

计算机

OpenAI 或将终止对中国的 API 服务

投资要点:

► 本周 AI 新闻速递

国内 AI: (1) 十年来 AI 领域首个国奖一等奖花落科大讯飞, 国产 AI 技术新的战略机遇展现。(2) 百度大模型应用升级, 助力企业加速数字化转型促进增收提效。

国外 AI: (1) 微软推出视觉基础模型 Florence-2, 促进 AI 视觉技术普及。(2) 全球首款 Transformer 专用集成电路(ASIC)芯片 Sohu 发布, 探索计算资源的更高效利用。(3) OpenAI 发布 CriticGPT, 审查 ChatGPT 代码输出错误。

► 宁畅发布自驱式相变液冷 AI 工作站

6月26日, 宁畅在第五届中国智算中心全栈技术大会上发布了自驱式相变液冷 AI 工作站 W350 G50, 将液冷技术应用于桌面端, 为终端算力带来新活力。面对压缩算力成本和结合 AIGC 技术创造应用价值的挑战, 宁畅提供了高效、低成本的算力解决方案。W350 G50 通过硬件优化和扩展能力提升, 将数据中心级算力带到桌面, 满足科研、制造、设计等领域的高端图形处理需求, 并针对 CAD、仿真模拟等应用进行优化。W350 G50 具备强劲性能、丰富拓展性和灵活管理特点。基于 NVIDIA RTX™ Ada 架构, 单 CPU 最高 56 核心 112 线程, 轻松处理复杂计算任务。支持全系列专业显卡, 最大 3 块双宽 GPU 卡, 满足复杂图形处理和计算加速需求。多种 I/O 接口方便连接外部设备。支持塔式和机架式切换, IPMI 远程管理, 冗余电源等, 提供服务器级专业体验。宁畅 W350 G50 是公司多年服务器研发经验的集大成之作, 从性能、功能、智能三维度定义 AI 工作站新标准。公司的"全局智算"战略和"全栈全液"AI 基础设施方案, 展示了对 AI 大模型时代需求的深刻理解, 推动行业发展和技术创新。这些方案的实施为智能算力领域带来新机遇, 支持各行业数字化转型。

► OpenAI 或将终止对中国的 API 服务

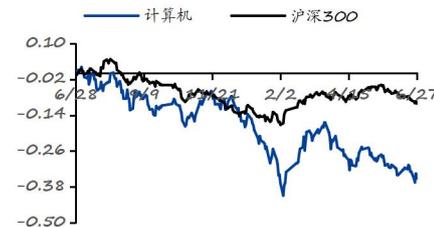
6月25日, OpenAI 通知部分开发者, 将从7月9日起阻止来自其不支持地区的 API 流量, 中国不在支持列表中, 意味着 API 服务将终止。此前, 用户可通过 OpenAI 官方 API 或微软 Azure 平台间接使用 OpenAI 技术, 但存在中断风险。消息一出, 国内大模型厂商推出迁移方案, 智谱 AI、百度智能云、阿里云等纷纷宣布零成本迁移服务, 提供替代方案和免费 tokens, 吸引用户转向国产大模型。同时, 厂商还通过降价或免费策略吸引用户。

► 风险提示

市场需求不及预期, 人工智能技术发展不及预期, 政策发布不及预期, 大模型商业落地不及预期的风险等。

强于大市 (维持评级)

一年内行业相对大盘走势



团队成员

分析师: 钱劲宇(S0210524040006)
QJY3773@hfzq.com.cn

相关报告

- 1、纯血鸿蒙 HarmonyOS NEXT 发布, 华为鸿蒙操作系统迎来全栈自研新时代——2024.06.23
- 2、特斯拉股东大会亮点频出, 武汉跟进车路云一体化示范项目——2024.06.16
- 3、英伟达披露最强 Rubin 架构, 北京百亿招标掀起车路云一体化试点应用大幕——2024.06.10



正文目录

| | |
|-------------------------------|---|
| 1 本周 AI 新闻速递..... | 3 |
| 1.1 国内 AI..... | 3 |
| 1.2 国外 AI..... | 4 |
| 2 宁畅发布自驱式相变液冷 AI 工作站..... | 5 |
| 3 OpenAI 或将终止对中国的 API 服务..... | 6 |
| 4 风险提示..... | 7 |

图表目录

| | |
|-----------------------------------|---|
| 图表 1：甄知全面重塑企业知识管理流程..... | 3 |
| 图表 2：CriticGPT 优化 ChatGPT 代码..... | 5 |
| 图表 3：宁畅 W350 G50 工作站..... | 6 |

