

CIC灼识咨询 大模型与AIGC蓝皮书



灼识咨询是一家知名咨询公司。其服务包括IPO行业咨询、商业尽职调查、战略咨询、专家网络服务等。其咨询团队长期追踪物流、互联网、消费品、大数据、高科技、能源电力、供应链、人工智能、金融服务、医疗、教育、文娱、环境和楼宇科技、化工、工业、制造业、农业等方面最新的市场趋势，并拥有上述行业最相关且有见地的市场信息。

灼识咨询通过运用各种资源进行一手研究和二手研究。一手研究包括访谈行业专家和业内人士。二手研究包括分析各种公开发布的数据资源，数据来源包括中华人民共和国国家统计局、上市公司公告等。灼识咨询使用内部数据分析模型对所收集的信息和数据进行分析，通过对使用各类研究方法收集的数据进行参考比对，以确保分析的准确性。

所有统计数据真实可靠，并是基于截至本报告发布日的可用信息。

若您希望获取CIC灼识咨询的详细资料、与灼识建立媒体/市场合作，或加入灼识行业交流群，欢迎扫码、致电021-23560288或致函marketing@cninsights.com。



I. 大模型如今已展现出卓越的能力，随着其不断演进和完善，必将引领AI的未来道路，开创智能时代的新纪元

II. 受益于大模型技术的日益成熟，AIGC有望掀起新一轮产业革命，开启人类发展的智能新时代

III. GPGPU高度并行的计算资源为大型模型的训练和推理提供了不可或缺的支持，推动了大模型与AI领域的不断进步和创新

IV. 彩蛋：生成式AI如何提升工作效率

全球AI市场预计将于2030年达到1万亿美元，并且全球超过15%的GDP将由AI驱动；AI在各垂直领域的应用渗透率也将继续提升。

AI在垂直领域中的应用								强/主要的	弱/次要的
行业	数据及AI基础设施建设			中国AI应用情况				AI应用的挑战	
	数据量	数据质量	IT系统成熟度	中国AI支出占AI市场比例，2022	AI占IT支出比例		AI支出CAGR，22-27E		
					2022	2027E			
金融				12.1%	11.8%	19.8%	28.1%	不同机构之间的数据孤岛 数据系统安全及隐私	
零售				5.7%	3.7%	5.7%	26.6%	传统零售商获取结构化/高度可用的数据 供应链机器学习算法优化	
制造业				8.6%	1.8%	4.4%	43.6%	专注单点技术突破，但无法跨场景规模化应用 数据整合程度低，管理欠佳 AI解决方案供应商分散	
能源				2.6%	5.1%	9.8%	41.4%	AI应用/转型进度较慢	
医疗				5.6%	5.5%	8.2%	27.6%	数据分散在各医院/管理机构 数据系统安全及隐私	
汽车				12.3%	19.5%	25.5%	21.0%	自动驾驶数据融合难度高	
政府				27.1%	21.6%	28.3%	18.7%	不同机构之间的数据孤岛 数据系统安全及隐私	

大模型是基于海量多源数据打造的模型，其是实现通用人工智能（AGI）的重要路径。大模型可以整合多种不同类型的数据和信息，实现多模态处理和分析，从而更全面地理解 and 解决复杂问题，其具备通用性、涌现性等诸多特点。

定义

- AI可分为ANI (Artificial Narrow Intelligence)、AGI (Artificial General Intelligence) 和ASI (Artificial Super Intelligence)。ANI是专注于执行某一领域任务的经过训练的人工智能，是目前大多数的AI形式；AGI是通用人工智能，是指一种具有与人类相当的认知能力的智能系统，能够理解、学习、计划和解决问题；ASI是超级人工智能，指在几乎所有领域都超过最优秀的人类所具备的智能、知识、创造力、智慧和社交能力的智能系统。目前，ANI已经广泛应用，AGI处于研发阶段，而大模型是实现AGI的重要路径。

AI的三种类型比较



大模型是实现AGI的重要路径，其具有诸多特点



大模型产业涵盖了行业应用、产品服务、模型工具 and 基础设施四个关键层面，广泛应用于各行业与垂直场景，前景广阔。

大模型相关产业图谱

行业应用

产品服务

模型与工具

基础设施



大模型的演变经历3个阶段：基础大模型具备大量知识储备，能理解并生成内容，质量差强人意；精调大模型能按照人类偏好及普世价值取向、并保证一定安全性的前提下，生成高度可用的内容；而出现能力涌现的大模型所生成的内容接近人类水平。

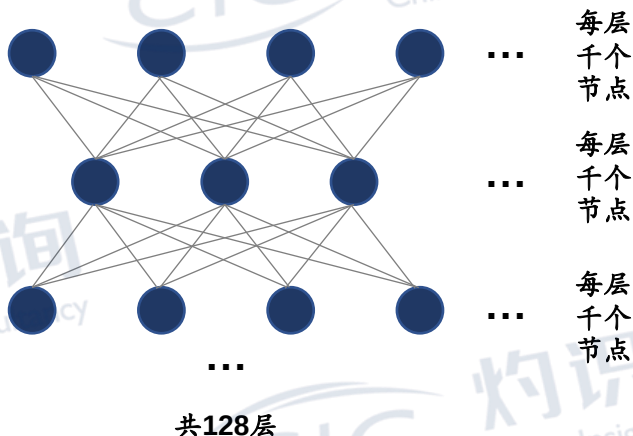
大模型演变概览

● 高/强

● 低/弱

1 基础大模型

模型示意图，以GPT-3模型为例



2 精调大模型

指令精调示意图

学习输出模板

人工对大量问题的回答形成模板供模型学习

建立评分机制

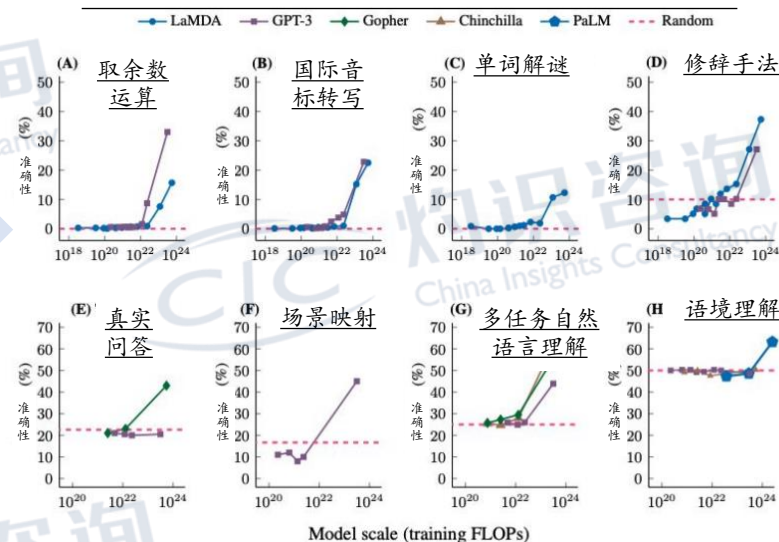
人工对模型的数个输出进行排序，建立奖励模型并为后续结果进行打分

根据结果重复优化过程

模型根据打分结果调整后输出，并不断重复上述过程以优化模型

3 能力涌现的大模型

能力涌现示意图



模型规模

模型变化

模型特征

- GPT-3采用了96层的多头Transformer，参数量达到1,750亿，并使用45TB数据进行训练。

- 指令精调在于确保模型输出结果的准确性及安全性，在该阶段需要使用较多的人工标注介入。在此基础上，引入奖励模型，让模型脱离人工指引，实现自优化。

- 随着模型参数的指数级增长，大模型能力呈现明显的爆发增长，呈现能力涌现的情况。

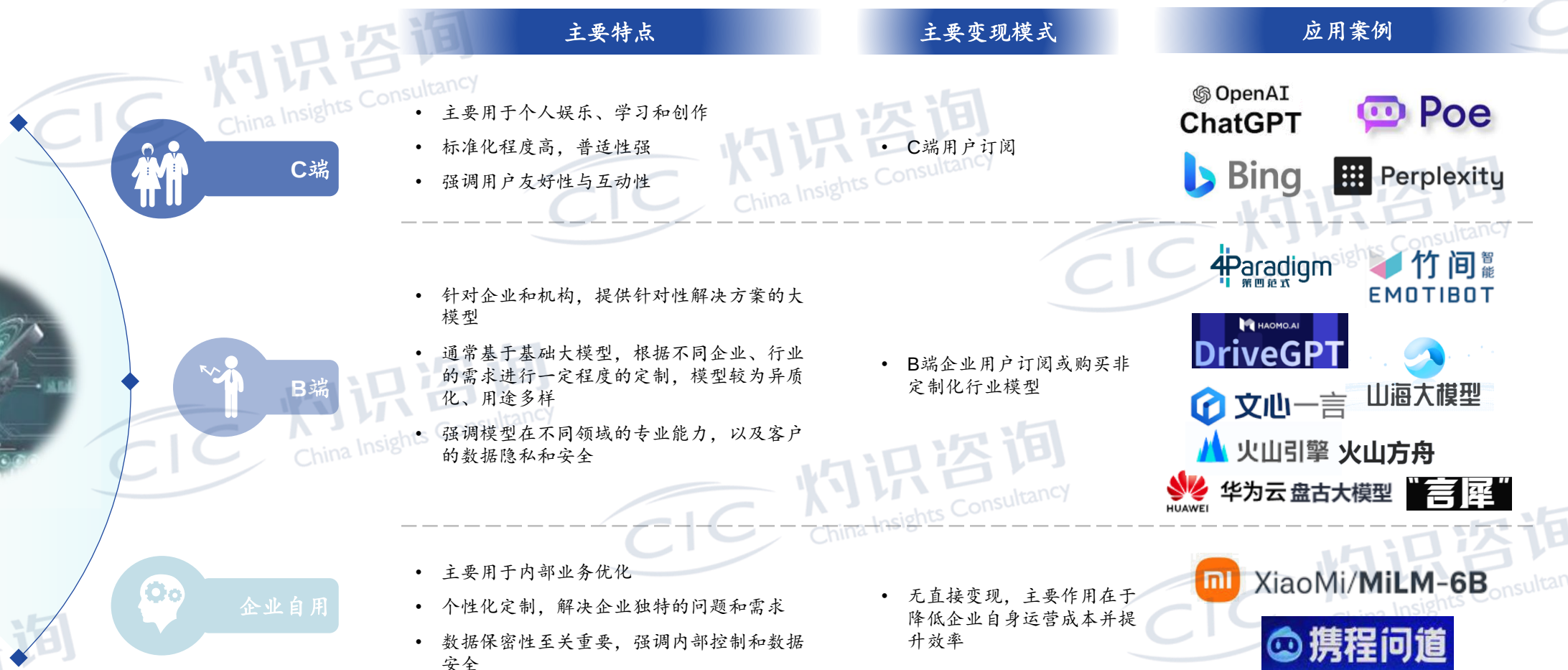
- 具备大量知识储备，能理解并生成语言，质量差强人意。

- 能按照人类偏好及普世价值取向，并保证一定安全性的前提下生成高度可用的内容。

- 具备逻辑推理能力及上下文理解能力，更接近人类水平。

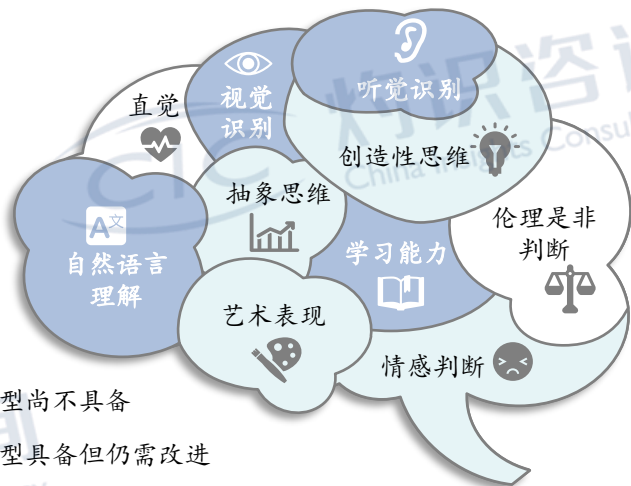
大模型的终端用户包括C端、B端和企业自用三类。C端用户一般使用标准化的模型产品，B端用户更倾向于选择能满足行业或企业特定需求的个性化解决方案，而企业自用的模型则旨在降低运营成本、提高运营效率、优化产品的用户体验。

