

# 上海市律师协会数字科技与人工智能专业委员会

## AI 前沿法律资讯

(2025 年 10 月)

本资讯将定期为读者分享关于人工智能（AI）领域的前沿资讯，包括 AI 国际视野、AI 国内监管、AI 行业融资、AI 技术革新等板块，本期资讯内容如下：

### 目录

一、AI 国际视野	2
1.2025 年全球人工智能发展趋势：技术加速、投资激增与治理挑战——基于《人工智能指数报告》的分析	2
2.自 2025 年 8 月 2 日起，通用目的人工智能模型提供者的义务正式生效	3
3.特朗普总统推动美国人工智能技术出口	3
4.2025 世界人工智能大会暨人工智能全球治理高级别会议发表《人工智能全球治理行动计划》	4
5.欧盟就通用型 AI（GPAI）发布一系列指导与自愿行为准则	5
二、AI 国内监管	7
1.国务院印发《关于深入实施“人工智能+”行动的意见》	7
2.9 月 1 日起，AI 生成合成内容必须添加标识	8
3.中央网信办专项整治 AI 技术滥用	9
4.中央网信办秘书局、市场监管总局办公厅联合印发《智能社会发展与治理标准化指引（2025 版）》	10
5.中央网信办、国家发展改革委近日联合印发《政务领域人工智能大模型部署应用指引》	11
三、AI 行业融资	12
1.万亿豪赌：OpenAI 结盟芯片、云巨头 AI 生态闭环了吗？	12
2.8 月人形机器人领域融资事件 21 起累计金额近 54 亿元人民币	12
3.法律工作提供自动化服务的初创公司 HarveyAI 估值增长至 50 亿美元	13
4.ReflectionAI 融资 20 亿美元，成为美国开放前沿 AI 实验室，挑战 DeepSeek	14
5.数据标签公司 SurgeAI 计划首次融资，筹集高达 10 亿美元	15
四、AI 技术革新	16
1.OpenAI 正式发布 GPT-5（统一多模态、高级推理与 agent 能力）	16
2.FigureAI（类似机器人公司）展示/报道其新一代类人/家庭用机器人进展	17
3.谷歌 DeepMind 声称在人工智能问题解决方面取得“历史性”突破	18
4.解决 LLM 限制：2025 年人工智能解决方案完整指南	19
5.OpenAI 与博通达成最新芯片交易，为其打造首款 AI 处理器	20

## 正文

### 一、AI 国际视野

#### 1.2025 年全球人工智能发展趋势：技术加速、投资激增与治理挑战——基于《人工智能指数报告》的分析



**摘要：**本文基于《2025 年人工智能指数报告》的综合数据，分析了当前全球人工智能（AI）在研究、性能、经济及治理领域的核心进展。研究显示，产业界持续主导标志性 AI 模型的研发，2024 年近 90% 的前沿模型源自产业界，而学术界在高影响力研究中仍居领先地位。

技术性能方面，AI 系统在 MMMU、GPQA 和 SWE-bench 等严苛基准测试中表现出显著进步，例如 SWE-bench 编程测试的解题能力在一年内从 4.4% 跃升至 71.7%。同时，开源模型迅速缩小与闭源模型的差距，且 AI 效率大幅提升，如达到 GPT-3.5 级别的推理成本在 18 个月内骤降逾 280 倍，极大降低了先进 AI 的应用门槛。在地缘政治格局上，美国在标志性模型数量上保持领先，但中国模型在性能质量上正迅速接近，中美模型能力差距大幅收窄。

在经济层面，全球私人 AI 投资创历史新高（同比增长 26%），生成式 AI 领域的投资激增，目前占所有私人 AI 投资的 20% 以上。企业 AI 技术采用率从 2023 年的 55% 激增至 2024 年的 78%，并开始是客户服务、供应链等职能中产生可量化的财务效益。值得注意的是，研究进一步证实 AI 能有效提高生产效率并有助于缩小劳动力技能差距。

在治理方面，尽管 AI 事件报告数量持续增加，但主流模型开发者采用标准化负责任 AI（RAI）评估的情况仍不普遍。各国政府正加速监管框架的建设，立法提及率同比增长 21.3%，同时展开大规模投资，共同塑造 AI 的未来发展方向。本报告全面揭示了 AI 快速、广泛演进的动态，及其在性能、普及和责任治理方面日益凸显的复杂挑战。

**参考链接：** <https://hai.stanford.edu/ai-index/2025-ai-index-report>

## 2.自 2025 年 8 月 2 日起，通用目的人工智能模型提供者的义务正式生效



**摘要：**自 2025 年 8 月 2 日起，通用目的人工智能模型提供者的义务正式生效。在此日期之后投放市场的通用目的人工智能模型必须合规，并需与欧盟 AI 办公室的技术团队进行非正式合作。特别是，针对最具先进性的模型（即存在系统性风险的模型），其提供者有法定义务向 AI 办公室进行通报。AI 办公室将为提供者提供合规支持，尤其是针对已签署《实践守则》的提供者。自 2026 年 8 月 2 日起，欧盟委员会的执法权力正式启用。委员会将对通用目的人工智能模型提供者的义务履行情况进行执法，包括处以罚款。至 2027 年 8 月 2 日，2025 年 8 月 2 日前已投放市场的通用目的人工智能模型提供者必须完成合规。