1 本周 AI 新闻速递

1.1 国内 AI

(1) 十年来 AI 领域首个国奖一等奖花落科大讯飞，国产 AI 技术新的战略机遇展现

6月24日，2023年度国家科学技术奖揭榜，科大讯飞牵头的“多语种智能语音关键技术及产业化”项目战到最后，并获得了国家科学技术进步奖一等奖。这是深度学习引发新一代人工智能浪潮以来，过去十年人工智能领域首个国家科学技术进步奖一等奖。科大讯飞也是19项国家科学技术进步奖特等奖、一等奖中唯一作为牵头单位获奖的民营企业。“多语种智能语音关键技术及产业化”，是集合了多语种语音识别、多语种语音合成、多语种翻译技术等智能语音技术和产业化为一体的综合项目。科大讯飞从四个创新点完成了技术的突破：复杂语音信号解耦建模，破解“鸡尾酒会”难题；多语种共享建模，多语言识别测评超 OpenAI；语音语义联合建模，读懂深层含义和专业知识；国产异构硬件平台，建立国产化集群。除了技术上的突破之外，在重要评审维度“产业化”上，科大讯飞围绕多语种智能语音技术构建了自主可控的产业生态，开创引领了我国语音产业的产业链，助力中国制造出海；促进全球语言互通和“一带一路”建设；并且在进一步打造覆盖全球的国际产业生态。十年以来，AI领域项目首次获得国家科技进步一等奖，表明了国家层面对AI产业在技术创新度、转化程度、示范带动作用等方面的盖章认可。

(2) 百度大模型应用升级，助力企业加速数字化转型促进增收提效

6月25日，百度智能云宣布三款核心大模型应用产品全面升级，包括百度智能云甄知知识管理平台、百度智能云客悦智能客服平台、百度智能云曦灵数字人平台。百度智能云甄知知识管理平台将私域知识纳入模型，有助于提升企业在私域方面知识的获取效率。传统的知识管理在企业知识管理中知识的接入、加工、组织、应用四大环节的每一个环节都面临着一些难点，升级后的甄知则针对性优化相关功能，从而全面重塑企业知识管理流程；百度智能云客悦平台进一步增强了智能客服的语义理解能力，在用户意图理解、对话控制等方面进行优化，提供更加拟人化的对话体验，优化智能客服处理复杂问题的能力；曦灵数字人平台则针对开发周期长、落地成本高、高门槛等痛点进行了全面升级。

图表 1: 甄知全面重塑企业知识管理流程



来源：智东西，华福证券研究所

1.2 国外 AI

(1) 微软推出视觉基础模型 Florence-2，促进 AI 视觉技术普及

近日，微软团队放出 Florence-2 权重和代码，任何人皆可试玩。在去年 11 月 Florence-2 首次发布之初，凭借惊艳的能力引发热议。只需要一个提示，就可以指示模型完成诸如字幕、对象检测、分割等各种各样的 CV 和 CV 语言任务。Florence-2 是一种具备多种感知能力的基础视觉模型，通过构建大规模多标注数据集 FLD-5B，并进行多任务预训练，赋予了模型强大的零样本和任务迁移能力。Florence-2 在诸多视觉任务上表现出色，推进了视觉基础模型的发展。

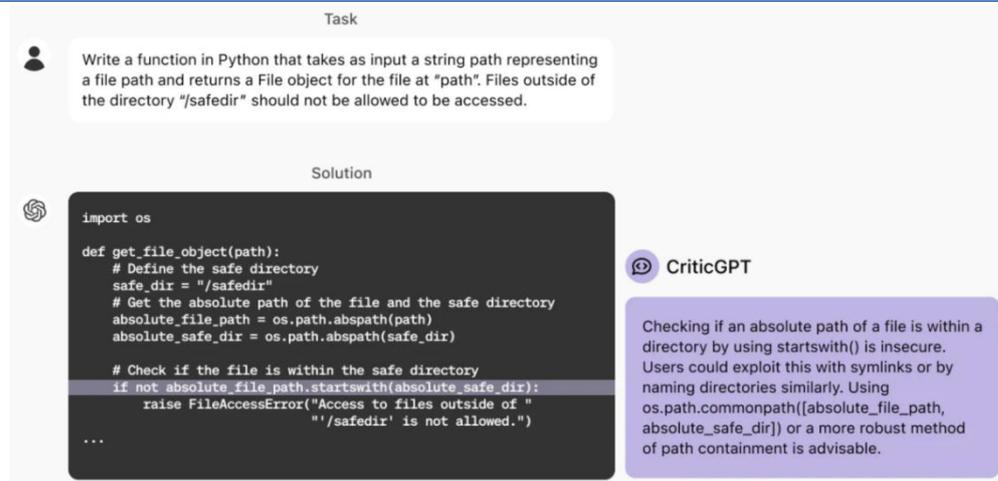
(2) 全球首款 Transformer 专用集成电路 (ASIC) 芯片 Sohu 发布，探索计算资源的更高效利用