大模型的终端用户分类



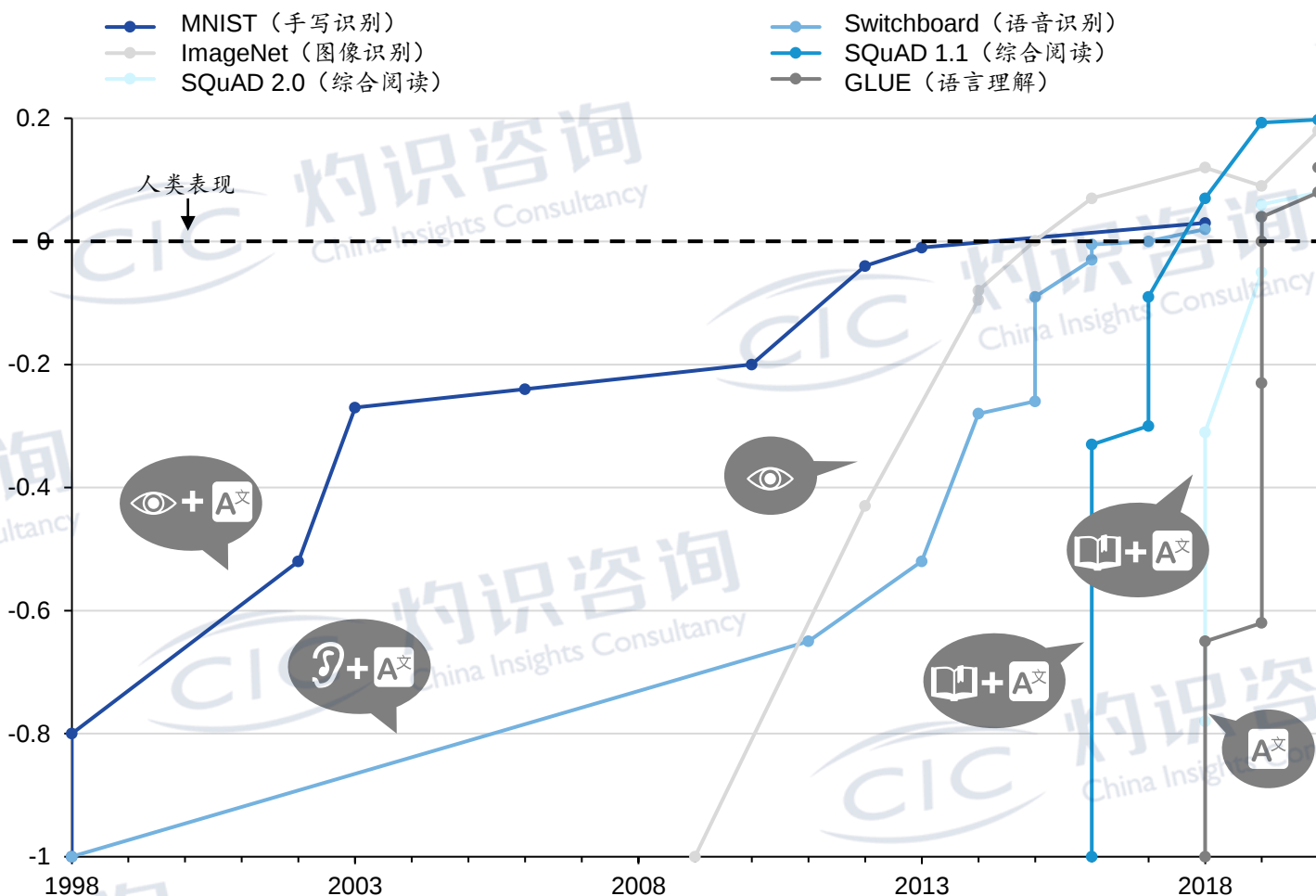
大模型在多领域表现卓越，在自然语言理解、学习能力、视觉听觉识别等领域可以媲美甚至超越人类。提高大模型的创造能力、减少对数据的依赖性、加强隐私保护等将会是大模型迭代的重点。

大模型在不同场景中与人类表现对比



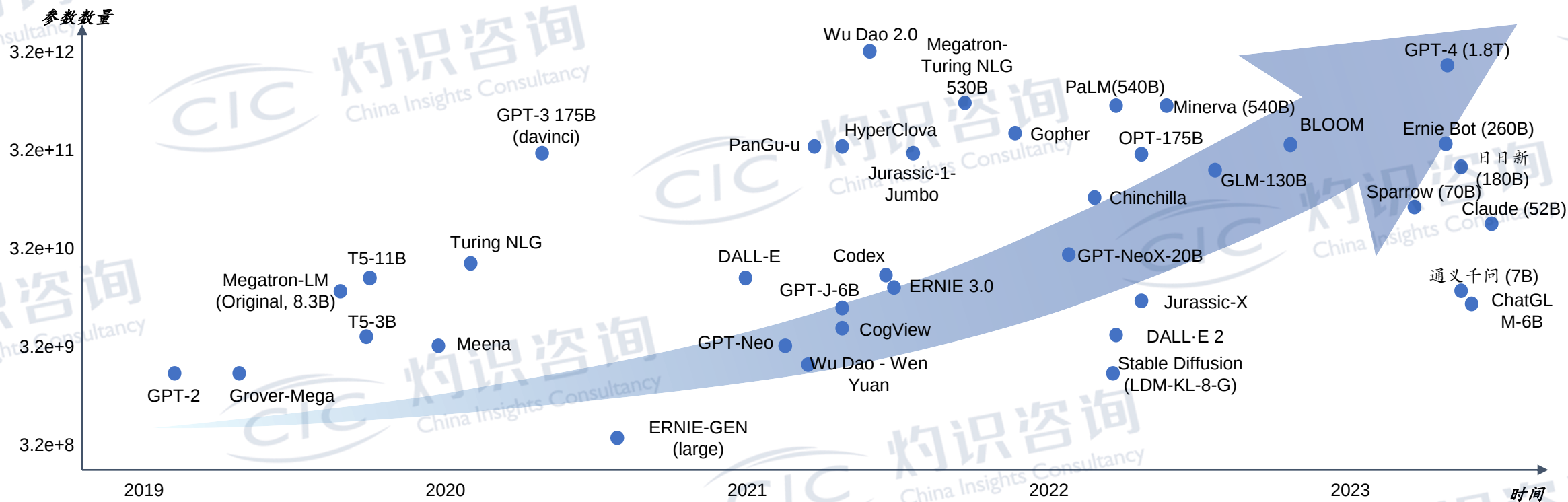
- 大模型尚不具备
- 大模型具备但仍需改进
- 大模型具备且可与人类媲美

- **视觉/听觉识别**：基本达到甚至一定程度上超越人类，准确率较高
- **学习能力**：具备自动学习能力，在特定任务和数据集上可以超越人类
- **自然语言理解**：熟练掌握多种人类语言，基本达到甚至一定程度上超越人类，存在一定语言不够自然（机械化）的问题
- **创造性思维**：能够生成创造性内容，但通常是在已知样本的基础上进行创作或是需要人类指导/二次修改
- **抽象思维**：相对有限，依赖于数据与模型参数
- **艺术表现**：可以生成艺术作品，但通常缺乏情感和创新
- **情感判断**：能够进行情感分析，但不具备真实情感体验
- **伦理是非判断**：不具备，可能引发错误或数据隐私等安全问题
- **直觉**：不具备



随着大模型的不断演进，它们的参数规模也呈现出指数级增长的趋势；与此同时，它们的复杂性和功能愈发提升，使得大模型能够在各领域担任更加多样和复杂的任务。

大模型的发展及参数量



分析

- 在大模型出现之前，机器学习算法的参数量以平均每5-6年翻一个数量级的速度快速增长，而大模型的出现使模型参数量的增长速度大幅提升。
- 举例而言，大模型及多模态模型的鼻祖之一GPT-2发布于2019年，参数量为15亿；GPT-3发布于2020年，参数量即达到了1,750亿，相比GPT-2增长了100多倍。粗略计算，在大模型兴起的前几年，大模型的参数量每年即可增长1-2个数量级。
- 在目前已公开参数量的大模型中，参数量最多的达到了1.75万亿。

大模型正在各应用场景崭露头角，其应用场景广泛且多样化。从自然语言处理到图像生成，从音频处理到视频、3D场景创建，大模型能够应对多领域的挑战。未来，大模型有望为更多应用场景带来更多机会和创新。

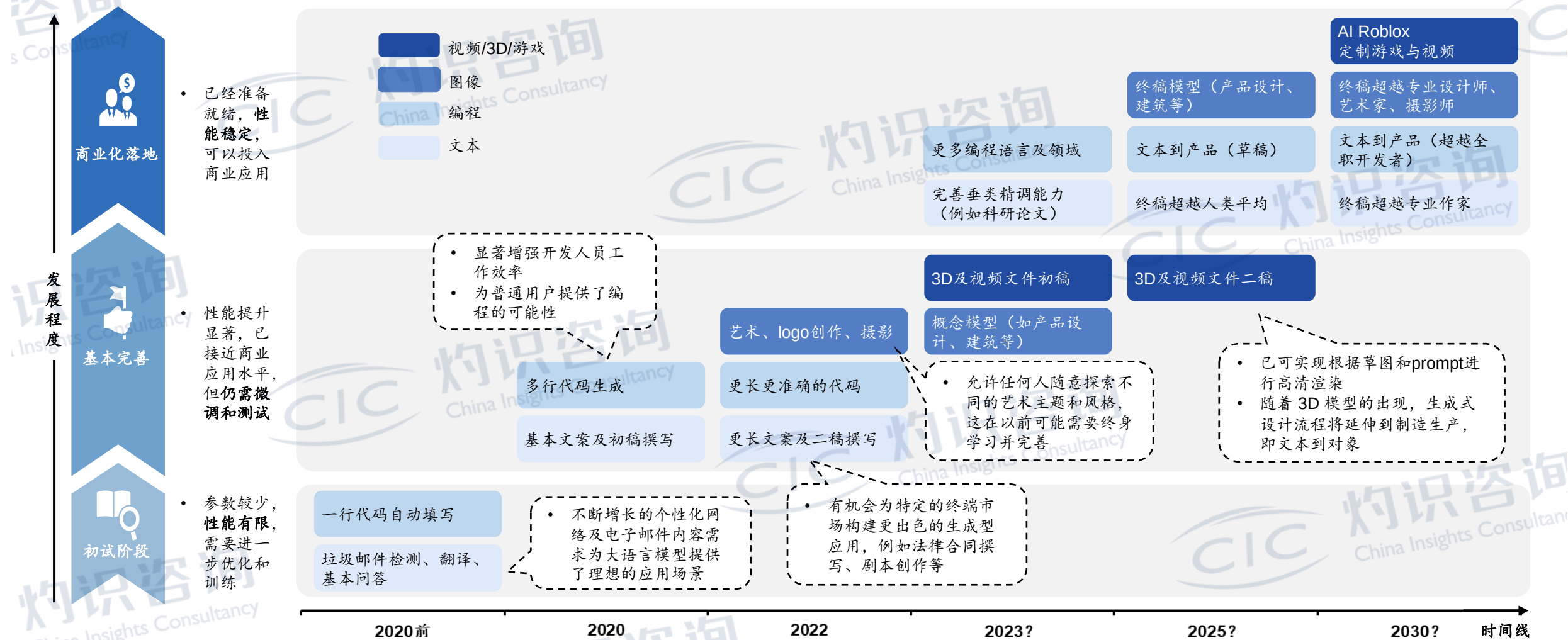
大模型的应用场景

应用场景	主要用途	相关大模型		描述	
 文本	- 营销（内容） - 销售（邮件） - 通用写作 - 记笔记	 OpenAI  DeepMind  facebook  Hugging Face	GPT Gopher OPT Bloom	 cohere  ANTHROPIC  AI2  Alibaba Yandex	- 模型比较擅长通用的短/中篇幅写作，通常用于初稿及更新稿撰写 - 模型能够理解上下文，生成更自然的文本，准确性逐渐接近人类水平，广泛应用于智能客服、文本摘要、内容生成等领域
 编程	- 代码生成 - 代码文档化 - 文本到SQL - 网页应用构建	 OpenAI  tabnine	GPT Tabnine	 stability.ai	- 代码生成可能在短期内对开发人员的生产力带来重大影响 - 能降低非开发人员编程的门槛
 图像	- 图像生成 - 消费者/社交应用 - 媒体/广告设计	 OpenAI  Stable Diffusion	Dall-E 2 Stable Diffusion	 craiyon	- 在图像识别、分割、风格转换等领域有广泛应用，能够处理复杂的视觉任务 - 不同风格的图像模型以及编辑和修改生成图像技术
 音频	语音合成	 OpenAI	OpenAI		对声音和语音的理解能力逐渐提高，生成的音频逐渐自然、不机械，且接近人类水平
 视频	视频编辑/生成	 Microsoft	X-CLIP	 Meta	用于自动剪辑、特效创作、噪音修复、虚拟场景制作等，大大提升了编辑效率
 3D	3D模型/场景搭建	 DREAM FUSION  NVIDIA	DreamFusion GET3D	MDM (Motion Diffusion Model)	在游戏、电影、虚拟现实、建筑和实体产品设计等大型创意市场极具潜力

