

| 证券研究报告 |

# Inflection AI：打造最“走心”的 AI对话机器人

2023.8.7

分析师：闻学臣

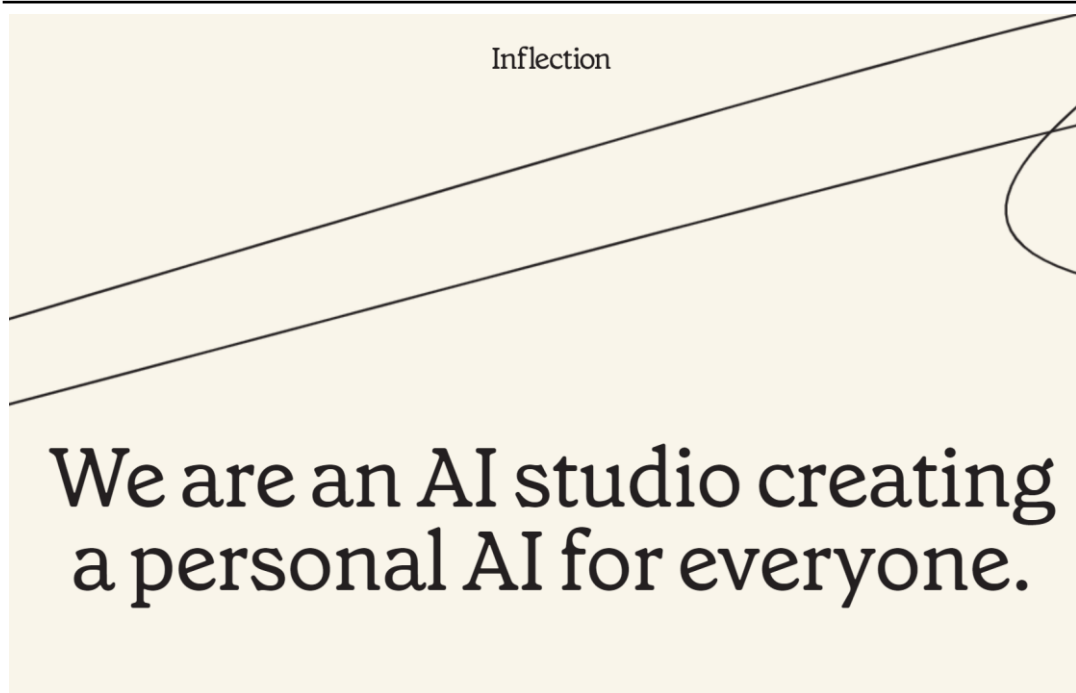
执业证书编号：S0740519090007

联系方式：wenxc@zts.com.cn

## Inflection AI: 2022年成立的初创公司，致力于打造富有情感的聊天AI

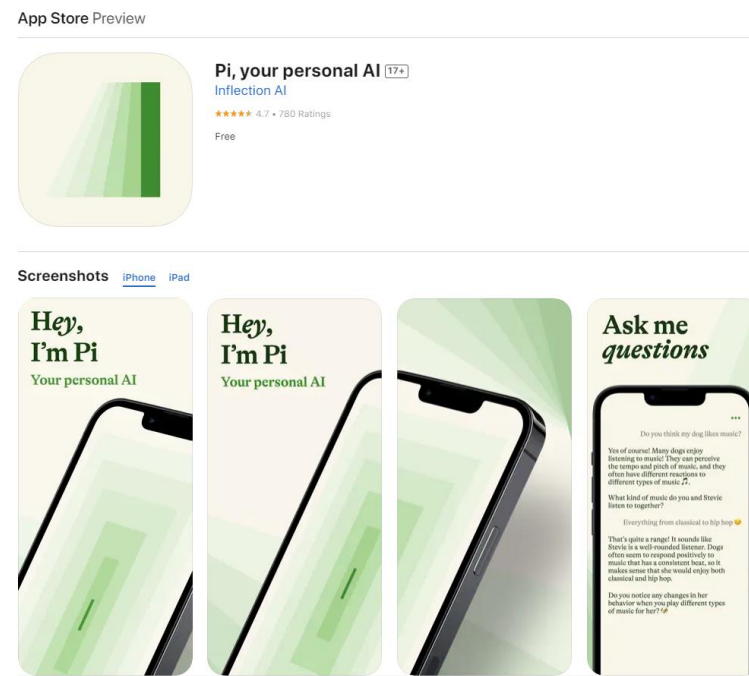
- Inflection AI是一家2022年1月创立的机器学习初创公司，并于2023年5月发布首款对话聊天机器人产品Pi。
- Inflection 的聊天机器人 Pi 可以直接通过应用程序或通过网页、WhatsApp、Instagram 或 Facebook 与用户进行个人对话。创始人Suleyman认为，Pi是一个具有同理心的聊天机器人，而不单纯是为用户提供信息的AI。