**参考链接：** <https://www.secrss.com/articles/81083>

## 3.特朗普总统推动美国人工智能技术出口

**摘要：**2025 年 7 月 23 日，特朗普总统今天签署了一项行政命令，旨在通过促进向全

球盟友及合作伙伴出口全栈美国人工智能技术产品，来支持美国的人工智能产业。特朗普总统正在推动美国在 AI 领域的引领地位，以确保经济增长、国家安全和全球竞争力。

AI 是一项基石技术，将在未来几十年塑造一个创新、国防和繁荣的未来。美国必须在 AI 技术、标准和治理模型的开发与部署方面发挥引领作用，以减少全球对来自敌对国家的有关系统的依赖。通过出口美国 AI 技术，美国将加强与盟友的关系，推进美国的标准和治理模式，并保持技术主导地位。这项举措通过促进对 AI 开发和基础设施的投资，支持包括小型企业在内的美国企业，确保美国在全球 AI 创新领域保持领先地位。让美国成为全球 AI 领导者：特朗普总统已将美国在 AI 领域的领导地位作为一项国家优先事项。

**参考链接：** <https://china.usembassy-china.org.cn/zh/fact-sheet-president-donald-j-trump-promotes-the-export-of-american-ai-technologies/>

#### 4.2025 世界人工智能大会暨人工智能全球治理高级别会议发表《人工智能全球治理行动计划》



## 人工智能全球治理行动计划（全文）

2025-07-26 12:54 来源：新华社

字号：默认 大 超大 | 打印 | 分享 | 收藏

新华社上海7月26日电 7月26日，2025世界人工智能大会暨人工智能全球治理高级别会议发表《人工智能全球治理行动计划》。全文如下：

### 人工智能全球治理行动计划

人工智能是人类发展的新领域，是新一轮科技革命和产业变革的重要驱动力量，也可以是造福人类的国际公共产品。人工智能带来前所未有的发展机遇，也带来前所未有风险挑战。智能时代，唯有同球共济，我们才能在充分发挥人工智能潜力的同时，确保其发展的安全性、可靠性、可控性和公平性，最终落实联合国《未来契约》及其附件《全球数字契约》有关承诺，为所有人创造包容、开放、可持续、公平、安全和可靠的数字和智能未来。

**摘要：**7月26日，2025世界人工智能大会暨人工智能全球治理高级别会议发表《人工智能全球治理行动计划》。人工智能带来前所未有发展机遇，也带来前所未有风险挑战。从技术自身看，人工智能具有不透明性、不可解释性和不可预测性，确保人工智能始终符合人类意愿，避免可能存在的失控风险，是当前必须直面的重大课题。从技术开发与应用

看，人工智能还可能给人类的生活、社会治理、伦理道德带来影响。比如，大语言模型处理、生成数据时，可能涉及个人隐私；相关技术被恶意使用，可能被用来制造虚假信息、诈骗等。兴利除弊，这需要加强技术前瞻研判，在发展和安全之间寻求平衡。

人工智能发展和应用带来的风险与挑战往往是跨越国界的，世界各国加强沟通与合作，才能在充分发挥人工智能潜力的同时，确保其发展的安全性、可靠性、可控性和公平性。

**参考链接：**

<http://www.news.cn/tech/20250804/536e869ca84648c5aaa39220ac354ed4/c.html>

## 5. 欧盟就通用型 AI（GPAI）发布一系列指导与自愿行为准则



**摘要：**该准则旨在帮助业界遵守《人工智能法案》关于通用人工智能的规则，该规则将于 2025 年 8 月 2 日生效。该规则一年后将由委员会人工智能办公室针对新模型强制执行，两年后将由委员会人工智能办公室针对现有模型强制执行。此举旨在确保投放欧洲市场的通用人工智能模型（包括最强大的模型）安全透明。

该准则由三章组成：透明度和版权，均针对所有通用人工智能模型提供商；安全和保障，仅与有限数量的最先进模型提供商相关。一旦《准则》获得成员国和委员会的认可，自愿签署《准则》的通用人工智能模型提供商将通过遵守《准则》证明其遵守了《人工智能法案》的相关义务。与通过其他方式证明合规性的提供商相比，签署《准则》的提供商将受益于更少的行政负担和更高的法律确定性。

参考链接: <https://digital-strategy.ec.europa.eu/en/news/general-purpose-ai-code-practice-now-available>