资料来源：红杉资本，灼识咨询

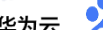
大模型在多个领域的应用上取得了显著进展，文本生成、编程、图像处理等领域的能力逐渐完善，有望商业化落地。然而，3D、视频、游戏等领域发展相对较慢，但潜力巨大，需要更多时间来成熟和商业化。

大模型在不同领域的进展



大模型也已开始逐步渗透进金融、医疗、汽车、制造、零售等行业，头部大厂及垂类解决方案服务商相继推出适用于各垂直行业的解决方案，助力行业变革。

大模型在垂直行业的应用

垂直行业	大模型应用情况	典型案例	代表玩家
金融	 在金融领域，大模型主要应用包括风险评估、量化交易、柜台业务办理等。以风险评估为例，大模型可用于预测信用风险、欺诈检测和市场趋势分析。	 第四范式携手中关村银行，以“式说”大模型为基础，联合开展多模态智慧金融平台研发及私有化部署，在行内规章制度及人员信息查询、行业术语通俗化解释、客户经理信贷管理、人工客服问答推荐、理财业务办理、AI应用快速研发等多个环节形成大模型能力。	 华为云      
医疗	 在医疗领域，大模型可应用于疾病诊断、药物发现等场景。以疾病诊断为例，大模型可用于医学图象识别，如X射线、MRI和CT扫描，以帮助医生诊断疾病。	 科大讯飞利用星火认知大模型打造“诊后康复管理平台”，将“专业的诊后管理和康复知道延伸到院外”，可根据患者健康画像自动分析，为患者生成个性化康复计划，并督促患者按计划进行。	 科大讯飞  商汤   华为云  阿里云  智云健康  云知声
汽车	 在汽车领域，大模型可应用于自动驾驶、虚拟设计和仿真、质量控制、供应链管理、销售和营销、定制设计、新能源汽车性能优化等。	 百度智能云为长安汽车提供了人工智能基础设施平台和数字人平台，双方正在开发一款基于文心大模型的生成式人工智能产品，以提升用户体验。同时，百度与吉利汽车一起打造汽车行业大模型，构建工厂数字化大脑，降低管理运维成本，提高资源利用效率。	 吉利汽车  百度  斑马智行  商汤  百度  HAOMO.AI
制造	 在制造业领域，大模型主要应用场景包括工业自动化、供应链优化等。以工业自动化为例，大模型可用于监控生产过程，预测设备故障，提高生产效率。	 华为云 华为云携手赛意信息，基于盘古大模型的自然语言大模型和视觉大模型，共同打造制造业大模型，为制造业客户提供针对工艺工程AI自动化、AI质检、供应链的智能优化等细分场景，提供“开箱即用”的模型服务。	    华为云  羚羊工业互联网
零售	 在零售行业，大模型可应用于个性化推荐、价格优化等场景。以个性化推荐为例，大模型可用于分析客户行为，提供个性化产品建议和广告走向。	 京东发布“言犀”大模型产品，支持语言、语音、视觉、多模态等应用。以AIGC创作内容营销为例，大模型的应用可使每套图的制作成本降低90%，制作周期从7天缩短至半天。	 京东云  华为云  腾讯云  阿里云  商汤  科大讯飞

随着大模型相关政策的出台，大模型的发展得到了进一步的推进和规范。这些政策不仅为大模型的研发和应用提供了指导和支持，同时也为大模型的推广和应用提供了更加广阔的空间和机遇。

大模型相关政策概览

2011 – 2015 “十二五” 规划

- 推进产业智能化改革
- 促进基于互联网的人工智能应用

2016 – 2020 “十三五” 规划

- 强调人工智能为战略前沿领域
- 制定人工智能发展目标

2021 – 2025 “十四五” 规划

- 明确人工智能应用的重点领域
- 规范生成式人工智能服务

2015年5月：国务院《中国制造 2025》

- 提出基于信息物理系统的智能制造技术引领制造方式变革
- 推进生产过程智能化

2015年7月：国务院《国务院关于积极推动“互联网+”行动的指导意见》

- 加速人工智能核心技术突破
- 促进人工智能在生活、工作、公共领域的创新应用，如智能家居、智能终端、服务型机器人等

2016年7月：国务院 《“十三五” 国家科技创新规划》

- 加强智能机器人等关键技术研发和应用推广
- 整合优化资源配置，瞄准引领未来发展的战略领域

2017年7月：国务院 《新一代人工智能发展规划》

- 2030年成为世界主要人工智能创新中心
- 深化人工智能在制造、农业、金融、教育、医疗、交通等领域的融合应用

2020年7月：中央网信办等五部门《国家新一代人工智能标准体系建设指南》

- 加强人工智能领域标准化顶层设计，推动人工智能产业技术研发和标准制定

2021年3月：全国人大 《中华人民共和国国民经济和社会发展第十四个五年规划和 2035年远景目标纲要》

- 瞄准人工智能等前沿领域，实施一批具有前瞻性、战略性的国家重大科技项目

2022年8月：科技部《关于支持建设新一代人工智能示范应用场景的通知》

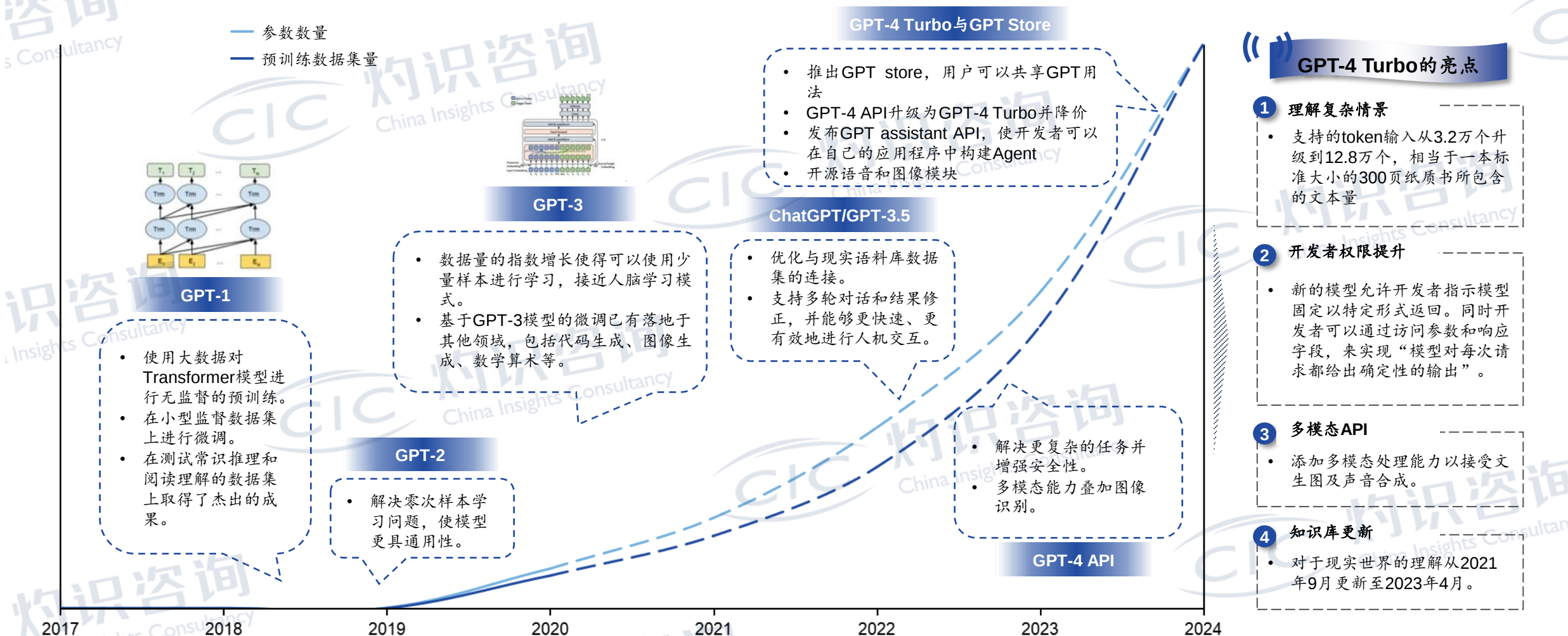
- 强调人工智能在重点领域的应用，包括公共安全、智慧城市、医疗健康、重大活动等需求迫切且普惠性强的领域