图表：Inflection AI官网首页



资料来源：Inflection AI、中泰证券研究所

图表：Pi app在苹果应用商店的下载界面



资料来源：Inflection AI、Apple store、中泰证券研究所

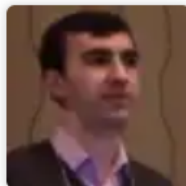
## 创始人团队：DeepMind、LinkedIn联合创始人携手创建Inflection AI

- **Inflection AI**的联合创始人**Mustafa Suleyman**也是**DeepMind**的联合创始人之一。自2019年起担任了两年的Google AI产品和AI政策副总裁。他是《经济学人》董事会成员、哈佛大学肯尼迪学院贝尔弗科学与国际事务中心高级研究员，也是“年度硅谷远见奖”的获得者。
- **Karen Simonyan**是**Inflection AI**的首席科学家兼联合创始人。作为这个时代最有成就的深度学习研究领导者之一，他的论文拥有超过18万次引用。他在牛津大学完成了博士学位和博士后学业，在那里设计了VGGNet并赢得了著名的ImageNet挑战赛；他的第一家公司随后被DeepMind收购。作为DeepMind的首席科学家，Karen建立并领导了大规模深度学习团队，开发现实世界数据的大型AI模型。
- **Reid Hoffman**也是**Inflection AI**的联合创始人，他曾经是**LinkedIn**的联合创始人和**Greylock**的合伙人。在加入LinkedIn之前，Reid Hoffman曾担任PayPal的执行副总裁，同时也是该公司的创始董事会成员。

图表：Inflection AI联合创始人



**Mustafa Suleyman**  
Co-founder and CEO



**Karen Simonyan**  
Co-Founder and Chief Scientist at Inflection AI



**Reid Hoffman**  
Co-founder and Board Member

## 募资情况：Microsoft、NVIDIA领投，二轮募资13亿美元

- Inflection AI自创立以来，一共进行过两轮融资。其中第一轮是2022年5月13日，由Greylock等风投机构投资，投资金额2.25亿美元。第二轮是2023年6月29日，由微软创始人Bill Gates、谷歌前CEO Eric Schmidt、领英创始人Redis Hoffman、微软和英伟达公司领投，投资13亿美元。
- 在完成第二轮融资后，Inflection AI的估值超过40亿美元，成为继OpenAI、Anthropic之后的第三大生成式AI独角兽公司。