6 月 26 日凌晨，美国芯片创业公司 Etched 推出了自己的第一块 AI 芯片 Sohu，它运行大模型的速度比英伟达 H100 要快 20 倍，比今年 3 月才推出的顶配芯片 B200 也要快上超过 10 倍。作为世界上首款 transformer ASIC (应用型专用集成电路) 芯片，一台集成了 8 块 Sohu 的服务器可以匹敌 160 块 H100 GPU。也即，Sohu 的运行速度是 H100 的 20 多倍。具体来讲，通过专门化，Sohu 具备了前所未有的性能。一台集成 8 块 Sohu 芯片的服务器每秒可以处理 50 万的 Llama 7B tokens。Sohu 拥有了更多的计算能力且利用率非常高，因此可以运行巨大的吞吐量，而不会出现内存带宽瓶颈。

(3) openAI 发布 CriticGPT，审查 ChatGPT 代码输出错误

6 月 28 日，OpenAI 发布了一款基于 GPT-4 训练的代码审查模型 CriticGPT，用于审查找到 ChatGPT 代码输出中的错误，表现比没用此工具的人要好 60%。OpenAI

开始将类似 CriticGPT 模型集成到其 RLHF 标注流水线中，为其训练师提供明确的 AI 帮助。CriticGPT 的建议并不总是正确的，但与相比没有 AI 帮助，它们有助于发现更多问题，并比模型单独工作时更少的幻觉错误。在 OpenAI 的实验中，第二个随机训练师在超过 60% 的时间里更喜欢来自 Human+CriticGPT 团队的批评，而不是来自未经 AI 帮助的人的批评。

图表 2: CriticGPT 优化 ChatGPT 代码


来源：智东西，华福证券研究所

2 宁畅发布自驱式相变液冷 AI 工作站

6月26日，宁畅在第五届中国智算中心全栈技术大会上发布了自驱式相变液冷 AI 工作站，将液冷技术从智算中心拓展至桌面端，为终端算力注入新的活力。

针对如何压缩算力成本，将先进的 AIGC 技术与实际业务场景相结合，创造有效的应用价值的难题。高效、低成本且易于落地应用的算力解决方案成为了市场的迫切期待。在这样的背景下，宁畅以桌面级 AI 工作站 W350 G50 交卷，通过对硬件的优化和对扩展能力的提升，将服务器或数据中心级别的算力带到桌面上。面向科研、制造、电影、设计、动画、建筑、机械等高端图形处理用户，宁畅 W350 G50 兼容主流的图形设计软件，并专门针对 CAD、仿真模拟、动画制作、CG 渲染、图形计算等应用进行了优化。总的来看，宁畅的桌面级 AI 工作站提供了一种创新的解决方案，通过硬件优化和扩展能力的提升，将原本属于服务器或数据中心的高性能算力，以更加经济的方式带给了中小企业，不仅响应了市场对于高效、低成本算力解决方案的需求，也为中小企业和开发者在 AIGC 领域的应用和发展提供了强有力的支持。

图表 3: 宁畅 W350 G50 工作站



来源：智东西，华福证券研究所

宁畅 W350 G50 具备三大亮点：强劲性能、丰富拓展、灵活管理。在性能方面，其基于 NVIDIA RTX™ Ada 架构，单 CPU 最高 56 核心 112 线程，能够轻松应对复杂的计算任务，无论是大规模的数据处理、深度学习模型的训练，还是高要求的图形渲染，都能迅速而高效地完成。在拓展能力方面，宁畅 W350 G50 支持全系列专业显卡，包括 RTX 5880 Ada、RTX A5000 等，具备广泛的适用性和升级潜力；最大支持 3 块全高、全长、双宽的专业图形卡或 GPU 卡，可用于复杂的图形处理和计算加速；多种形态的 I/O 接口方便用户连接各类外部设备，满足不同工作场景的特殊需求。在管理方面，宁畅 W350 G50 支持塔式和机架式灵活切换，既能够作为高端桌面工作站，也可作为微型服务器转换为数据中心模式；同时支持 IPMI 远程管理，管理员可以在异地对工作站的运行状态进行实时监测和控制，及时处理可能出现的问题，减少维护成本和停机时间；支持冗余电源、企业级部件等，能带来服务器级别的专业体验。