2023年4月：国家网信办《生成式人工智能服务管理办法（征求意见稿）》

- 规定了提供生成式人工智能产品或服务应当遵守法律法规的要求
- 利用生成式人工智能生成的内容应当体现社会主义核心价值观

ChatGPT：建立在OpenAI引入的深度学习模型GPT基础上，它已经成为迄今为止增长最快的APP之一。

GPT的发展历程



百度文心：中国领先大模型之一，且对于中文有更深入的理解。

百度文心一言概览

文心一言4.0

VS

文心一言3.5

文心一言4.0相比于3.5版本，显著提升了
理解、生成、逻辑、记忆四大能力

逻辑能力提升幅度是
理解能力提升幅度的
300%

能够理解乱序、模
糊意图、潜台词等
复杂提示词

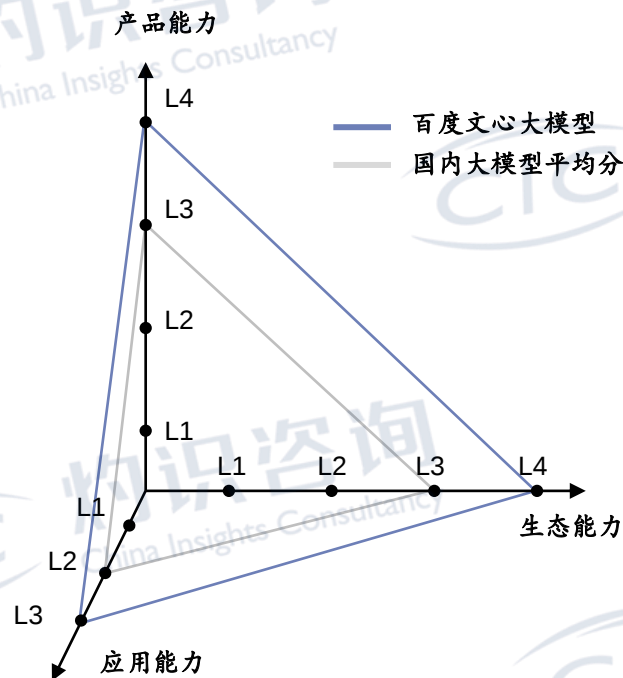
逻辑能力提升，
解答数学难题并
总结知识点

记忆能力提升幅度是
理解能力提升幅度的
200%

几分钟内生成
海报、营销视频等复
杂图像和视频

能够完成完成
千字小说
撰写和角色、情节设置

文心一言4.0的亮点



1 国内市场格局领先

- 根据IDC公布的数据，百度文心大模型在产品能力、应用能力、生态能力方面在中国领先。

2 对中文的深入理解

- 百度作为中国搜索引擎龙头，拥有更多的中文语料数据进行模型训练。因此，百度文心大模型对中文有更深入的理解，能够熟练使用成语、写藏头诗等。

注：目前尚未有厂商在产品能力和生态能力方面达到L5，应用能力尚未有厂商达到L4，因此不在本次评估中凸显



I. 大模型如今已展现出卓越的能力，随着其不断演进和完善，必将引领AI的未来道路，开创智能时代的新纪元

II. 受益于大模型技术的日益成熟，**AIGC**有望掀起新一轮产业革命，开启人类发展的智能新时代

III. GPGPU高度并行的计算资源为大型模型的训练和推理提供了不可或缺的支持，推动了大模型与AI领域的不断进步和创新

IV. 彩蛋：生成式AI如何提升工作效率

AIGC是继PGC和UGC之后，经由AI自动生成内容的新型内容生产方式。作为一种先进的生产力变革，AIGC的兴起让世界进入了智能创作时代。

内容生产阶段及比较，2000年至今

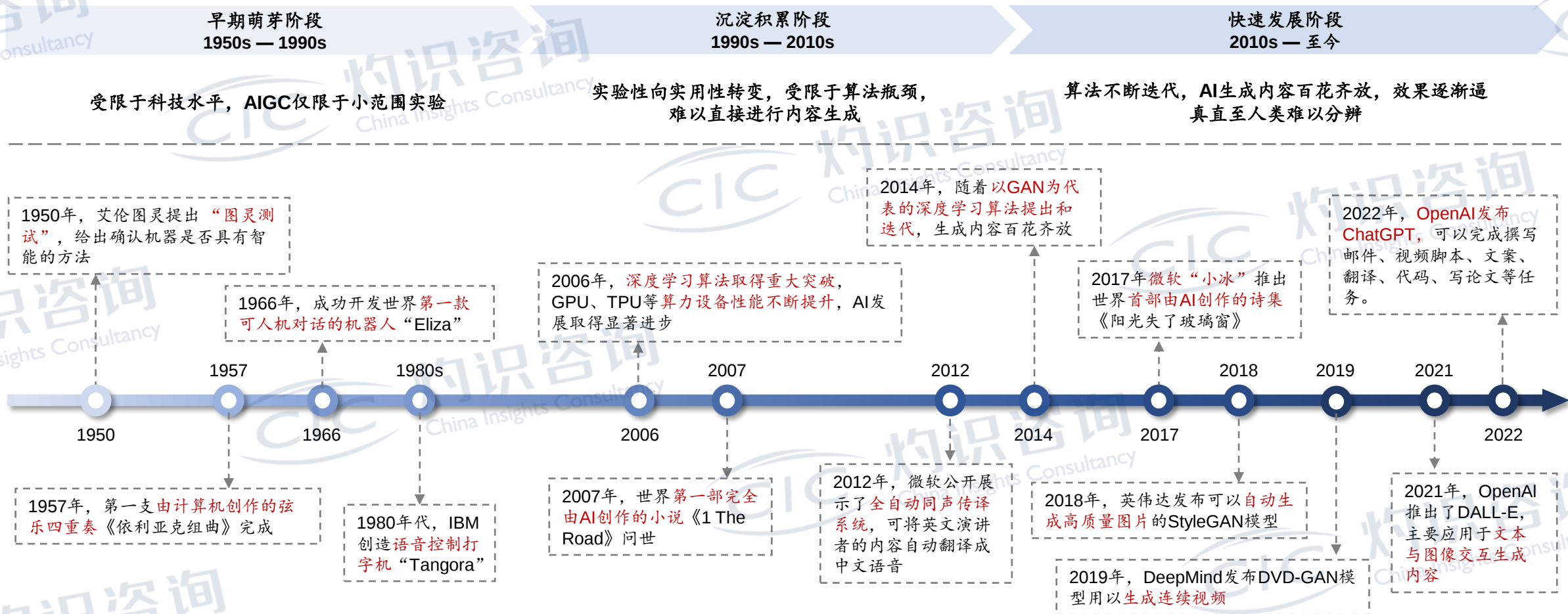
● 高/强

● 低/弱

	 PGC（2000-2010年） 专业化的内容生产	 UGC（2010-2017年） 用户生产内容	 AIGC 1.0（2017-2022年） AI辅助内容生产	 AIGC 2.0（2022-未来） 未来：全AI生产内容
内容生产方式				
定义	官媒时代，由专业的内容创作者或团队进行创作、编辑或发布的内容	由普通用户或受众参与创作、编辑和发布的内容，微博微信的诞生为UGC发展提供土壤	基于PGC、UGC生成创作框架，下达指令让AI自动生成内容，指导AI完成编码、绘画、建模等复杂任务	OpenAI发布，大模型时代开启，创作突破人为限制，提升到自主创作的层次，创造更加丰富多样的内容
内容生产者	专业用户	一般用户	AI基于指令创作	大模型逐步实现自主创作
内容生产成本	为保障质量，一般PGC的内容制作成本较高，需要投入大量的人力、物力和财力	用户个性化、自主创作，通常制作成本较低	随着生成内容数量增加，边际成本大幅下降	知识性、创造性工作的边际成本将降至无限接近0，产生巨大生产效率和经济价值
内容生产效率	包括主题选定、创意构思、策划、撰写、编辑、排版、发布等多个环节，需要长时间投入	通常为个人兴趣爱好、在自媒体平台、社交网络等自发的创作行为，创作流程简单直接	通过程序训练生成，流程自动化，仅需进行算法训练与参数调整，效率较高	随着技术进步，AIGC将越来越多用于快速生成内容及自动化创作，效率进一步提升
内容生产质量	内容生产者拥有专业知识、内容相关领域资质和一定权威，因此内容相对专业、权威	内容丰富但质量参差不齐，通常需要平台设计规则加以约束或遴选出优质内容	发展初期存在使用门槛高、内容生成简单和质量偏低等问题	随着AI大模型从大量数据中进一步学习，内容质量进一步提升
内容可控性	内容生产者及生产程序专业性内容经过筛选，从而带来强内容可控性	每天有大量用户自由创作，需要依靠机器及人工对内容进行审核，内容可控性弱	AI大模型作为辅助人进行内容生产的工具，可根据指令对内容进行编辑及优化，可控性较强	通过深度学习技术训练模型，使其能够学习到输入数据的内在规律，并根据规律创作内容，可控性强

结合AI的演进历程，AIGC的发展大致可分为早期萌芽、沉淀积累、快速发展三个阶段。

AIGC发展阶段、发展特点和典型事件



对于企业而言，AIGC在降低内容制作成本、加快内容制作效率的同时提升了内容的多样性及质量，AIGC与不同产业融合互动亦可孕育出新业态新模式；对个人而言，AIGC则显著降低了内容创作门槛。

1 降本增效

- **降低内容制作成本**：例如游戏行业，美术创作者可以使用AIGC寻找灵感，降低初期成本投入
- **提升效率**：承担信息挖掘、素材调用、复刻编辑等基础性机械劳动，从技术层面实现以低边际成本、高效率的方式满足

3 提升内容创作质量

- 有助于企业提升交付工作成果，近年来AI模型在手写、语音和图像识别、阅读理解和语言理解方面的表现逐渐超过了人类的基准水平。

2 增加内容多样性

- **生产多模态内容**：有助于企业多元化交付工作成果，例如通过文本形成图片、基于文本生成视频短片，满足海量个性化需求

4 新商业模式

- 通过支持AIGC与其他产业的多维互动、融合渗透从而孕育新业态新模式，为各行各业创造新的商业模式，提供价值增长新动能

5 降低内容创作门槛

- AIGC降低了用户的内容制作门槛，例如没有绘画基础的用户也能借助AIGC工具快速作画，内容创作将迎来爆发期。



底层生成算法和开发架构是实现AIGC所必不可缺的部分。底层生成算法是实现内容生成的基石，通常是深度学习模型，而开发架构是支撑底层算法的基础设施，以便开发人员较易集成多种算法组成。

AIGC技术架构概览



AIGC 关键底层算法模型

资料来源：灼识咨询

AIGC产业仍处于早期发展阶段，在数据、算法和算力的基础之上，核心模型以关键企业研究成果为主导，AIGC以AI生成文本、图像、视频和音频为主流形式。

AIGC产业链概览

上游



基础层

数据

- 包括业务数据联通、素材数据搜集等

算法及模型

- 包括神经网络、深度学习、自然语言处理技术、扩散模型、云渲染等开源算法。
- 以及基于算法通过数据训练生成的机器学习模型、深度学习模型及大模型等。



算力

- 包括服务器、GPU、HPC、云等提供基本计算能力的基础设施



中游



应用层

基于模块分类

文本生成



图像与视频生成



音频生成



下游



终端用户层

各类内容创作及分发平台



内容终端生产厂商



第三方内容服务机构



AIGC产品形态多元，如文本生成、音频生成、图象生成、视频生成、代码生成、3D模型等，赋能“数字化程度高+内容需求丰富”的行业，如娱乐、传媒、影视、电商等，驱动各行业创新。

AIGC应用场景概览

AIGC应用场景

AIGC产品形态



AIGC应用场

娱乐领域

全民娱乐

- 人脸美妆/融合
- 黑白图像上色
- 图像风格转换

社交互动

- 虚拟歌姬
- 虚拟直播

偶像养成

- 用户数字化身

传媒领域

信息采集

- 采访助手
- 多语言翻译

内容编辑

- 写稿机器人
- 剪辑视频集锦
- 视频字幕自动生成

内容播报

- AI主播
- 智能播报

影视领域

前期创作

- 剧本创作
- 场景设计辅助

中期拍摄

- 数字复活已故演员
- 高难度动作合成
- 演员角色年龄的跨越

后期制作

- 替换劣迹艺人
- 影视作品修复
- 影视内容2D转3D

电商领域

商品展示

- 智能商品详情
- 商品3D模型生成

主播打造

- 品牌虚拟主播

交易场景

- 虚拟商城构建

其他

教育

- 合成历史人物
- 视频虚拟教授
- 线上课程音视频处理

工业

- 辅助工程设计
- 数字孪生系统构建

金融

- 营销视频合成
- 智能客服
- VR金融场景构建

医疗

- 医学图像处理
- 智能病例录入
- 合成肢体投影

AIGC+娱乐：扩展辐射边界，获得发展动能

- 实现趣味性图像或音视频生成，激发用户参与热情
- 打造虚拟偶像，释放IP价值
- 开发C端用户数字化身，布局消费元宇宙

AIGC+传媒：人机协同生产，推动媒体融合

- 采编环节：采访录音语音转写，提升传媒工作者的工作体验；智能新闻写作，提升新闻资讯的时效；智能视频剪辑，提升视频内容价值
- 传播环节：AIGC应用主要集中在以AI合成主播为核心的新闻播报等领域

AIGC+影视：拓展创作空间，提升作品质量

- 为剧本创作提供新思路
- 扩展角色和场景创作空间
- 赋能影视剪辑，升级后期制作

AIGC+电商：推进虚实交融，营造沉浸体验

- 生成商品3D模型用于商品展示和虚拟试用，提升线上购物体验
- 打造虚拟主播，赋能直播带货
- 赋能线上商城和线下秀场加速演变，为消费者提供全新的购物场景

AIGC+其他：推进数实融合，加快产业升级

- 教育领域：AIGC赋予教育材料新活力
- 金融领域：AIGC主力实现降本增效
- 医疗领域：AIGC赋能诊疗全过程
- 工业领域：AIGC提升产业效率和价值

不断迭代升级的算法可以实现AI模型的性能稳定、功能进阶，进而使得AIGC产品表现愈加出色。

AIGC算法不断迭代升级

AIGC算法技术升级步入深化阶段

传统机器学习算法

算法

回归算法：线性、回归树等

分类算法：逻辑回归等

聚类算法：仿射传播等

算法特点：

- 不具备强大的学习能力，依赖于预先定义的统计模型或专家系统执行特定的任务，可以完成简单线条、文本和旋律的生成

算法缺点：

- 算力挑战：传统机器学习方法的维度诅咒等导致算力挑战
- 缺乏感知：生成内容空洞、刻板、文不对题

算法示例：

- 通过定义复杂的函数方程组，计算机所绘出的函数曲线具备某种美学图样

深度神经网络

早期神经网络

卷积神经网络 CNN

递归神经网络 RNN

生成对抗网络 GAN

深度强化学习

深度变分自编码器

流模型

扩散模型

深度残差网络

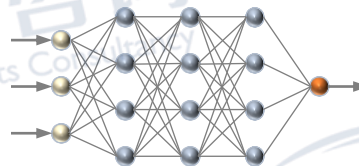
Transformer

学习范式
+
网络结构不断
迭代

- 损失函数和梯度下降算法可以灵活快速的调整深度神经网络中的参数，从而实现从数据中进行学习的功能

CNN/RNN 神经网络与Transformer对比

CNN/RNN 神经网络



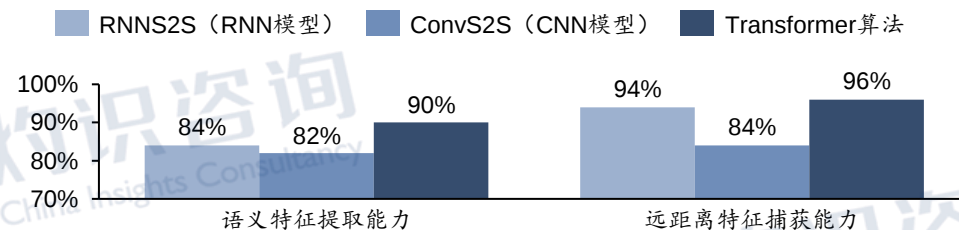
- 引入大规模并行处理，模拟人脑分类学习和适应性联想。

Transformer



- 使用注意力机制进行优化，使语句中的词语有了重点区分。GPT-3、InstructGPT、ChatGPT都是在此模型上迭代、演进的。

Transformer与RNN、CNN准确度对比



关键
分析

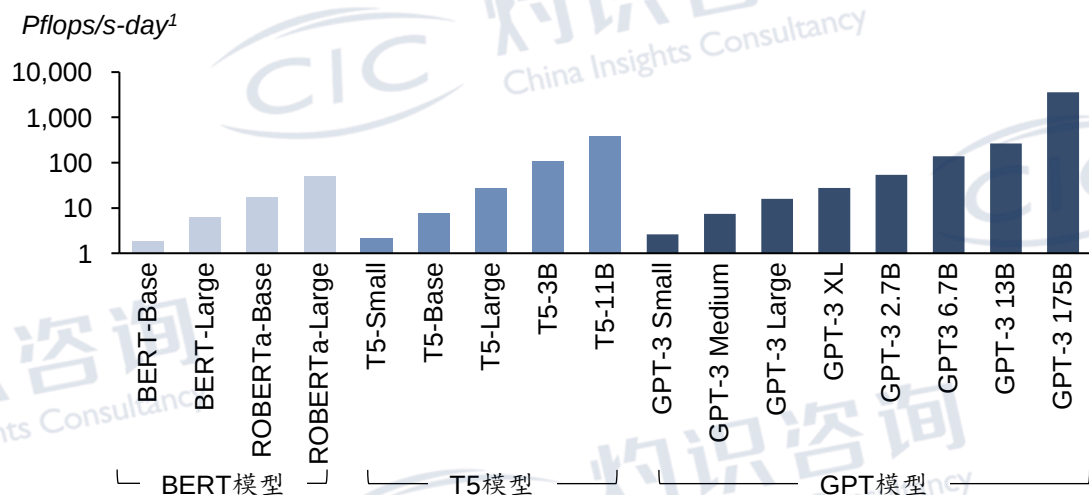
- Transformer算法具备跨时代意义的原因是算法上添加了注意力机制。这种机制突破了RNN模型不能并行计算的限制；相比CNN模型，其关联所需的操作次数不随距离增长。
- Transformer算法在语义特征提取能力、远距离特捕获能力等方面明显增强。