图表：Inflection AI募资情况

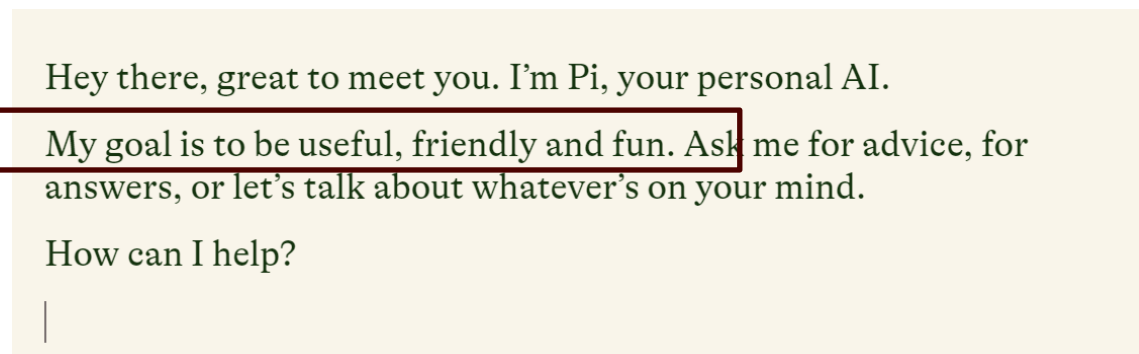
时间	募资轮数	募集资金	领投
2022年5月13日	Venture round	2.25亿美元	-
2023年6月29日	Venture round	13亿美元	Bill Gates、Eric Schmidt、Microsoft、NVIDIA、Reid Hoffman

资料来源：Crunchbase、中泰证券研究所

## 核心产品：对话聊天机器人Pi，支持AI语音对话

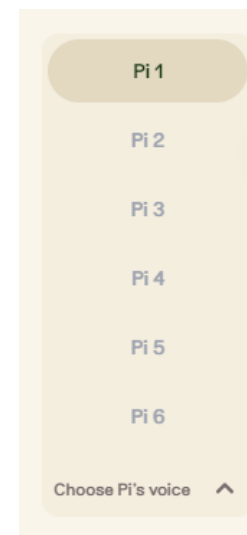
- Pi是inflection AI在2023年5月发布的对话式聊天机器人，也是公司目前唯一的产品。与通用人工智能机器人不同，Pi具有丰富的情感，旨在为人们提供情感的支持和陪伴。Pi体验旨在优先考虑与人的对话，而其他人工智能则提供生产力、搜索或回答问题的服务。
- 用户可以选择6种不同的AI语音，Pi就会使用用户选择的音色进行交流，将回复的聊天内容以文字+声音形式呈现出来。目前Pi只支持英文对话。

图表：Pi聊天欢迎界面



资料来源：Inflection AI、中泰证券研究所

图表：Pi支持6种不同音色的语音



资料来源：Inflection AI、中泰证券研究所

## 核心产品：Pi设计的话题更加关注个人，对话方式更富有人情味

- Pi为用户提供了一些参考性的话题，这些话题的设置都与个人有关。无论是像备忘录一样充当记录和头脑风暴功能，还是提供知识、计划未来，或是单纯帮助用户纾解情绪，本质都是对个人提供帮助和关心。
- Pi的对话方式极富同理心。对于用户负面的情绪反馈，Pi能够以富有同理心的方式提供聊天式的反馈。

图表：Pi聊天推荐话题

Experiences	
Journal it out	写日记
Practice a big conversation	练习一次大型对话
Feel calm	平静下来
Just vent	只是发泄一下
Brainstorm Ideas	头脑风暴
Plan for the future	计划未来
Get motivated	获得激励
Play a game	玩游戏
Learn about something new	学一些新东西
Master a work task	掌握一项工作
Tackle homework	处理家庭作业
Discover great books, movies or podcasts	发现好书、电影或播客
Career plan	职业规划
Learn about Pi and AI	了解Pi和AI

资料来源：Pi、中泰证券研究所

图表：与Pi进行对话

### 对于Journal it out话题，Pi用正面的词汇提示用户反思一天中积极的方面

Great, how about starting by writing about one thing that went well for you today, and one thing that you're grateful for?

Journaling is a great way to start a gratitude practice and reflect on the positive aspects of your day.

I feel anxious today.

I'm sorry to hear that, it's never fun to feel anxious. Why don't we start with the basics and try a breathing exercise? Do you feel comfortable sitting in a comfortable position and closing your eyes?