## 二、AI 国内监管

### 1. 国务院印发《关于深入实施“人工智能+”行动的意见》



中华人民共和国中央人民政府

www.gov.cn



首页 | 简 | 繁 | EN | 登录 | 邮箱

首页 > 国务院公报 > 2025年第25号

字号: 默认 大 超大

| 打印



#### 国务院关于深入实施“人工智能+”行动的意见

国发〔2025〕11号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

为深入实施“人工智能+”行动，推动人工智能与经济社会各行业各领域广泛深度融合，重塑人类生产生活范式，促进生产力革命性跃迁和生产关系深层次变革，加快形成人机协同、跨界融合、共创分享的智能经济和智能社会新形态，现提出如下意见。

##### 一、总体要求

以习近平新时代中国特色社会主义思想为指导，完整准确全面贯彻新发展理念，坚持以人民为中心的发展思想，充分发挥我国数据资源丰富、产业体系完备、应用场景广阔等优势，强化前瞻谋划、系统布局、分业施策、开放共享、安全可控，以科技、产业、消费、民生、治理、全球合作等领域为重点，深入实施“人工智能+”行动，涌现一批新基础设施、新技术体系、新产业生态、新就业岗位等，加快培育发展新质生产力，使全体人民共享人工智能发展成果，更好服务中国式现代化建设。

**摘要：**8月21日，国务院印发《关于深入实施“人工智能+”行动的意见》，提出加快实施6大重点行动。一是“人工智能+”科学技术，加速科学发现进程，驱动技术研发模式创新和效能提升，创新哲学社会科学研究方法。二是“人工智能+”产业发展，培育智能原生新模式新业态，推进工业全要素智能化发展，加快农业数智化转型升级，创新服务业发展新模式。三是“人工智能+”消费提质，拓展服务消费新场景，培育产品消费新业态。四是“人工智能+”民生福祉，创造更加智能的工作方式，推行更富成效的学习方式，打造更有品质的美好生活。五是“人工智能+”治理能力，开创社会治理人机共生新图景，打造安全治理多元共治新格局，共绘美丽中国生态治理新画卷。六是“人工智能+”全球合作，推动人工智能普惠共享，共建人工智能全球治理体系。

**参考链接：** [https://www.gov.cn/gongbao/2025/issue\\_12266/202509/content\\_7039598.html](https://www.gov.cn/gongbao/2025/issue_12266/202509/content_7039598.html)

## 2.9月1日起，AI生成合成内容必须添加标识



首页

组织机构

新闻发布

政务公开

政务服务

互动交流

工信数据

首页 > 新闻发布 > 媒体报道 > 文字报道

### 9月1日起，AI生成合成内容必须添加标识

发布时间：2025-03-15 11:43 来源：“网信中国”微信公众号、工信微报

近日，国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合发布《人工智能生成合成内容标识办法》（以下简称《标识办法》），自2025年9月1日起施行。国家互联网信息办公室有关负责人表示，《标识办法》旨在促进人工智能健康发展，规范人工智能生成合成内容标识，保护公民、法人和其他组织合法权益，维护社会公共利益。

**摘要：**国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合发布《人工智能生成合成内容标识办法》（以下简称《标识办法》），自2025年9月1日起施行。《标识办法》明确，人工智能生成合成内容标识主要包括显式标识和隐式标识两种形式，显式标识是指在生成合成内容或者交互场景界面中添加的，以文字、声音、图形等方式呈现并可以被用户明显感知到的标识；隐式标识是指采取技术措施在生成合成内容文件数据中添加的，不易被用户明显感知到的标识。《标识办法》强调，任何组织和个人不得恶意删除、篡改、伪造、隐匿本办法规定的生成合成内容标识，不得为他人实施上述恶意行为提供工具或者服务，不得通过不正当标识手段损害他人合法权益。

#### 参考链接：

[https://wap.miit.gov.cn/xwfb/mtbd/wzbd/art/2025/art\\_5e46c60f9a7141cdb584eb139f476ce9.htm](https://wap.miit.gov.cn/xwfb/mtbd/wzbd/art/2025/art_5e46c60f9a7141cdb584eb139f476ce9.htm)

1

### 3.中央网信办专项整治 AI 技术滥用



#### 中央网信办深入开展“清朗·整治AI技术滥用”专项行动第一阶段工作

2025年06月20日 17:00

来源: 中国网信网

【打印】【纠错】



**摘要:** 为规范 AI 服务和应用,促进行业健康有序发展,保障公民合法权益,中央网信办在全国范围内部署开展为期 3 个月的“清朗·整治 AI 技术滥用”专项行动。专项行动将分两个阶段开展,重点整治违规 AI 产品、重点领域安全风险、侵害未成年人权益等突出问题。专项行动第一阶段强化 AI 技术源头治理,重点整治违规 AI 产品,传授、售卖违规 AI 产品教程和商品,训练语料管理不严,安全管理措施薄弱,未落实内容标识要求等突出问题。第二阶段聚焦利用 AI 制作发布谣言、不实信息、色情低俗内容,假冒他人实施侵权违法行为、从事网络水军活动等突出问题,集中清理相关违法不良信息,处置处罚违规账号、MCN 机构和网站平台。