算力提升是AIGC发展的重要驱动力之一，愈大的算力规模可以支撑更大规模的模型运转。



算力是打造大模型生态的必备基础，AIGC拉动算力需求指数级增长

不同模型训练需要的算力



算力是打造大模型生态的必备基础

- 算力是指计算设备执行算法、处理数据的能力，包括CPU、GPU等。算力的应用包括数据中心、分布式计算、云计算、边缘计算等。AIGC背景下，大模型是AI发展的必然趋势，而算力是打造大模型生态的必备基础。

AIGC拉动算力需求指数级增长

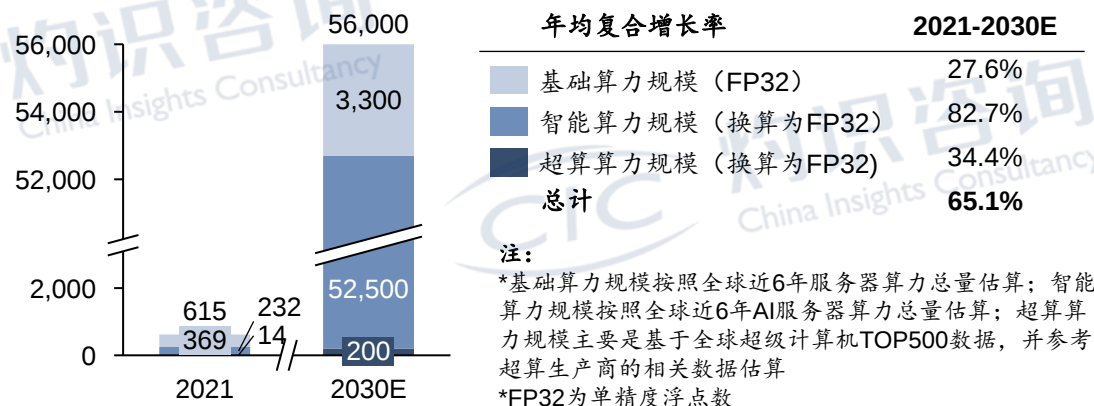
- 大模型训练所需算力每3-4个月增长1倍，增速远超摩尔定律（18-24个月/倍）。模型预训练过程是消耗算力的最主要场景，AI训练所需算力呈现指数增长态势。



算力芯片、AI服务器有望持续放量，数据中心建设提速，全球算力总规模将保持高速稳定增长态势

全球计算设备算力总规模，以FP32²计，2021-2030E

Eflops (10¹⁸次)



算力芯片、AI服务器有望持续放量，数据中心建设提速

- 需求拉动算力芯片及AI服务器的技术迭代，预计未来将持续放量；同时，各国纷纷提速数据中心建设，我国“东数西算”工程投资规划逐步清晰。

全球算力规模将保持高速稳定增长态势

- 预计2030年人类将迎来YB数据时代，全球算力规模达到56ZFlops (10²¹次)。预计智能算力2021至2030年年均复合增长率将超65%，以支撑庞大的新增非结构化数据（文本、图片、语音、视频等）。

注：1. Pflops/s-day：以10¹⁵次每秒浮点运算次数的速度进行计算，需要消耗的天数；2.单精度浮点数，用8bit表示指数，23bit表示小数。

AIGC通常可同时面向企业和个人变现，随着用户方需求持续升级和AIGC产品愈加成熟，诸多潜在的变现方式也将进一步刺激AIGC行业蓬勃发展。

AIGC变现模式概览



潜在变现方式

- 1 作为底层平台收费**
 - 作为底层平台接入其他产品对外开放，按照数据请求量和实际计算量计算
 - 如GPT-4对外提供API接口，采用的四种模型分别采用不同的按量收费方式
- 2 产出内容收费**
 - 按平台和工具产出内容量收费
 - 大多数AI图像生成平台大多按照图像张数收费
- 3 提供软件服务收费**
 - 主要服务于C端企业用户，大部分C端AIGC工具则以约80人民币/月的价格对外出售
- 4 模型训练收费**
 - 模型训练费用，适用于NPC训练等个性化定制需求较强的领域
- 5 具体属性收费**
 - 针对对特殊性质的内容自动生成服务进行专门化收费方式设计

OpenAI 变现模式概览

2C 个人用户端

ChatGPT Plus

产品简介：AI技术驱动的自然语言处理工具
收费标准：每月20美元

DALL·E

产品简介：AI技术驱动的文本生成图像模型
收费标准：15美元可购买115个积分，约能生成460张图片，并获得图片的使用权，包括商业用途

GPT-4 API接口

产品简介：AI技术驱动的大规模多模态模型，可以接受图像和文本输入，产生文本输出
收费标准：不超过8K文本长度时，定价为每1K个提示令牌0.03美元，每1K个补全令牌0.06美元

ChatGPT API 接口

产品简介：AI技术驱动的自然语言处理工具
收费标准：每1K个提示令牌0.002美元

InstructGPT API 接口

产品简介：专注于解决指导型对话
收费标准：每1K个提示令牌0.0004美元-0.02美元

DALL·E API 接口 文本生成图像

Whisper API 接口 自动语音识别

微调模型
增加分类器或特征映射微调
GPT-3模型

嵌入模型
将一个内容实体映射为低维向量，从而可以获得内容之间的相似度

2B 企业端

2C 个人用户端

按产出收费

OpenAI、DALL·E、Deep Dream Generator等AI图像生成平台大多按照图像张数收费，收费方式更符合用户预期

会员制

用户可免费享受部分功能，可以通过成为会员获取更优质的内容服务，主要包含定制化、更高质量及多模态的输出内容

按次收费

微软、亚马逊、谷歌的人工智能平台存在按API盈利的模式，用户调用的API次数越多，收费越高。

2B 企业端

按次收费

客户根据需求按次收费，通常是按照API调用次数进行收费

会员制

客户可以按月会员制付费，得到相应的服务和升级，内容输出根据方案的不同呈现阶梯式



I. 大模型如今已展现出卓越的能力，随着其不断演进和完善，必将引领AI的未来道路，开创智能时代的新纪元

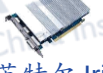
II. 受益于大模型技术的日益成熟，AIGC有望掀起新一轮产业革命，开启人类发展的智能新时代

III. GPGPU高度并行的计算资源为大型模型的训练和推理提供了不可或缺的支持，推动了大模型与AI领域的不断进步和创新

IV. 彩蛋：生成式AI如何提升工作效率

GPU可根据其功能和集成方法进行分类。GPGPU出色的处理和并行计算能力被广泛应用于大模型、高性能计算和其他领域，也是未来AI发展的重要基础设施。

GPU分类

	定义	特征	应用	典型产品
通过功能分类	GPGPU GPGPU是执行原本由CPU处理完成的通用计算任务的处理器。	GPGPU具有GPU高度并行的非专用计算能力，并针对通用计算任务进行了优化，尤其适合大规模数据集的训练和推理。	- 大模型 - 数据中心 - 高性能计算 ...	 AMD MI300 英特尔 Ponte Vecchio
	图形GPU 图形GPU是专门处理图形计算任务的处理器。	图形GPU专注于传统的图形处理计算，如渲染图像、视频和游戏视觉效果制作等。	- 非线性编辑 - 视频编解码 - 图形加速可视化 3D设计 ...	 AMD Radeon PRO 英特尔 ARC
	独立GPU（商用） 独立GPU是显示芯片及相关器件制作成一个独立于电脑主板的板卡，成为专业的图形处理硬件设施。	通过独显插到主板相应接口上，具备高位宽，高频独立显存和更多的处理单元，性能优越，不仅可用于一般性的工作，还具有完善的2D效果和很强的3D水平。	常用于高性能台式机和笔记本电脑以及服务器	 AMD X1250 英特尔 GMA X4500HD
通过集成方式分类	集成GPU 集成GPU是指一般不带显存，而是使用系统的一部分主内存，可以被整合进主板作为芯片的一部分，也可以和CPU集成。	集成在CPU里面的图形处理单元，构成CPU的一部分，其价格低，兼容性较好，但是性能相对较差，功耗低。	常用于笔记本电脑	 英特尔 Iris Xe DG1

在GPGPU产业链中，上游供应商主要为中游的芯片设计公司提供服务，在GPGPU生产出来后再被转交给下游的终端用户。随着GPGPU在更多新兴应用场景中的普及，预计未来终端用户的覆盖范围将进一步扩大。

GPGPU产业链

关键分析

上游

中游

下游

半导体IP

innosilicon
芯动科技

arm

Imagination

硅片设计

GUC
国家电子系统有限公司

VeriSilicon

alchip

Bröte
semiconductor

FARADAY

EDA工具

synopsys

Empress

华大九天

cadence

楷伦电子
PROPLUS

芯片设计

NVIDIA

AMD

intel

晶圆制造

GLOBAL
FOUNDRIES

tsmc

SMIC

力积电

Tower
Semiconductor

封装组配

Amkor
Technology

TF 通富微电

SPIL 矽品精密
SiliconwareKYEC
The Testing Industry Network

集成测试

ASUS

msi

ASRock

Gisdom

终端用户

运营商

Tencent

Google

Google

HUAWEI

云厂商

at&t

verizon

BT

中国联通

中国移动

...

PC

Lenovo

acer

ASUS

产业链上游:

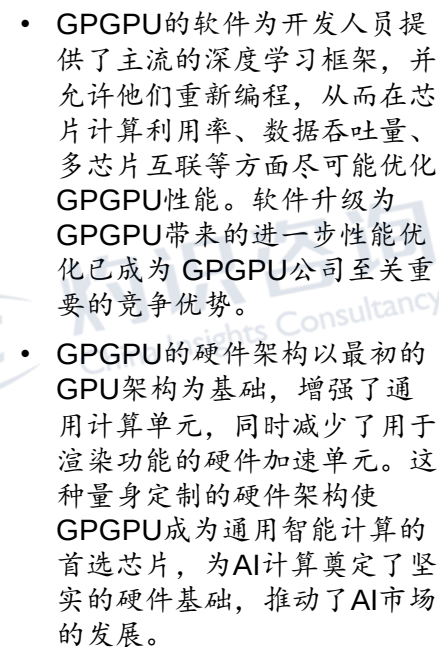
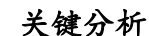
- 半导体IP供应商提供搭建SoC所需的核心功能模块。
- 硅片设计服务供应商提供各个研发环节部分或全部的研发服务及后续晶圆制造、封装及测试的委外管理
- EDA工具供应商则主要提供芯片设计所需的自动化软件工具等。

产业链中游:

- 芯片设计指根据芯片规格要求，通过主系统涉及、逻辑设计、电路设计、物理设计，最终形成设计版图。
- 晶圆制造环节是根据设计版图将电路图形信息蚀刻在硅片上，并形成电路的过程。
- 封装则是将芯片在框架上布局、粘贴、固定及连接，引出接线端子并通过可塑性绝缘介质灌封固定，构成整体立体结构的工艺。
- 封装完成后，再对芯片进行功能和性能测试。

产业链下游:

- 下游主要为终端厂商，如运营商、云厂商、AI厂商、互联网、行业AI、政府AIDC等。如今，GPGPU已被用于越来越多的新兴领域，推动各行业快速发展。



- GPGPU应用
- GPGPU软件
- GPGPU硬件
- 其他硬件

计算能力、热设计功耗、能效比、制程、内存和内存带宽等是GPGPU的关键参数。

GPGPU性能指标示意（以英伟达A100为例）

1 计算能力

- 624 TOPS INT8
- 312 TFLOPS BF16
- 156 TFLOPS TF32

2 热设计功耗

- 400W

3 能效比

- 1.6 TOPS/W INT8
- 0.8 TFLOPS/W BF16
- 0.4 TFLOPS/W TF32

4 制程

- 7nm

5 内存

- 80GB HBM2E Memory

6 内存带宽

- 1,935 GB/s



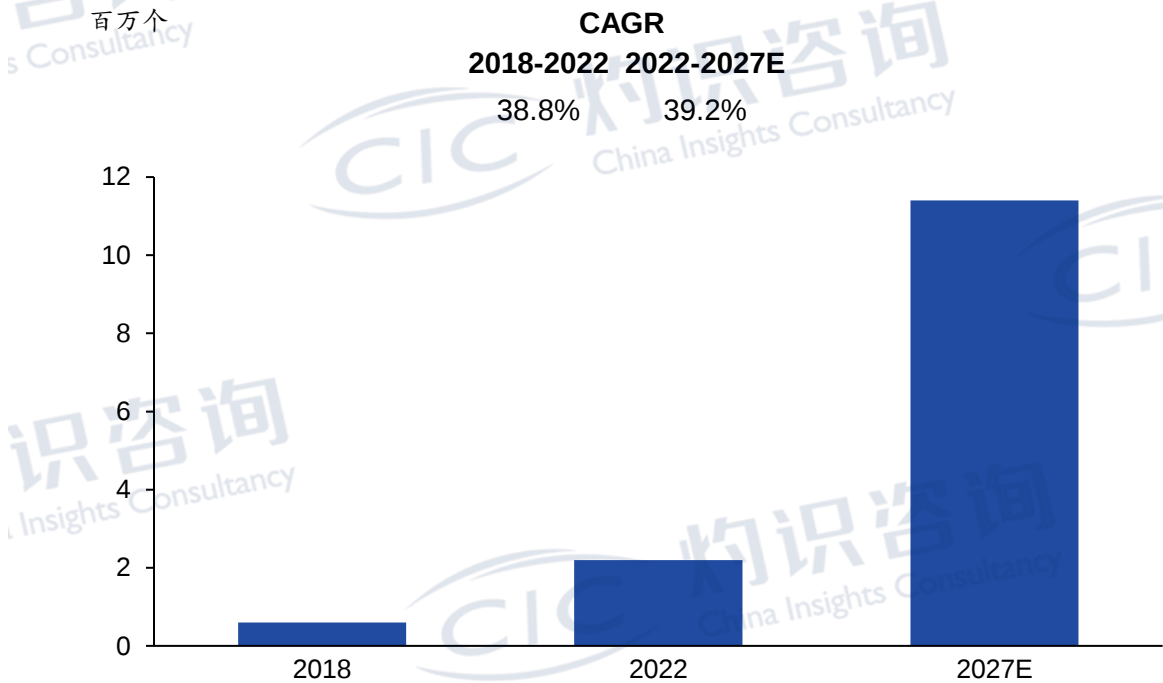
资料来源：灼识咨询

GPGPU关键参数

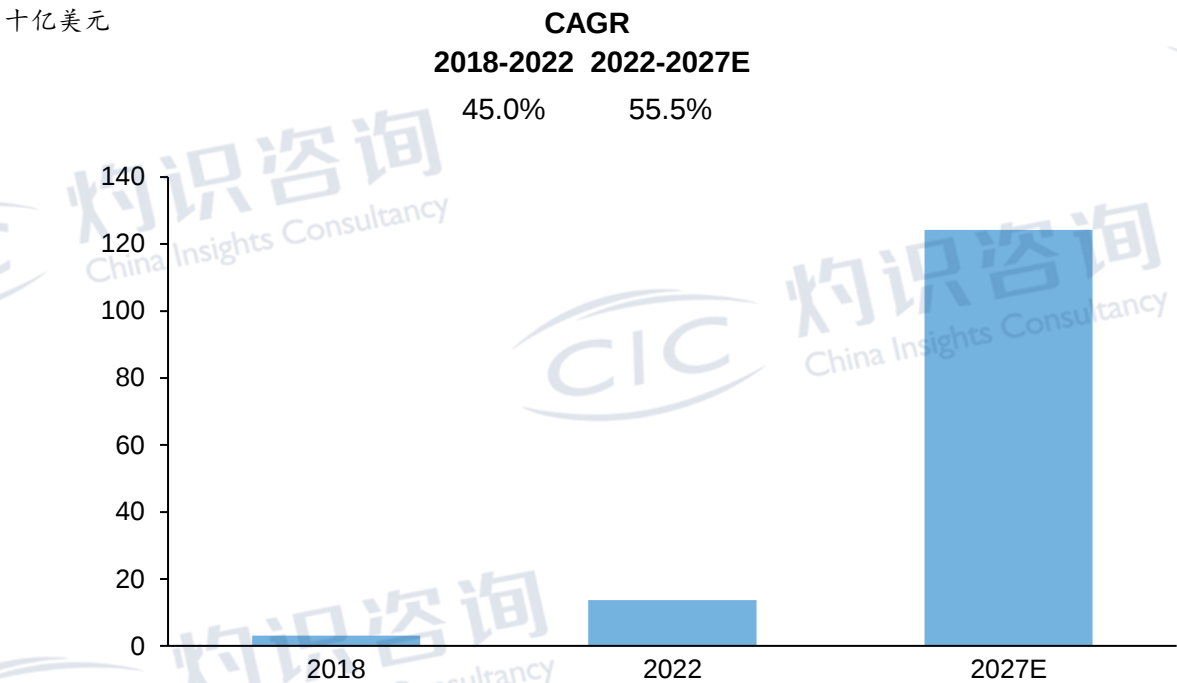
关键参数	描述
TF32, TFLOPS	TFLOPS 是指每秒一万亿次浮点运算次数。
1 BF16, TFLOPS	TF32指TensorFlow-32，指一种单精度浮点格式。BF16指Brain Float-16，指一种数据精度低于TF32的浮点数据格式。INT8指8位二进制数据格式。
INT8, TOPS	数字越大，每种格式下的芯片计算能力越强。
2 W	热设计功耗是指GPGPU的最大功耗。
TF32, TFLOPS/W	
3 BF16, TFLOPS/W	一般将计算能力除以热设计功耗计算得出，反映了在相同功耗下GPGPU可提供的计算能力。
INT8, TOPS/W	
4 nm	- 用于制造芯片晶片的核心工艺。 - 数字越小，工艺技术越先进，性能也越强。
5 GB	- 内存是指GPGPU可以临时存储和快速访问的数据大小。 - 内存越大，GPGPU可同时处理的数据就越多，性能也就越强。
6 GB/S	- 内存带宽是指芯片与内存之间的数据传输速度。 - 内存带宽越大，GPGPU 从内存读取数据的速度就越快。

GPGPU技术在大模型、AI和其他领域的应用日趋成熟，推动全球GPGPU市场持续快速发展。

全球GPGPU出货量, 2018-2027E



全球GPGPU市场规模, 2018-2027E



关键分析

- 尽管GPGPU最初是为了完成图像相关的任务而发明的，但它的架构并非是定制的，这使它们能够广泛应用于各种不同的计算任务，在计算场景不断发展和多样化的当今世界，这是一个至关重要的优势。因此，GPGPU已成为通用计算芯片的主流选择，并将在未来继续保持这一优势。
- 推理的算力需求通常会随着用户使用AI模型次数的增加而增加。受益于AI模型用户的不断增长，推理所需的计算能力未来将持续增长。
- 同时，AI模型的参数越多，完成一次训练所需的计算能力就越高。随着大模型不断发展，对能够处理此类大规模计算的GPGPU的需求将越来越大。未来，大模型的在各行业应用更加多样化，长期促进训练所需的计算能力的增长，推动GPGPU市场的发展。

资料来源：灼识咨询



I. 大模型如今已展现出卓越的能力，随着其不断演进和完善，必将引领AI的未来道路，开创智能时代的新纪元

II. 受益于大模型技术的日益成熟，AIGC有望掀起新一轮产业革命，开启人类发展的智能新时代

III. GPGPU高度并行的计算资源为大型模型的训练和推理提供了不可或缺的支持，推动了大模型与AI领域的不断进步和创新

IV. 彩蛋：生成式AI如何提升工作效率

市面上AI工具五花八门，有文字、图像、音视频、编程、办公等众多类型。**GPT爆火后**，各行各业都在开发大模型，有些是蹭热度，有些是基于已有产品进行升级，但目前具有颠覆性的仍是我们熟知的**GPT、Midjourney**等文字、图像类生成式AI工具。

1 文字类（对话、写作等）



2 图像类



3 音视频类



4 编程类



5 办公类



注：以上截图来自网站- <https://ai-bot.cn/>。

文字 - 常用大模型帮助撰写一些分析，如行业驱动因素等，经过简单改写或让AI基于之前的回复改写后，生成的内容能够达到及格水平；也会用来翻译中英文、改写并总结文字。以修改英语为例，虽然简单但仍有一定的技巧，合适的prompt可以更高效地完成工作。

Prompt 1

要求:你是一个英语为母语且是一个专业的分析师。你现在要修订报告中的内容，语言是英文。我会把每一段文字复制给你，你需要检查语法、错别字、表达不当，返回给我修订后的版本

Prompt 1: In February 2020 we set out our ambition to be a net zero company by 2050 or sooner and to help the world get to net zero. This ambition is supported by 10 aims: five to help us become a net zero company, and five to help the world meet net zero.

Prompt 2: We believe that our net zero ambition and aims, taken together, set out a path for bp that is consistent with the goals of the Paris Agreement.

回复



In February 2020 we set out our ambition to be a net zero company by 2050 or sooner and to help the world get to net zero. This ambition is supported by 10 aims: five to help us become a net zero company, and five to help the world meet net zero.



In February 2020, we articulated our ambition to become a net-zero company by 2050 or sooner, and to assist the world in achieving net zero. This aspiration is underpinned by 10 objectives: five focused on transforming us into a net-zero entity, and another five dedicated to aiding the global transition to net zero.



We believe that our net zero ambition and aims, taken together, set out a path for bp that is consistent with the goals of the Paris Agreement.



We believe that our net-zero ambition and objectives, when considered collectively, outline a trajectory for BP that aligns with the goals of the Paris Agreement.

Prompt 2

要求:你是一个英语为母语且是一个专业的分析师。你现在要修订报告中的内容，语言是英文。我会把每一段文字复制给你，你需要检查语法、错别字、表达不当，返回给我一个修订后的版本，再返回给我一个修订后的版本中体现所有修改痕迹，删除的部分加上删除线，更新的部分加粗显示。必须按照我的要求体现所有修改痕迹。

Prompt 1: In February 2020 we set out our ambition to be a net zero company by 2050 or sooner and to help the world get to net zero. This ambition is supported by 10 aims: five to help us become a net zero company, and five to help the world meet net zero.

Prompt 2: We believe that our net zero ambition and aims, taken together, set out a path for bp that is consistent with the goals of the Paris Agreement.

回复



In February 2020 we set out our ambition to be a net zero company by 2050 or sooner and to help the world get to net zero. This ambition is supported by 10 aims: five to help us become a net zero company, and five to help the world meet net zero.



Revised Version:

In February 2020, we established our goal to become a net-zero company by 2050 or earlier, while also contributing to global efforts to achieve net-zero emissions. This goal is underpinned by 10 objectives: five focused on transforming us into a net-zero company, and another five aimed at assisting the world in reaching the net-zero target.

Version Showing Changes:

In February 2020, we **set out established** our ambition **goal** to be a net zero company by 2050 or sooner **earlier**, and to **help while also contributing to** the world **get to achieve** net zero. This ambition **goal** is supported by **underpinned by** 10 aims **objectives**: five to **help us become focused on transforming us into** a net zero company, and five to **help another five aimed at assisting** the world **meet in reaching** net zero.

文字 - 在使用文字生成式AI时，promot要清晰、目的明确、且与话题相关，主要有帮助翻译、发散思路、和扩写想法三种提高效率的方式。

帮助翻译

1 角色赋予

给予它一个明确身份帮助它使用你想要的语序和表达习惯

“假如你是一个英国人，请帮我翻译‘...’”



Certainly! In English, that phrase would be translated as: "ChatGPT can provide reasonable and effective assistance during its operation."

2 鼓励反馈

鼓励 ChatGPT 提出问题或提供反馈，以保持对话的互动性和吸引力。

“从母语为英语表达者的习惯来看，这个句子有什么问题吗？可以帮我修改吗？”



从母语为英语的表达者的角度来看，原句没有明显的语法错误，但可以稍作修改以更流畅地表达：

"ChatGPT can offer practical and effective assistance during its operation."

3 针对修改

具体、有针对性的提示来关注语言的具体方面，如语法、词汇或发音

“我不喜欢effective这个词，帮我用一个高级一些的词汇替换它”



当然，请使用更高级的词汇来替换 "effective"，您可以选择使用 "efficacious"，这是一个更正式和高级的表达方式：

"ChatGPT can provide practical and efficacious assistance during its operation."

发散思路

1 保持正轨

误解或者跑偏时直白明确地指出错误，它可以很快做出修正，

“我想要的不是.....，而是.....”



抱歉对不起，我误解了你的请求。以下是一首中文诗：

2 扩展性提问

- 要求它对其中一条做出解释
- 扩写其中你认为较好的一条

“请展开详细解释一下.....这一条.....”