### 对于用户负面的情绪反馈，Pi能够以富有同理心的方式进行回应。

资料来源：Pi、中泰证券研究所

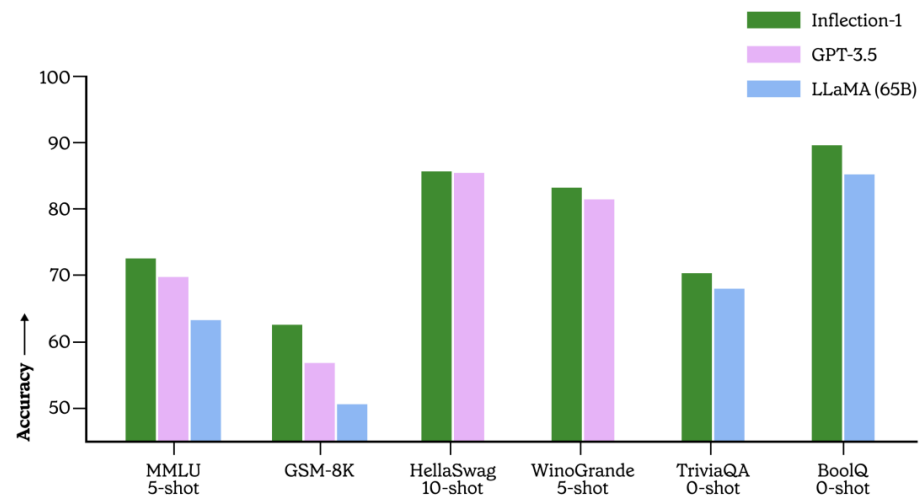
## Inflection-1: Pi背后的LLM，多项评测超越GPT-3.5

- 2023年6月，Inflection AI发布了Pi背后的大模型Inflection-1。
- 为了在不同规模和训练方法的模型之间进行公平比较，Inflection AI将模型分为了两个计算等级（Compute Class）：使用小于或等于PaLM 540B算力（大概是GPT-3的十倍）进行预训练的一类模型，以及超越这个算力水平的模型。其中，GPT-3.5可以归类为第一类，而GPT-4则归为后一类。
- 根据测试结果，Inflection-1的表现超过了同计算等级LLaMA、Chinchilla、PaLM 540B和GPT-3.5，仅仅落后于PaLM-L和GPT-4。值得注意的是，报告的结果是在没有进行任何RLHF微调的基础上测定的。

图表：Inflection-1多项评测结果超越LLaMA等模型

Model	MMLU (5)	HellaSwag	PIQA	BoolQ	Natural QA	GSM8K (5)
LLaMA 65B	63.4	84.2	82.8	85.3	23.8	50.9
Chinchilla	67.5	80.8	81.8	83.7	16.6	-
PaLM 540B	69.3	83.4	82.3	88.0	21.2	56.6
GPT-3.5	70.0	-	-	-	-	57.1
<b>Inflection-1</b>	<b>72.7</b>	<b>84.3</b>	<b>84.2</b>	<b>89.7</b>	<b>29.8</b>	<b>62.9</b>