基于多年服务器研发、生产经验，宁畅 W350 G50 成为一款集大成的终端产品，从性能、功能、智能三维度带来“AI 工作站标杆之作”。“全局智算”战略以及“全栈全液”AI 基础设施方案，体现了宁畅对当前 AI 大模型时代需求的深刻理解，也展现了其在推动行业发展和技术创新方面的领导力。随着这些方案的深入实施和不断完善，宁畅为智能算力领域带来了新的发展机遇，也为各行业的数字化转型提供了有力的支持。

3 OpenAI 或将终止对中国的 API 服务

OpenAI 官方邮件显示 OpenAI 或将终止对中国提供 API 服务。6 月 25 日，部分开发者收到了来自 OpenAI 官方的邮件。邮件显示，“贵组织的 API（接口）流量来自 OpenAI 目前不支持的地区，我们将从 7 月 9 日开始采取额外措施，阻止来自不



在我们支持的国家和地区列表中的地区的 API 流量。”据悉，目前 OpenAI 的 API 向 161 个国家和地区开放，但中国未包含在其中。这也意味着，OpenAI 将终止对中国提供 API 服务。在利用 OpenAI 的能力方面，此前主要存在两个被广泛认可的途径。一方面，用户可以直接对接 OpenAI 官方提供的 API 接口；另一方面，则是通过与微软 Azure 平台合作，间接利用 OpenAI 的技术支持。业内人士表示，在国内直接对接 OpenAI 官方 API 以使用其能力的方式，一直存在被突然中断或封禁的风险。

在消息传开后，国内大模型厂商纷纷下场“抢食”OpenAI 份额，公布 OpenAI 用户迁移方案。智谱 AI 首先有所行动，发文宣布“面向 OpenAI API 用户提供特别搬家计划”，帮助用户切换至国产大模型。下午，百度智能云官方发文《故乡的云·大模型普惠计划》，表示“0 成本切换”，会为新注册企业用户提供 0 元调用、0 元训练、0 元迁移等服务。阿里云同样在下午四点发文称“丝滑迁移通义 API”，阿里云还表示将为 OpenAI API 用户提供最具性价比的中国大模型替代方案，并为中国开发者提供 2200 万免费 tokens 和专属迁移服务。随后，MINIMAX 同样宣布推出“零成本保姆式迁移方案”、零万物随之也面向 OpenAI 用户推出了平滑迁移至 Yi 系列大模型的服务，针对接入 OpenAI 的不同模型的用户提供替换方案。除了公布 OpenAI 用户迁移方案以“抢食”OpenAI 份额，今年 5 月，国内多家大模型厂商还通过降价或免费吸引大模型使用者。

4 风险提示

市场需求不及预期，人工智能技术发展不及预期，政策发布不及预期，大模型商业落地不及预期的风险等。



分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

一般声明

华福证券有限责任公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的信息均来源于本公司认为可信的公开资料，该等公开资料的准确性及完整性由其发布者负责，本公司及其研究人员对该等信息不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，之后可能会随情况的变化而调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

在任何情况下，本报告所载的信息或所做出的任何建议、意见及推测并不构成所述证券买卖的出价或询价，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司仅承诺以勤勉的职业态度，独立、客观地出具本报告以供投资者参考，但不就本报告中的任何内容对任何投资做出任何形式的承诺或担保。投资者应自行决策，自担投资风险。

本报告版权归“华福证券有限责任公司”所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。

特别声明

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。投资者请勿将本报告视为投资或其他决定的唯一参考依据。

投资评级声明

| 类别 | 评级 | 评级说明 |
|------|------|------------------------------------|
| 公司评级 | 买入 | 未来 6 个月内，个股相对市场基准指数涨幅在 20%以上 |
| | 持有 | 未来 6 个月内，个股相对市场基准指数涨幅介于 10%与 20%之间 |
| | 中性 | 未来 6 个月内，个股相对市场基准指数涨幅介于-10%与 10%之间 |
| | 回避 | 未来 6 个月内，个股相对市场基准指数涨幅介于-20%与-10%之间 |
| | 卖出 | 未来 6 个月内，个股相对市场基准指数涨幅在-20%以下 |
| 行业评级 | 强于大市 | 未来 6 个月内，行业整体回报高于市场基准指数 5%以上 |
| | 跟随大市 | 未来 6 个月内，行业整体回报介于市场基准指数-5%与 5%之间 |
| | 弱于大市 | 未来 6 个月内，行业整体回报低于市场基准指数-5%以下 |

备注：评级标准为报告发布日后的 6~12 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；香港市场以恒生指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准（另有说明的除外）

联系方式

华福证券研究所 上海

公司地址：上海市浦东新区浦明路 1436 号陆家嘴滨江中心 MT 座 20 层

邮编：200120

邮箱：hfjys@hfzq.com.cn