**参考链接:** [https://www.cac.gov.cn/2025-06/20/c\\_1752129980667315.htm](https://www.cac.gov.cn/2025-06/20/c_1752129980667315.htm)

#### 4.中央网信办秘书局、市场监管总局办公厅联合印发《智能社会发展与治理标准化指引（2025版）》



国家标准化管理委员会  
National Standardization Administration

请输入关键字



本站热词：标准 通知

首页

机构概况

新闻

信息公开

办事服务

互动交流

专题

政策文件

标题：中央网信办秘书局、市场监管总局办公厅联合印发《智能社会发展与治理标准化指引（2025版）》

文号：

发文机关：中央网信办秘书局、市场监管总局办公厅

发文日期：2025年6月11日

所属机构：中央网信办秘书局、市场监管总局办公厅

**摘要：**《指引》旨在建立健全科学合理的智能社会发展与治理标准研究制定、实施反馈、优化完善的工作机制，建成覆盖智能技术主要社会应用场景、有效保障技术全生命周期良性健康发展的标准体系，从而适应技术创新需要、满足产业发展需求、支撑智能社会建设，助力国家治理体系和治理能力现代化。《指引》提出了智能社会发展与治理的基本原则和要求，明确了常见智能技术应用场景、社会影响及其观测评估指标，规定了人工智能社会实验的一般程序和要求，构建了包括基础通用、发展与治理原则、场景应用、技术和方法、效果评价等五部分内容的智能社会发展与治理标准体系，为各地方、各部门、科研院所、企事业单位等开展智能社会发展与治理理论研究和实践活动提供技术支撑和规范指引。

**参考链接：**

[https://www.sac.gov.cn/xxgk/zcwj/art/2025/art\\_3536416555b84e22919e3d32e46460e8.html](https://www.sac.gov.cn/xxgk/zcwj/art/2025/art_3536416555b84e22919e3d32e46460e8.html)

## 5.中央网信办、国家发展改革委近日联合印发《政务领域人工智能大模型部署应用指引》

# 两部门印发《政务领域人工智能大模型部署应用指引》

2025-10-11 08:03 来源：新华社

字号：默认 大 超大 | 打印  |   

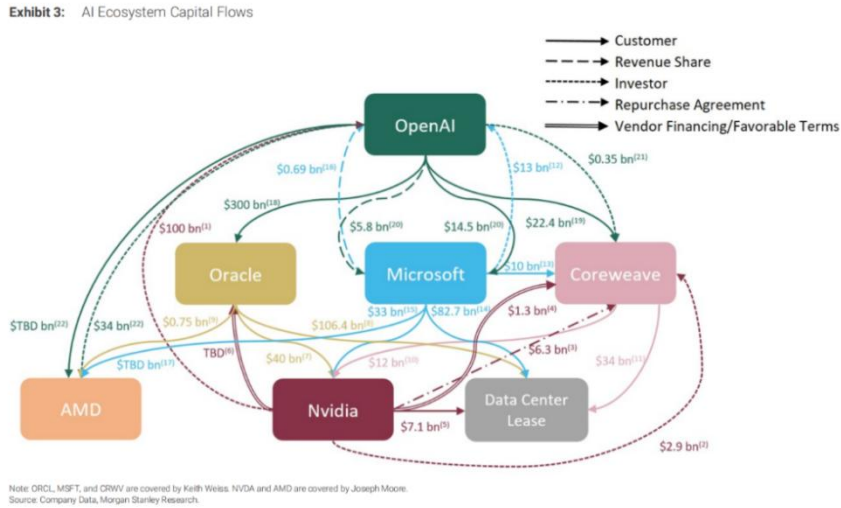
新华社北京10月10日电 记者10日从中央网信办获悉，为安全稳妥有序推进政务领域人工智能大模型部署应用，中央网信办、国家发展改革委近日联合印发《政务领域人工智能大模型部署应用指引》，从应用场景、规范部署、运行管理等方面，为各级政务部门提供人工智能大模型部署应用的工作导向和基本参照。

**摘要：**指引提出，政务部门可围绕政务服务、社会治理、机关办公和辅助决策等工作中的共性、高频需求，因地制宜、结合实际，选择典型场景进行人工智能大模型探索应用。应探索构建“一地建设、多地多部门复用”的集约化部署模式，统筹推进政务大模型部署应用，防止形成“模型孤岛”。指引指出，政务部门应统筹减负和赋能，避免盲目追求技术领先、概念创新，切实防范“数字形式主义”。应建立健全全周期管理体系，明确应用方式和边界，落实人工智能大模型“辅助型”定位，防范模型“幻觉”等风险。

**参考链接：** [https://www.gov.cn/lianbo/bumen/202510/content\\_7043864.htm](https://www.gov.cn/lianbo/bumen/202510/content_7043864.htm)