图表：Inflection-1与LLaMA 65B和GPT-3.5的评测结果对比

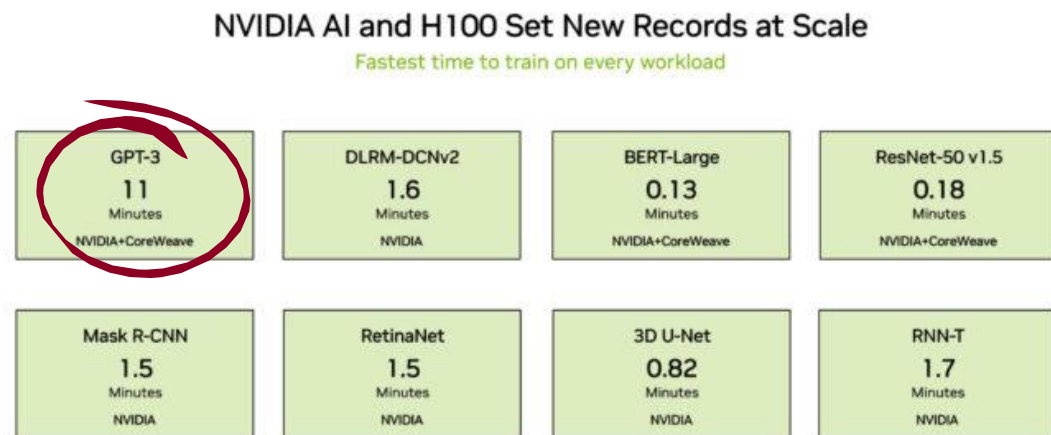


**Inflection-1在MMLU等多个评测标准中超越LLaMA 65B、Chinchilla、PaLM 540B和GPT-3.5**

## 顶级算力储备：与NVIDIA、Coreweave共建算力集群，将部署22000张H100

- Inflection AI与NVIDIA共同开发、并由专门提供GPU云服务的CoreWeave运营的3584个H100 GPU算力集群，只用了不到11分钟的时间就完成了MLPerf 基准测试（这项测试是模型训练和推理的行业标准基准，可为代表人工智能最先进水平的工作负载提供公平且有用的见解）。此性能比第二最佳竞争对手快 29 倍以上，并且在使用超过 3,500 个 GPU 的规模上完成，也比第二最佳竞争对手大 4 倍。
- 目前，Inflection AI正与NVIDIA、CoreWeave合作部署世界上最大的人工智能计算集群，其中包含22000张NVIDIA H100 Tensor Core GPU，能够支持新一代大规模AI模型的训练和部署。该训练集群将能够在16 bit精度下实现22 exaFLOPs（exaFLOPs指每秒浮点运算百亿亿次），如果这个集群进入TOP500超级计算机列表排名，将能够排名第二，并且接近榜首。