"大量参数"是指模型具有大量需要通过训练数据进行学习的可调整权重。这些参数用于定义模型的结构，以便能够更好地拟合训练数据和进行准确的预测。下面详细解释"大量参数"这一特征：

3 持续补充

可以要求他补充发散一些灵感，这个补充可以是没有目的的，也可以是针对某个方面进行发散

“我还想听到五个有关如何指导chatgpt给出我想要的文风的原则”



指导ChatGPT产生特定文风的关键原则包括：

- 明确表达：**在提问或指导时，尽量清晰明确地表达你的意图。使用具体的语言和示例，以确保ChatGPT能够更准确地理解你的期望。
- 简明扼要：**尽量保持指导简洁和扼要。避免过于复杂或模糊的表达，以减少歧义，并使ChatGPT更好地理解你的指导。

扩写想法

1 规定关键词

- 明确你希望它在扩写的过程当中提到的关键词或扩写方向
- 你可以明确告诉它一些行业术语从而确保它生成的内容与特定领域相关

“请写一段描述说明‘...’，我希望这段描述中包括ChatGPT在工作中的应用场景和使用后提高工作效率的原因”

2 提供样例句子

你可以要求他进行模仿描写，帮助它更熟悉你想要的文字风格和内容方向

“请仿照”烹调锅具的多元化在满足更多厨房场景需求的同时，也成为了带动全球烹调锅具市场发展的重要增量来源之一。”帮我写一段话描述“...”

3 善用总结

一般chatgpt都会给你一段很长的内容，但不需要在一开始就过多的限制她的长度，善用总结

“把这段话缩写到10字以内”

“再优化一下语言，使得表述简洁明确没有重复内容”

二手 – New Bing可以联网，功能强大且UI更好，但回复较慢；虽然GPT-4的训练数据还是基于21年，但可以通过插件联网收集二手信息，只是效果一般，可以基于二手信息再用GPT-4的其他功能进行处理。

Prompt (New Bing)

Prompt 1: 帮我列出下面21家的咨询公司的信息并返回表格，第一列是企业名称，第二列是成立年份，第三列是总部所在地区，第四列是企业简介（企业简介不要超过30字），第五列是企业的官网（如果没有就填写None）。这些公司是：Tata Group, Deloitte, PricewaterhouseCoopers, Accenture, Ernst Young, Cognizant, Infosys Public Services, Boston Consulting Group, McKinsey Company, CGI Group, Booz Allen Hamilton, Bain Company, Gartner, Mercer, Capgemini Invent, KPMG, Oliver Wyman, A.T. Kearney, Roland Berger, Arthur D. Little, China insights consultancy

Prompt 2: 进入前三个公司的网站，总结他们的成立时间、总部所在地、发展历史、主要产品。返回表格就可以，第一列公司名称，第二列成立时间、第三列总部所在地、第三列发展历史（不要超过30字）、主要产品（不要超过30字）。

回复

1	Roland Berger	1967年（成立）	德国慕尼黑市	全球最大的独立咨询公司之一，提供战略和运营等服务。	Roland Berger
	Arthur D. Little	1886年（成立）	美国马萨诸塞州波士顿市	全球最早的咨询公司之一，提供战略和技术等服务。	Arthur D. Little
	China insights consultancy	2011年（成立）	中国北京市朝阳区酒仙桥路甲10号院2号楼B座3层301室	提供商业咨询和研究等服务。	China insights consultancy

好的，我会为您提供这些公司的信息。请注意，我将只提供前三个公司的信息。

企业名称	成立年份	总部所在地区	发展历史	主要产品
Tata Group	1868年	印度孟买	印度最大的跨国企业之一，涉及多个行业，包括钢铁、汽车、信息技术、能源、化学品、饮料和酒店等。	钢铁、汽车、信息技术、能源、化学品、饮料和酒店等。
Deloitte	1845年	美国纽约市	四大会计师事务所之一，提供审计、税务、咨询和风险管理等服务。	审计、税务、咨询和风险管理等服务。
PricewaterhouseCoopers	1998年（合并）	英国伦敦	四大会计师事务所之一，提供审计、税务、咨询和风险管理等服务。	审计、税务、咨询和风险管理等服务。

请注意，这些公司的官方网站可能会有更详细的信息。

Prompt (GPT-4通过web pilot插件联网)

Prompt 1: 帮我列出下面21家的咨询公司的信息并返回表格，第一列是企业名称，第二列是成立年份，第三列是总部所在地区，第四列是企业简介（企业简介不要超过30字），第五列是企业的官网（如果没有就填写None）。这些公司是：Tata Group, Deloitte, PricewaterhouseCoopers, Accenture, Ernst Young, Cognizant, Infosys Public Services, Boston Consulting Group, McKinsey Company, CGI Group, Booz Allen Hamilton, Bain Company, Gartner, Mercer, Capgemini Invent, KPMG, Oliver Wyman, A.T. Kearney, Roland Berger, Arthur D. Little, China insights consultancy

Prompt 2: 你需要回复一个表格就可以，你发太多东西了

Prompt 3: 进入前三个公司的网站，总结他们的成立时间、总部所在地、发展历史、主要产品。返回表格就可以，第一列公司名称，第二列成立时间、第三列总部所在地、第三列发展历史（不要超过30字）、主要产品（不要超过30字）。

回复

2	Roland Berger	1967	德国慕尼黑	全球领先的独立咨询公司，为客户提供管理咨询服务	链接
	Arthur D. Little	1886	美国马萨诸塞州波士顿	全球最早的管理咨询公司，专注于链接战略、创新和转型	链接
	China insights consultancy	无法获取	无法获取	无法获取	无法获取

公司名称	成立时间	总部所在地	发展历史	主要产品
Tata Group	1968	印度孟买	从1968年开始，Tata Group已经发展成为提供信息技术、咨询和业务解决方案服务的全球领导者	提供信息技术、咨询和业务解决方案服务
Deloitte	1845	英国伦敦	Deloitte自1845年以来，已经发展成为全球领先的专业服务网络	提供审计、咨询、税务和咨询服务
PricewaterhouseCoopers	1998	英国伦敦	PwC自1998年以来，已经发展成为全球最大的专业服务公司之一	提供审计、税务、咨询和其他专业服务

以上信息来源于各公司的官方网站。

二手 – Perplexity是基于大模型的应用，做二手非常强大，而且可以试用。

GPT4



According to the IPCC report, with the aim to limit global warming to 1.5°C, global carbon emissions need to be kept within 500 billion tons from 2020 to the mid-century. Is this true?



Used KeyMate AI Search

According to the information available, the statement that global carbon emissions need to be kept within 500 billion tons from 2020 to mid-century to limit global warming to 1.5°C is not explicitly confirmed by the IPCC report. The IPCC does emphasize the urgency of reducing carbon emissions to limit global warming to 1.5°C, but the specific figure of 500 billion tons does not appear to be a direct recommendation from the IPCC.

It's crucial to note that the IPCC reports are comprehensive and cover a range of scenarios and recommendations. They often provide a range of carbon budgets that are compatible with different temperature targets, taking into account various uncertainties. Therefore, while the goal is to limit global warming, the exact carbon budget may vary depending on multiple factors, including technological advancements, policy measures, and international cooperation.

Would you like to delve deeper into this topic for more precise information? [Source](#)

Claude2

According to the IPCC report, with the aim to limit global warming to 1.5°C, global carbon emissions need to be kept within 500 billion tons from 2020 to the mid-century. Is this true?

I do not have enough context to definitively confirm or deny that specific claim from the IPCC report. However, here is some relevant information on the IPCC's statements about global carbon budgets:

- In 2018, the IPCC released a special report on the impacts of global warming of 1.5°C. In it, they estimated the remaining global carbon budget from 2018 onwards to limit warming to 1.5°C with a 66% probability at around 420 billion tons of CO₂.
- For a 50% probability of limiting warming to 1.5°C, the remaining carbon budget from 2018 was estimated to be around 580 billion tons of CO₂.
- These budgets are for CO₂ emissions only, not all greenhouse gases. When factoring in non-CO₂ emissions, the total carbon budget is reduced somewhat further.
- The IPCC carbon budgets generally refer to cumulative global CO₂ emissions from a certain start date. Your question seems to refer specifically to the 2020-2050 period, which I do not have exact numbers for from the IPCC reports.

So in summary, the IPCC has estimated a global carbon budget well below 500 billion tons of CO₂ to have a good chance of meeting the 1.5°C target, but I cannot confirm or deny the exact 500 billion ton figure stated in your question without more specifics on where that number comes from. The IPCC budgets are however in a similar range for the critical period leading up to mid-century.

Copy Retry

Perplexity

According to the IPCC report, with the aim to limit global warming to 1.5°C, global carbon emissions need to be kept within 500 billion tons from 2020 to the mid-century. Is this true?

✧ Copilot

Q Searching web

IPCC report global warming 1.5 degrees carbon emissions limit 500 billion tons 2020 mid-century

Considering 6 sources

Sources

Global Warming of 1.5 °C
— - IPCC

ipcc • 1

Chapter 2 — Global
Warming of 1.5 °C - IPCC

ipcc • 2

[PDF] Climate Change
2022 - IPCC

ipcc • 3

Chapter 3 — Global
Warming of 1.5 °C - IPCC

ipcc • 4

Summary for
Policymakers — Global...

ipcc • 5

[PDF] Climate Change
2021 - IPCC

ipcc • 6

Answer

Yes, according to the IPCC report, the current central estimate of the remaining carbon budget from 2020 onwards for limiting warming to 1.5°C with a probability of 50% has been assessed as 500 billion tons (GtCO₂)³. This means that to have a 50% chance of limiting global warming to 1.5°C, global carbon emissions should be kept within 500 billion tons from 2020 to mid-century.

🚩 [] 🗑️ ✎️

图片 - 日常工作中，常找一些封面页的背景图片，生成式AI可以生成非常好且合适主题的图片；另外，常常要找一些icon和插画，生成式AI可以生成简单的icon和插画。当然，使用Midjourney需要一定的技巧，目前上手有一定的难度。

1 找到合适的Prompt



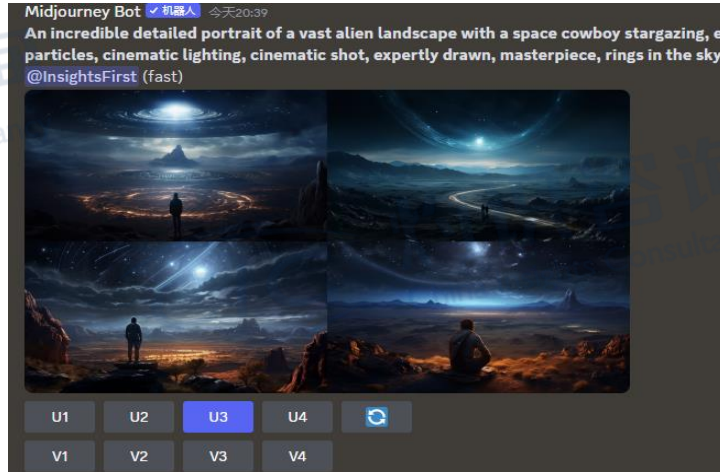
An incredible detailed portrait of a vast alien landscape with a space cowboy stargazing, 8k, extreme detail, photorealistic, dirt particles, cinematic lighting, cinematic shot, expertly drawn, masterpiece, rings in the sky

Negative prompt: lowers, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worst quality, low quality, normal quality, jpeg artifacts, signature, watermark, username, blurry, artist name, deformed, missing limb, poorly drawn hands, floating limbs, malformed hands, out of frame, too many eyes, too many limbs, too many legs, too many arms, too many nostrils, multiple nostrils, model, plastic

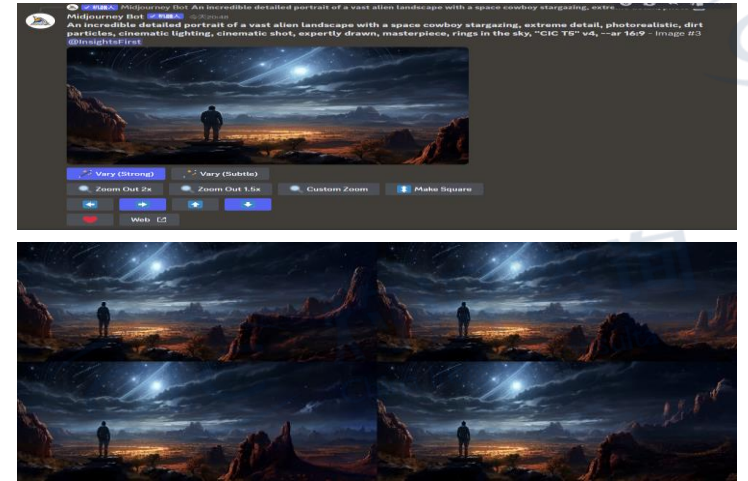
Create Copy Share



2 把Prompt发给Midjourney



3 微调并下载



1 找到合适的Prompt并替换内容



Prompt: Simple logo designs that represent children's symbols can only be designed by Line, plane 2d, vector, company logo, plane, clean, Concise, combined with the geographical characteristics or iconic buildings of Chongqing.