### 三、AI 行业融资

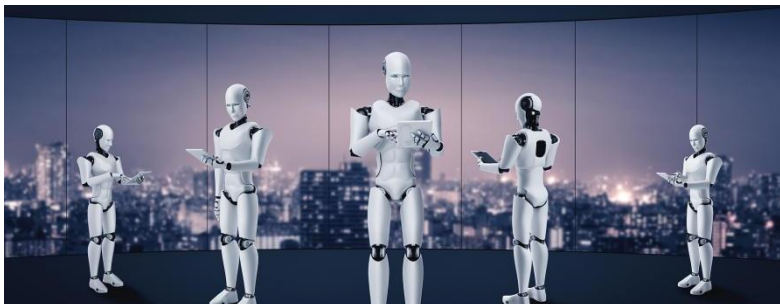
#### 1. 万亿豪赌：OpenAI 结盟芯片、云巨头 AI 生态闭环了吗？



**摘要：**10月13日，OpenAI 与博通宣布达成战略合作，将共同推出总容量达 10GW（千兆瓦）的定制芯片。OpenAI 将负责芯片设计，博通则从 2026 年下半年开始进行芯片开发与部署。作为全球估值最高的 AI 初创公司，OpenAI 已连续抛出重磅合作，与甲骨文、英伟达、AMD 等巨头签署总额约 1 万亿美元协议，覆盖算力、云服务等多个关键领域。这一系列布局预计将使 OpenAI 在未来十年获得超 30GW 的算力支撑。

**参考链接：** <https://guba.eastmoney.com/news/usnvda,1609897597.html>

#### 2.8 月人形机器人领域融资事件 21 起累计金额近 54 亿元人民币

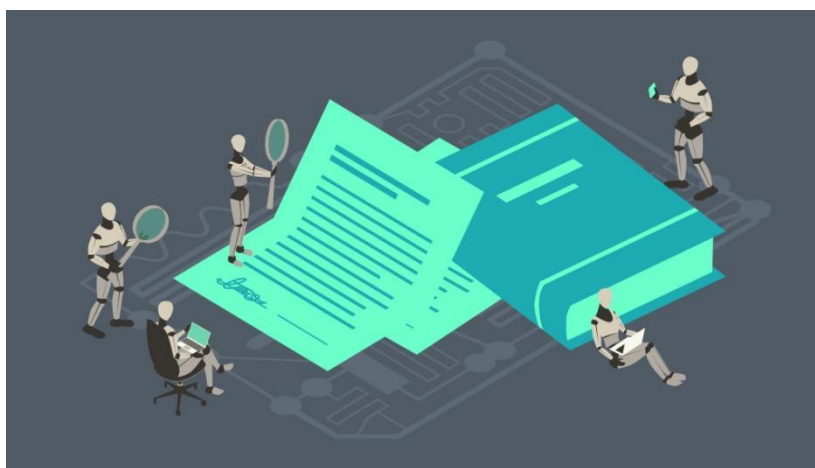


**摘要：**全球人形机器人竞赛已经进入白热化阶段，中美不同的发展路径各有优势。国际市场聚焦基础技术突破，国内则注重全产业链生态构建。这种差异化的竞争格局将为全

球人形机器人技术进步提供更多元化的解决方案，最终加速整个产业的发展进程。从规模来看，数据显示，8月国内单笔融资近5亿元的仅1起，由梅卡曼德拿下，资金将用于加速公司具身智能“眼脑手”全栈技术的持续进化，拓展产品线与场景应用，并进一步提升全球商业化与客户服务能力。梅卡曼德由清华团队创办，是国家级专精特新“小巨人”企业。从细分领域来看，8月融资覆盖本体、大模型、灵巧手、触觉/视触觉感知、运控系统、关节模组、仿生面部交互等赛道。其中，融资项目中含“本体/大模型/本体+大模型”的共有10起，融资事件数量占比50%。

**参考链接：** <https://www.stcn.com/article/detail/3324855.html>

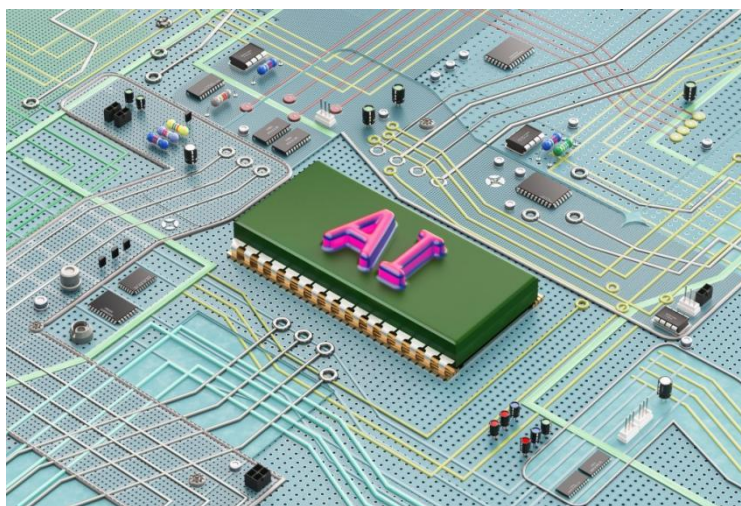
### 3.法律工作提供自动化服务的初创公司 HarveyAI 估值增长至 50 亿美元