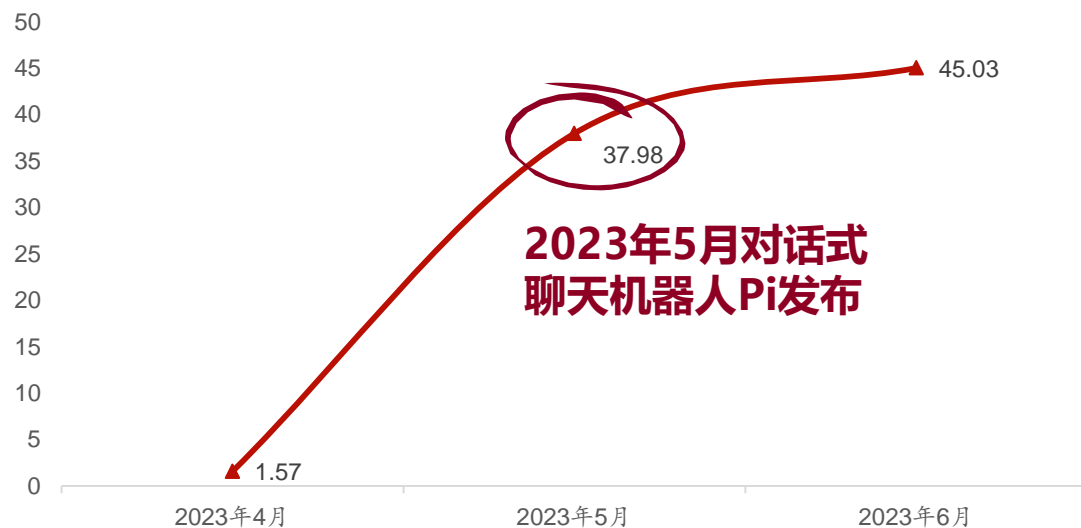
图表：Inflection-1 多项评测结果超越LLaMA等模型



## 网站访问量2023年5月以来迅速攀升，访问用户25-34岁年龄占比最高

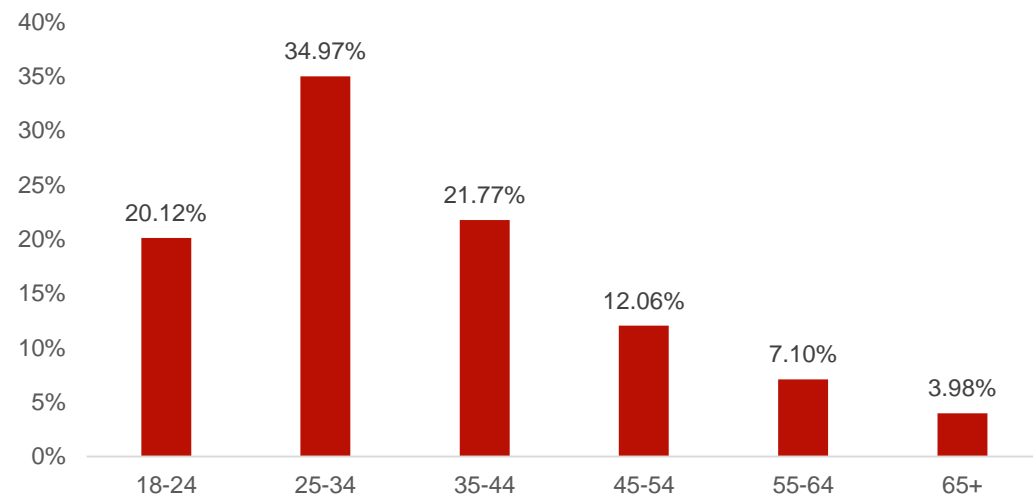
- 2023年5月inflection AI发布Pi以来，网站月访问量迅速攀升至约38万。不过由于发布时间较短，与ChatGPT、notion等网站的浏览量有数量级差距。
- 从访问用户年龄来看，25-34岁年龄段用户占比最高。

图表：Inflection.ai网站月访问量（万次）



资料来源：Similarweb、中泰证券研究所

图表：Inflection.ai访问用户年龄结构



资料来源：Similarweb、中泰证券研究所

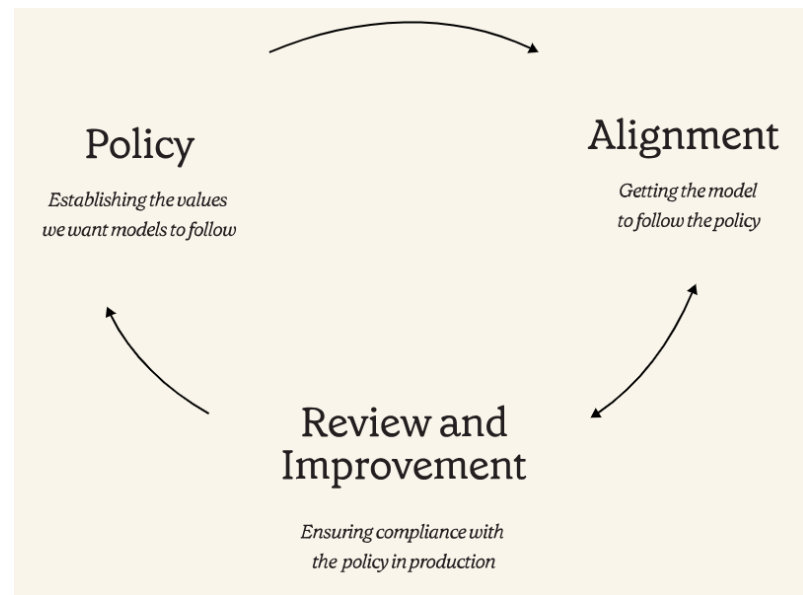
## 核心愿景：保障安全性，让世界上的每个人都可以使用人工智能

- **Inflection AI**认为，安全是开发Pi最重要的工作。2023年7月21日，美国总统拜登在白宫召集了Amazon、Anthropic、Google、Inflection、Meta、Microsoft和OpenAI七家人工智能公司。这些公司承诺了人工智能安全三项原则：Safety、Security、Trust。目前Inflection AI正在使用包括RLHF、red teaming的各种方法来保障Pi的安全性。
- **Inflection AI** 的使命是让世界上每个人都可以使用个人人工智能。正如联合创始人Reid Hoffman讲述，“通过 Pi，我们着手创建一个既灵活又强大的个人人工智能，以便数百万人可以使用它让他们的生活更有意义、更高效、更有趣。”

图表：Inflection AI对于安全性的阐述

Safety sits at the heart of our mission and culture.