Prompt: Generate a set of clean, corporate - style icons representing various buildings. Please include office buildings, skyscrapers, and other commercial structures. The icons should have a professional look and feel, with straight lines, minimalist design, and consistent proportions. The color palette should be neutral and muted, suitable for corporate branding. Each icon should be easily recognizable and distinguishable, portraying the essence of the respective building type.

2 把Prompt发给Midjourney

Prompt:
Generate a clean set of oil production industry icons representing the various segments of the oil production chain. Icons should have a professional look and feel with straight lines, minimalist design and consistent proportions. The color palette should be neutral and muted, suitable for a consultant's output report. Each icon should be easily recognizable and distinguishable, reflecting the essence of the respective segment.

Prompt:
Generate 4 illustrations of the oil extraction industry, illustrations should have a professional look and feel with straight lines, minimalist design and consistent proportions. The color palette should be neutral and muted, suitable for a consultant's output report. Each illustration should be easily recognizable and distinguishable, reflecting the essence of the respective segment.

3 微调并下载



行业分析 – MindOs是一个基于大模型的应用，有一个角色是“行业分析师”，告诉它想要分析哪个行业即可生成报告。生成的报告质量不敢恭维且很模板化，但思路很有意思，即先通过网页搜索，找到最合适的信息源，再总结输出为报告。

思考过程

Thought Process ^

Generating reports for you: Solar PV, China

Preparing infographics about Solar PV...

1 result

"market size of Solar PV industry in China infographic"

✓

Searching for different aspects of Solar PV...

1 result

Solar PV industry overview China, Solar PV industry forecast China, Solar PV competitive landscape China, Risks and challenges in Solar PV industry China, Solar PV industry growth potential China

✓

Searching for infographics: "market size of Solar PV industry in China..."

10 results ✓



Searching for["Solar PV industry overview China","Solar PV industry forecast China","Sola...

Visualizing market size report...

1 result

Report: Solar PV Market Share Tables

✓

Finding best sources...

1 result

<https://www.china-briefing.com/news/powering-the-future-investing-in-chinas-photovoltaic-industry/>

✓

Extracting webpages' content...

Deep diving into industry overview...

1 result

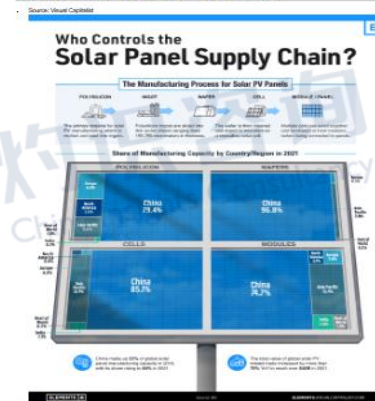
注：以上截图来自网站 - <https://mindos.com/>。

输出的报告

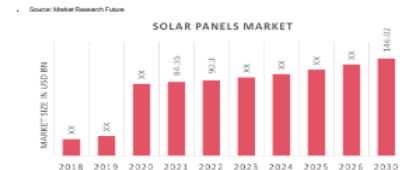
Industry Report of Solar PV

Report: Solar PV Market Share Tables

1. China's Dominance in the Solar Panel Supply Chain



2. Solar Panels Market Size, Share, Industry Growth 2023-2030



3. Solar Photovoltaic Market size to grow by USD 47.16 billion ...



Industry Report

Solar PV Industry Overview in China

Overview of the Solar PV Industry in China

China has emerged as a dominant force in the global photovoltaic (PV) industry, showcasing remarkable achievements in both installed capacity and manufacturing prowess. The PV industry harnesses solar energy through photovoltaic conversion, and China, with its abundance of solar resources, has become a thriving hub for PV production and innovation.

China's Position in the Global PV Industry

China has established itself as a global leader in the PV industry, with a significant share of the global production capacity. By the end of 2022, China installed more than four times as much solar PV capacity as the second-largest market, the United States. China's additions in 2022 surpassed the combined capacity added by the other top nine markets. The country accounted for 40% of the total annual PV installations worldwide in 2022 and held a 34% share of the world's total solar capacity.

Advantages of Investing in China's PV Industry

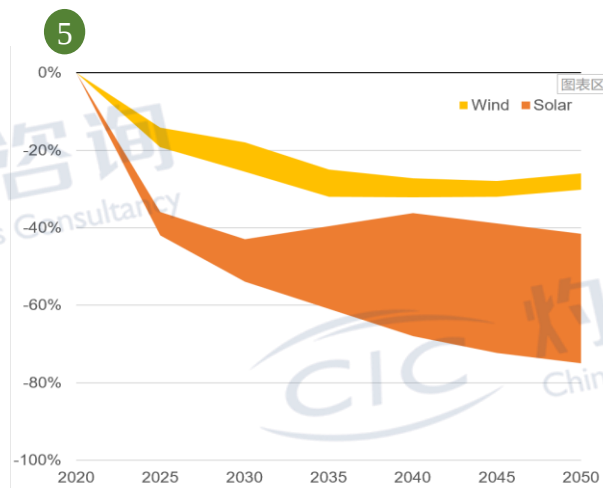
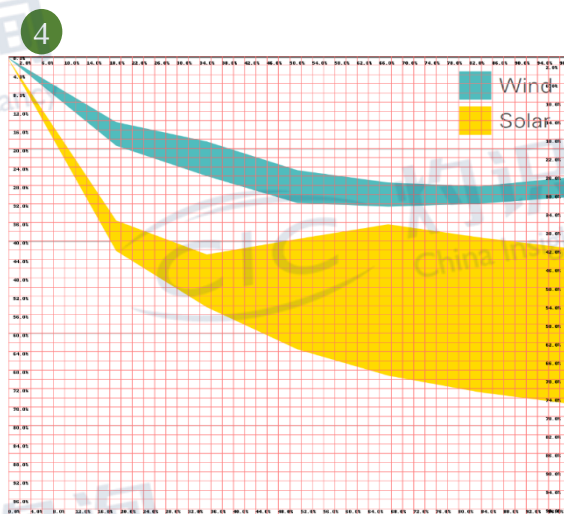
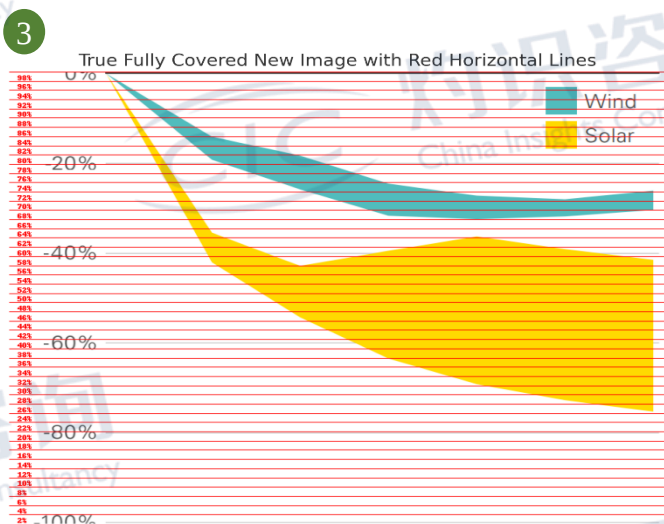
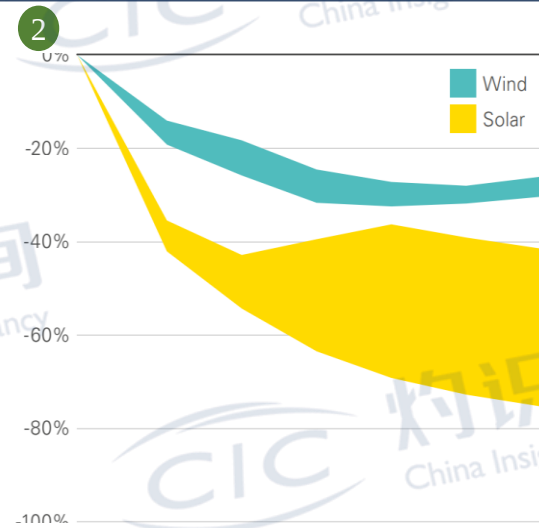
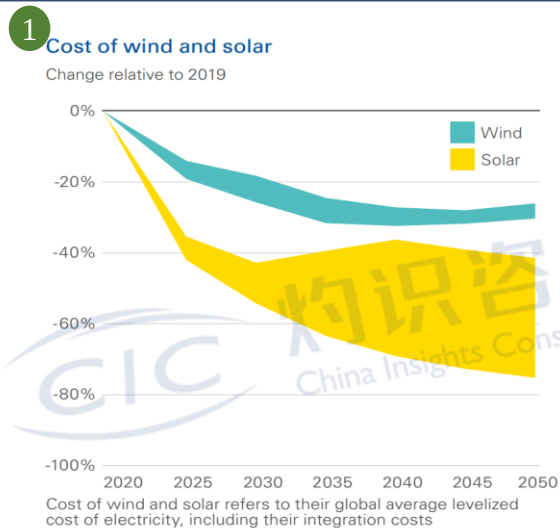
Abundant Solar Energy Resources

China possesses abundant solar energy resources, primarily concentrated in regions such as Inner Mongolia, Gansu, Qinghai, and Tibet. Approximately 68.8% of the land in China has the potential to be cost-effective sites for solar power stations. This availability of solar resources makes China an ideal region for PV industry development.

High Level of PV Investment Intensity

图像处理 - 有个比较难用肉眼识别的图想复现一下，基于GPT-4的Advanced Data Analysis可以给图像加上横线、竖线，帮助判断具体数字。

“请在我上传的图像上添加50条红色半透明横线，以便于我手动估算图表中的数值。横线应该从图像顶部到底部精确均匀分布，每一条线的间隔必须精确的为100%除以横线的数量，从0%到100%全覆盖。竖线应该从图像左部到右部均匀分布，每条线的间隔应精确的为100%除以竖线数量，从0%到100%全覆盖。请确保每条横线和竖线的位置是精确计算出来的，以避免任何不等间距的情况。横线的标签应该一半一半分布在图像的左侧和右侧以避免视觉混乱，竖线的标签一半一半分布在上方和下方。处理好的图像直接生成出来让我看见并给我一个附上下载链接。”



生成式AI能够提升我们的工作效率，但目前仍有不少问题，如：1) 应用端不成熟；2) 无法突破数据、敏感性的限制；3) 使用受限；4) 对用户要求较高。

1

应用端还不成熟

- 需要不断探索方法并总结和标准化才能真正的提升效率，因为很多场景对我们日常工作是低频的，例如画图，想要通过生成式AI生成符合我们要求的图需要不断调试，前期要花费大量的时间。

2

无法突破数据、敏感性的限制

- 以做二手为例，底层逻辑是通过企查查、天眼查、维基百科等发起请求并给予查询到的内容进行整理，很多信息在搜索引擎上查找起来也是比较费劲的
- 以MindOS为例，它所查询的数据基本是非常公开的数据，而很多我们常用的数据在公开渠道查找比较困难，未来即使有行业大模型，受版权等限制这些数据仍不会是公开数据，或许私有化的垂类大模型能够做到整合私有化的数据并提供输出

3

使用受限

- 我们虽然都自带“魔法”，但想要使用GPT-4、Midjourney十分困难
 - 首先，这2家对“魔法”要求比较高，得有特定的线路和支持全局的“魔法”
 - 其次，付款限制严格，GPT-4不支持中国境内发行的卡（包括外币卡）
 - 再次，价格高，GPT-4每月\$20，Midjourney最基础版本每月\$10
- 国内目前也已经推出了不少大模型，在文字输出上尚可，其他方面的能力略有不足。

4

对用户要求较高

- 除去国内特有的一些限制及价格较高的限制外，目前生成式AI实际对用户知识边界仍有不低的要求，增效仅局限于已有的知识，想要通过生成式AI达到不懂行业的及格线非常困难（需要不断的扩自己的知识边界才能更全面提升）
 - 以GPT-4的Code Interpreter为例，不需要用户会写代码，但想要实现更复杂的工作，还是需要用户对编程有一定的了解且能够大致读懂代码，否则使用起来非常费劲。
 - 以Midjourney为例，生成一张比较复杂的图片，需要不断调整，而且还要善于理解特定的prompt。不懂的用户其实根本就没法调整，例如普通用户对各种画风可能一无所知。



扫码关注公众号「灼识CIC」



扫码添加CIC灼识小助手

CIC灼识咨询

电话: +86 21 2356 0288

地址: 上海市静安区普济路88号静安国际中心B座10楼

如需更多信息, 请访问: www.cninsights.com

敬请致函: marketing@cninsights.com