**摘要：**为法律工作提供自动化服务的初创公司 HarveyAI 向《财富》杂志透露，该公司已完成 3 亿美元 E 轮融资，估值达 50 亿美元。本轮融资由凯鹏华盈和 Coatue 联合领投，现有投资者 Conviction、EladGil、OpenAStartupFund 和红杉资本等也参与其中。就在此次融资的四个月前，哈维宣布红杉资本领投了 3 亿美元的 D 轮融资，估值为 30 亿美元。许多人工智能公司都力求保持低员工规模，而 Harvey 却在迅速扩张员工队伍。据《财富》报道，这家成立三年的初创公司已拥有 340 名员工，并计划利用新融资将员工人数翻一番。部分新员工将帮助 Harvey 打造面向法律以外专业服务的人工智能产品，包括税务会计。该公司的人工智能解决方案可协助律师审查文件和起草合同

**参考链接：** <https://techcrunch.com/2025/06/23/four-months-after-a-3b-valuation-harvey-ai->

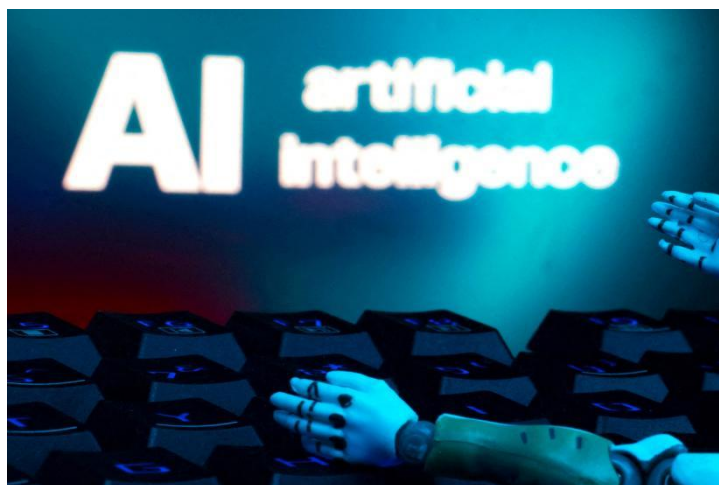
#### 4.ReflectionAI 融资 20 亿美元，成为美国开放前沿 AI 实验室，挑战 DeepSeek



**摘要：**这家初创公司于 2024 年 3 月由 DeepMind Gemini 项目奖励建模负责人 Misha Laskin 和 AlphaGo 联合创始人 Ioannis Antonoglou 创立，AlphaGo 是 2016 年围棋比赛中击败世界冠军的人工智能系统。他们开发这些非常先进的人工智能系统的背景是他们的核心，即合适的人工智能人才可以在成熟的科技巨头之外构建前沿模型。在新一轮融资的同时，ReflectionAI 宣布已从 DeepMind 和 OpenAI 招募了一支顶尖人才团队，并构建了一个先进的 AI 训练堆栈，并承诺将向所有人开放。或许最重要的是，ReflectionAI 表示，它“已经确定了一个符合我们开放智能战略的可扩展商业模式”。ReflectionAI 团队目前约有 60 人，据公司首席执行官 Laskin 介绍，他们主要负责基础设施、数据训练和算法开发方面的 AI 研究人员和工程师。他告诉 TechCrunch，ReflectionAI 已经获得了一个计算集群，并希望明年发布一个基于“数十万亿个 token”进行训练的前沿语言模型。

**参考链接：** [https://techcrunch.com/2025/10/09/reflection-raises-2b-to-be-americas-open-frontier-ai-lab-challenging-deepseek/?utm\\_source=chatgpt.com](https://techcrunch.com/2025/10/09/reflection-raises-2b-to-be-americas-open-frontier-ai-lab-challenging-deepseek/?utm_source=chatgpt.com)