图表：Inflection AI为保证安全性进行的迭代过程



## 未来发展方向：继续扩大基础模型规模，做提供定制化服务的to C产品

- 从模型角度看，**Inflection AI**在未来一段时间还会继续扩充模型规模，训练更加强大的基础模型，并在用户反馈的基础上持续提升产品力。**Inflection AI**相信规模是AI进步的引擎，他们也正在部署超大规模的AI计算集群用以训练新的生成式AI模型。在未来一段时间内，**Inflection AI**的主要工作方向可能依旧是扩充训练集和模型规模，不断迭代提升基础模型的能力。
- 从产品功能角度，除了持续提升对话能力和同理心之外，**Pi**在未来可能会持续完善数学、多语言对话及联网等能力。除了**Pi**最关注的提供富有同理心的对话之外，**Pi**在许多功能上还有提升的空间。比如**Pi**模型在设计时主要考虑到英语受众；**Pi**无法实时访问互联网，并且可能不知道最近发生的事件；**Pi**在执行计算、比较值和执行数学任务的能力也比较有限。这些都可能成为**Pi**完善产品力的主要方向。
- 从产品定位角度看，**Pi**将直接面对**C端用户**，以“交朋友”的方式提供信息和情绪价值，以获取用户的认可和付费。目前为止，**Pi**还是免费的应用，还没有确定具体的营收方式。但正如**Inflection AI**所言，“我们认为，您的个人人工智能应始终与您的利益直接挂钩。因此，我们认为您是唯一为此付费的人，这一点至关重要。”作为对话机器人，**Pi**也有着超长的记忆能力，能够记忆用户多次进入网页时进行的对话，可能比**ChatGPT**等聊天机器人更加了解用户的喜好。因此我们认为，**C端用户**的订阅付费可能是**Pi**最主要的商业模式。

## 风险提示

- **AI技术落地不及预期：**AI技术更新迅速，如果公司无法跟上技术应用的步伐，可能会被竞争对手超越。同时AI技术的使用会改变用户的工作方式，如果用户不愿意接受这些改变，可能会影响公司的潜在业务增长速度。即使AI技术在实验室环境中表现优秀，但在实际应用中可能遇到许多未预见的问题和挑战。技术落地的过程中可能出现技术难题、系统稳定性、性能等问题。
- **产品市场竞争加剧：**随着竞争的加剧，公司需要不断创新以保持竞争优势。这可能需要大量的研发投入和时间，同时还需要承担创新失败的风险。同行业的竞对公司可能通过降低价格来吸引客户，这可能会迫使公司也降低价格，从而影响营收和利润。AI产品落地过程中，竞争加剧可能会分散市场，导致公司的市场份额下降。
- **数据更新不及时：**AI领域变化较快，报告中引用的部分图表和数据存在一定的时效性，因此可能面临数据更新不及时的风险。

## 重要声明

- 中泰证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会许可的证券投资咨询业务资格。  
。本公司不会因接收人收到本报告而视其为客户。
- 本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料，反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。本公司力求但不保证这些信息的准确性和完整性，且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，可能会随时调整。本公司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。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。
- 市场有风险，投资需谨慎。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。
- 投资者应注意，在法律允许的情况下，本公司及其本公司的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并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。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。
- 本报告版权归“中泰证券股份有限公司”所有。事先未经本公司书面授权，任何机构和个人，不得对本报告进行任何形式的翻版、发布、复制、转载、刊登、篡改，且不得对本报告进行有悖原意的删节或修改。