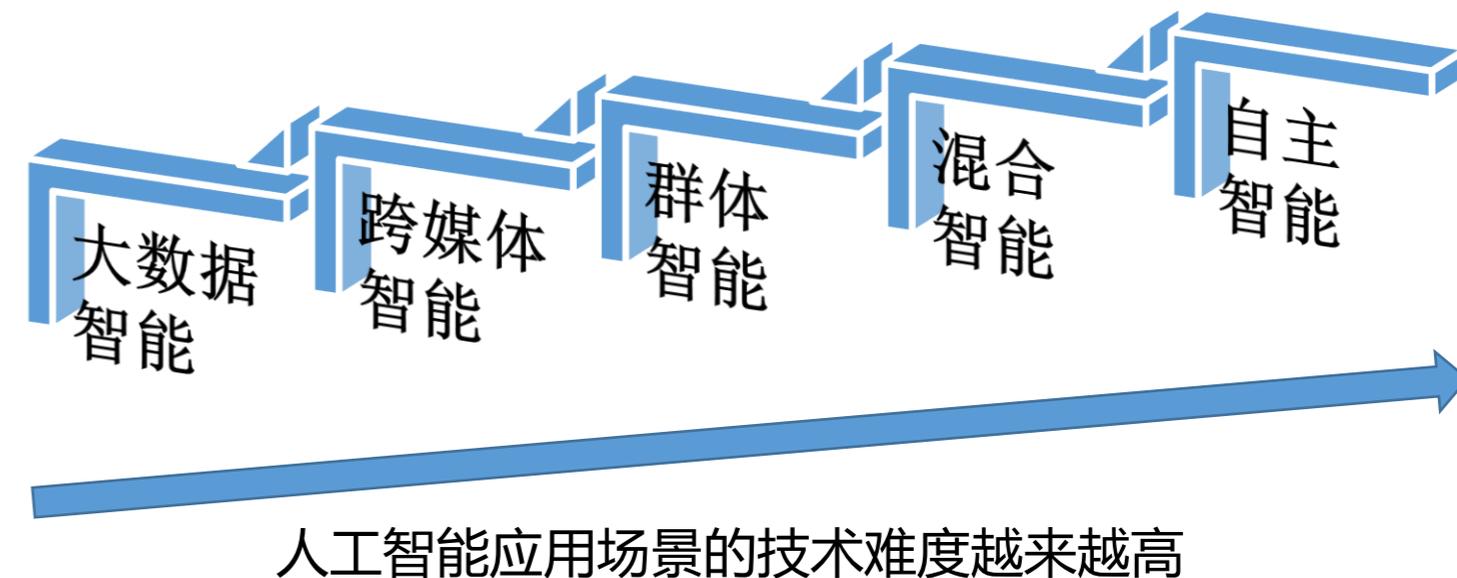
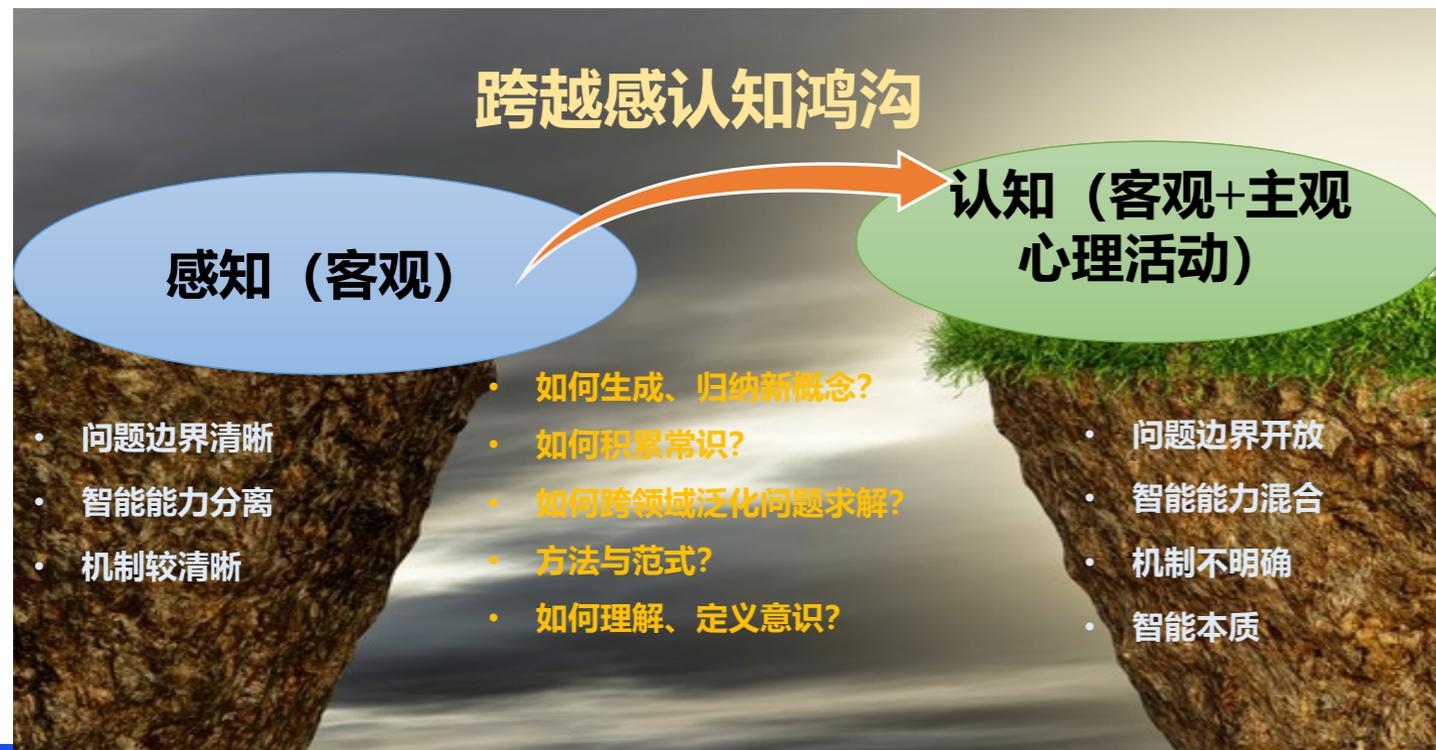
AI应用服务公司研发成本降低

可以重点发展落地应用

未来趋势：从语言智能走向现实世界多模态具身智能

- 虽然人工智能感知能力很强，但是其**认知泛化能力不足**，需要重点探索类人多模态感知、人机混合智能理论与方法，实现动态、开放环境下的**更通用认知能力**。

- 以大数据、跨媒体为代表的信息智能在**单源信息场景**得到广泛应用，达到实用成熟阶段，正向**多源信息复杂场景**及**自主无人系统**等决策/行为智能方向发展，实现从**认知世界**向**融入环境**的跨越。



感谢聆听