## 5.数据标签公司 SurgeAI 计划首次融资，筹集高达 10 亿美元



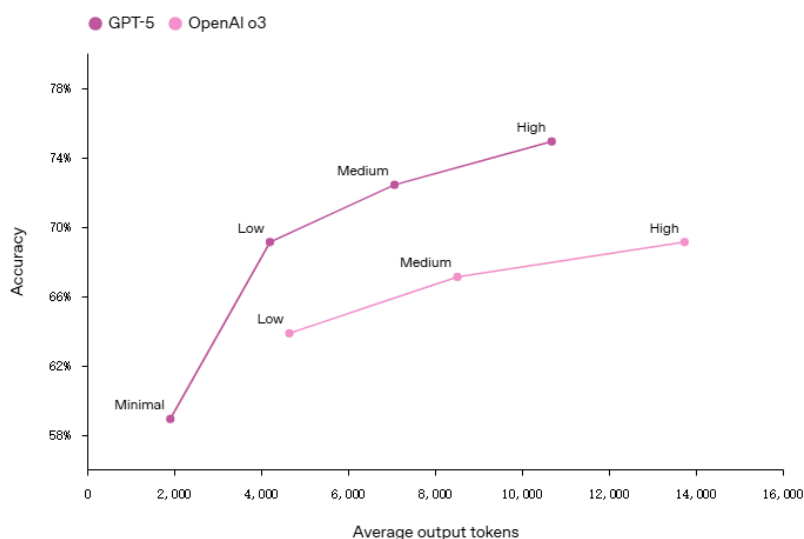
**摘要：**在当下最火热的生成式人工智能大模型中，数据标注公司为其提供了大量“干净、精准”的训练数据。人工标记的数据对模型性能影响深远，是 AI 产业中最基础、最不可或缺的角色。正如英伟达掌控了算力命脉，数据标注公司则卡住了 AI 的数据入口，因此数据标注也被业内戏称为“赛博富士康”，AI 产业的“卖铲人”。比起竞争对手 ScaleAI，SurgeAI 显得十分低调，公司主页没有炫酷的互动，只有一堆纯文字介绍，最鲜明的一句是“数据质量决定抱负上限”。urges 声称其方法不同于旧式的数据标注模式，旧式的数据标注模式往往雇佣全球欠发达国家的人员，以每小时几美分的报酬让他们坐在电脑前区分猫和狗。而 Surge 则创造了一种人机协同的变通方案：由 AI 生成数据并自行标注，但由人类评估其表现。就是这样一家数据标注公司，成立 5 年，员工 250 名（包括全职、兼职和顾问），0 融资，就实现了 2024 年营收 12 亿美元，超过了 ScaleAI 公司同期 8.7 亿美元的营收，还拿下谷歌、Meta、微软以及 AI 实验室 Anthropic 和 Mistral 等大客户。仅 Meta 一家，其生成式 AI 部门去年在 SurgeAI 的全流程标注服务上的支出就超过了 1.5 亿美元。按收入计算，ScaleAI 已经成为该行业最大的企业。

**参考链接：** <https://info.51.ca/articles/1473986>

---

## 四、AI 技术革新

### 1. OpenAI 正式发布 GPT-5（统一多模态、高级推理与 agent 能力）



**摘要：**OpenAI 于 2025 年 8 月 7 日正式发布了 GPT-5。“推理”功能是 GPT-5 的一项核心更新，旨在提升处理复杂问题的能力。系统可以自动识别需要深入思考的查询，并以一系列步骤化的消息展示其“思考过程”，这种机制在数学和编程任务上能有效提升答案质量。但 OpenAI 和外部专家都指出，这并非真正揭示了 AI 的内部工作原理，更像是一种为难题分配更多计算时间的策略。值得注意的是，该功能由系统自动触发，用户无法主动选择，这意味着平台在答案生成路径和资源调配上拥有更大的控制权，用户在透明度和可控性上有所让步。

**参考链接：** <https://chatgpt.com/c/68edf272-dc54-8323-9576-45921f6bdae6>

## 2. FigureAI（类似机器人公司）展示/报道其新一代类人/家庭用机器人进展



**摘要：**当今人工智能的核心事实是不可预测。神经网络如此强大的原因——它们不是从指令中学习，而是从海量数据中学习——也正是这一点让它们如此难以控制。例如，在聊天机器人中，这导致“越狱”模型能够编写恐怖分子手册。但总的来说，文本模型可用于实施伤害的手段有限。机器人则不然。家用机器人甚至能接触到你的菜刀。聊天机器人的幻觉令人讨厌；机器人的幻觉则可能致命。Adcock 声称自己非常重视安全。他正在自己家里测试一台 Figure03，家里还有年幼的孩子。（他称，这台机器需要“严格照看”。）他还补充说，Figure 内部实现了艾萨克·阿西莫夫著名的机器人三定律。

**参考链接：** [https://time.com/7324233/figure-03-robot-humanoid-reveal/?utm\\_source=chatgpt.com](https://time.com/7324233/figure-03-robot-humanoid-reveal/?utm_source=chatgpt.com)

### 3. 谷歌 DeepMind 声称在人工智能问题解决方面取得“历史性”突破

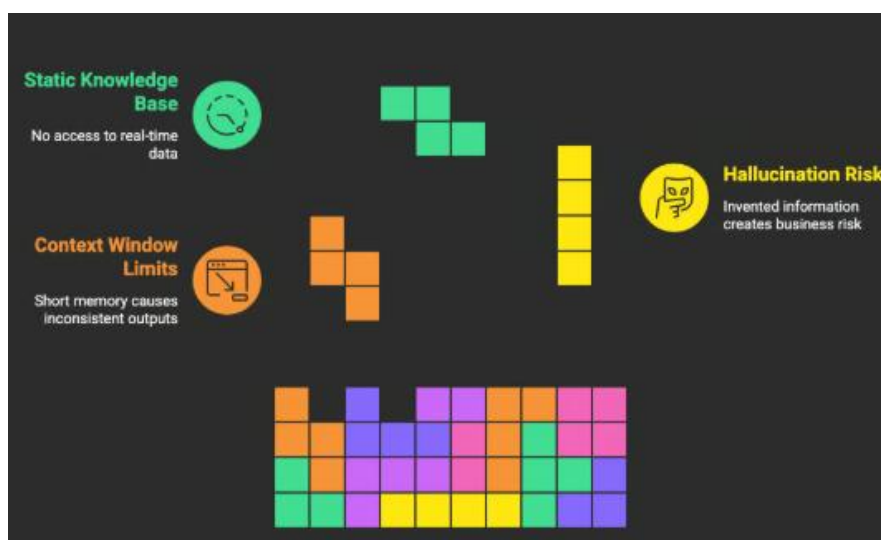


**摘要：**谷歌 DeepMind 声称，它取得了人工智能的“历史性”突破，类似于 1997 年深蓝计算机在国际象棋中击败加里卡斯帕罗夫，以及 2016 年人工智能击败人类围棋冠军。该公司的 Gemini2.5 人工智能模型解决了困扰人类计算机程序员的复杂现实问题，成为本月早些时候在阿塞拜疆举行的国际编程大赛中第一个获得金牌的人工智能模型。

这家科技公司称，这场演示是“抽象问题解决能力的一次重大飞跃”。演示者在不到半小时的时间内，就计算出如何权衡无数种可能性，将液体通过管道网络输送到一组相互连接的储液器。目标是尽快完成液体的分配。包括来自俄罗斯、中国和日本大学的顶尖选手在内的所有人类队伍都没有答对。尽管该人工智能在设定的 12 项任务中失败了两项，但其整体表现在全球 139 位顶尖大学水平的计算机程序员中排名第二。谷歌表示，这是“迈向通用人工智能（AGI）的历史性时刻”。通用人工智能被广泛认为在各种任务上都拥有与人类水平相当的智能。

**参考链接：** [https://www.theguardian.com/technology/2025/sep/17/google-deepmind-claims-historic-ai-breakthrough-in-problem-solving?utm\\_source=chatgpt.com](https://www.theguardian.com/technology/2025/sep/17/google-deepmind-claims-historic-ai-breakthrough-in-problem-solving?utm_source=chatgpt.com)

## 4. 解决 LLM 限制：2025 年人工智能解决方案完整指南



**摘要：**大型语言模型具有革命性，但并非万能。为了有效地部署它们，你必须清楚地了解其固有的局限性。承认这些限制是克服这些限制的第一步。第一个主要障碍是上下文窗口。LLM 的记忆很短，它一次只能处理有限量的信息。一旦在冗长的文档或对话中超出此限制，模型就会忘记之前的内容，导致输出不一致或不完整。第二个问题是幻觉问题。由于 LLM 是概率词汇预测器，而非事实核查器，它们可能会生成看似令人信服但实际上完全错误的信息。在未经验证的情况下依赖其输出将带来巨大的商业风险。第三，他们的知识是静态的。LLM 的知识被冻结在时间中，只感知到它所训练的数据。它无法获取实时信息、突发新闻以及公司最新的内部数据。

**参考链接：** <https://www.firstaimovers.com/p/llm-limits-solved-ai-workarounds-guide-2025>

## 5. OpenAI 与博通达成最新芯片交易，为其打造首款 AI 处理器



**摘要：**OpenAI 宣布与博通(AVGO.US)达成战略合作，共同开发其首批内部人工智能处理器，这是 ChatGPT 制造商在应对服务需求激增背景下，保障计算能力供给的最新芯片合作举措。根据协议，OpenAI 将主导芯片设计工作，博通则负责从 2026 年下半年起开发并部署这些定制芯片，新的定制芯片将于 2029 年底前完成全部部署，这一计划建立在双方现有的联合开发与供应协议之上。据了解，双方将推出算力规模达 10 GW 的定制芯片集群，其功耗相当于 800 万户美国家庭同时用电，或约五座胡佛大坝满发功率。此次合作是科技行业持续加码 AI 芯片投资的最新例证，反映出行业对算力需求的指数级增长——企业正竞相构建达到或超越人类智能水平的 AI 系统。

**参考链接：** [https://www.reuters.com/business/openai-taps-broadcom-build-its-first-ai-processor-latest-chip-deal-2025-10-13/?utm\\_source=chatgpt.com](https://www.reuters.com/business/openai-taps-broadcom-build-its-first-ai-processor-latest-chip-deal-2025-10-13/?utm_source=chatgpt.com)