



AI 的未来： Perspectives Startups

2025



目录

S

[02 前言](#)

[04 人工智能预测](#)

[08 对创始人的建议](#)

[14 人工智能的下一步：行业领袖的观点](#)

[阿明·瓦德哈特](#)
VP/GMML, 系统与云AI, 谷歌云

[Apoorv Agrawal](#)
帕口口, altimeter口本

[阿文德·贾因](#)
创始人兼首席执行官, Glean

[查马特·帕利哈皮蒂亚](#)
创始人兼首席执行官, 社交资本,
及联合创始人兼首席执行官, 8090

[黄晶](#)
General伙伴, GV

[大卫·弗里德伯格](#)
首席执行官, OhaloGenetics

[多夫·基耶拉](#)
CEO, ContextualAI

[迪伦·福克斯](#)
创始人兼首席执行官, AssemblyAI

[爱德华·利伯特](#)
创始人兼首席执行官, Pinecone

[EladGil](#)
首席执行官, 吉尔资本

[哈里森·奇斯](#)
首席执行官兼联合创始人, LangChain

[詹姆斯·特朗普](#)
总经理, Web3, Google云

[李嘉欣](#)
通用伙伴, a16z

[JerryChen](#)
帕口口, 格雷洛克

[贾莉](#)
联合创始人、总裁和首席人工智能官, LiveXAI

[JillGreenbergChase](#)
投资伙伴, 资本G

[马赫流夫](#)
联合创始人兼首席执行官, Potoroom

[MayadaGonimah](#)
首席技术官和联合创始人, ThreadAI

[拉维拉杰·贾因](#)
帕口口口, Lightspeed

[萨拉姆·特贾](#)
帕口ner, 激口冒口, 以及董事会成员, 艾斯佩克特生物系统公司, 承诺机器人公司, 无畏实验室

[SarahGuo](#)
创始人兼合伙人, 信念

[MikeVernal](#) 伴侣, 信念

[YoavShoham](#)
斯坦福大学计算机科学荣休教授, AI21Labs联合创始人

[72 用谷歌云构建未来](#)



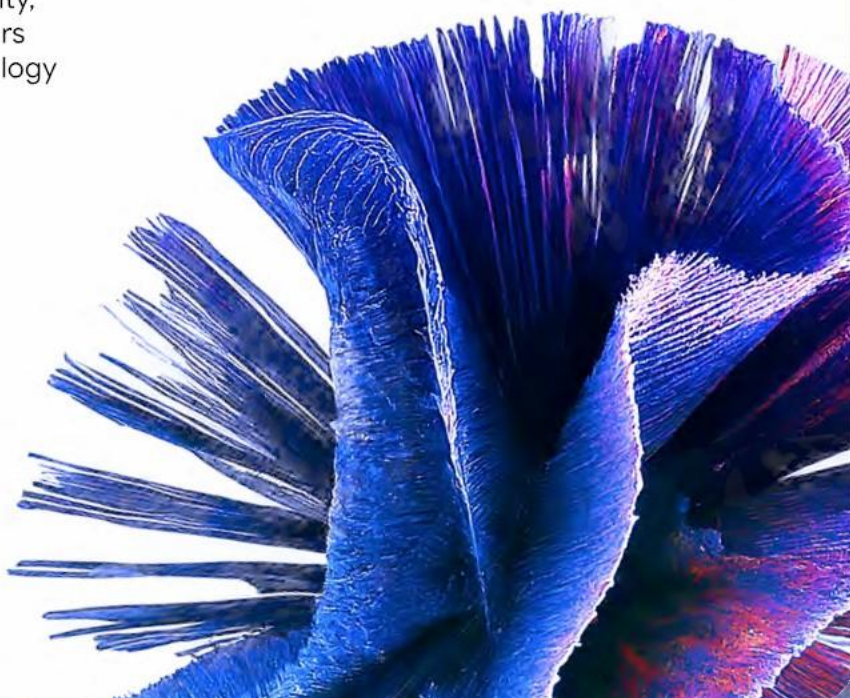
前言

“

人工智能正在改变全球每个组织，并代表着解决复杂问题、推动增长、创造效率以及开辟新的商业机会的前所未有的机会。这对初创公司尤其如此，它们正迅速利用人工智能来应对新的市场机会。谷歌云是人工智能创新的核心，我们为我们持续推动客户可能性的边界的技术领导力感到自豪，包括全球超过60%的获得资金生成人工智能初创公司。我们很兴奋与初创公司、风险投资界和行业领导者合作，确保创始人及其团队能够获得将帮助他们重新定义未来的技术。

ive
ner
ity,
rs
logy

托马斯·库里亚 CEO
，谷歌云





“

谷歌正在构建人工智能技术栈的所有组件，从定制芯片到数据中心到前沿模型。因此，我们新的 Gemini 2.0 模型比以前的版本更具能力、更快、更高效。这些模型是原生的多模态的——它们能够处理文本、图像、音频和视频。它们还可以生成图像和可控制文本转语音音频。凭借高达200万个token的长上下文窗口，Gemini 可以驱动需要深度理解和记忆的高级应用。

此外，思考模型能够展示解决复杂问题的推理能力，这在数学和科学领域尤其有用。Gemini 还可以原生使用 Google 搜索等工具来获取实时信息，而 DeepMind 的 Project Mariner 已证明，使用 Gemini 模型构建的智能体可以通过网络浏览器完成任务。现在，可以使用 Gemini 多模态实时 API 构建对话体验，该 API 接受音频和视频流输入。这些能力的结合能够实现一类新的代理式体验，我们很期待看到 2025 年使用 Gemini 构建的初创公司会带来什么。

David Thacker
首席产品官，谷歌深度学习公司





[01](#)

[02](#)

[03](#)

A 预测方程式

**阿明·瓦德哈特**

VP/GMML，系统与云AI，谷歌云

紧密同步和巨大的计算需求将把基础设施推向前所未有的计算密度和能力水平。

**Apoorv Agrawal**

帕□□，altimeter□本

通过结合语音、视觉和自然语言，多模态AI将减少对计算机和手机等设备的需求，并使与数字世界的互动更加无缝。

**阿文德·贾因**

创始人兼首席执行官，Glean

人工智能将继续作为增强人类能力的工具，而非取代人类。基于人工智能的员工的概念延续了一种有限的视角，阻碍了人工智能和人类智能的真正潜力。

**查马特·帕利哈皮蒂亚**

创始人兼首席执行官，社交资本，及联合创始人兼首席执行官，8090

软件的未来在于用更少做更多。通过人工智能自动化，软件行业将变得更加高效，随着企业得到更多支付更少，标普500的平均利润率将翻倍。

**黄晶**

General伙伴，GV

随着构建生成式AI应用的工具越来越容易获得并因此变得商品化，AI领域的微繁荣和衰退将不可避免。

**大卫·弗里德伯格**

首席执行官，OhaloGenetics

我认为最容易受到人工智能驱动转型的行业是媒体、SaaS和生物（治疗药物和农业）。例如，基因组语言模型将能够预测实现任何期望的植物性状或生物药物所需的精确DNA序列，这将革新农业和人类健康。

**多夫·基耶拉**

CEO，ContextualAI

长上下文和RAG将趋同。模型可能会学习何时使用长上下文以及何时使用RAG来优化准确性和效率。

**迪伦·福克斯**

创始人兼首席执行官，AssemblyAI

广泛应用人工智能的企业时间表将比人们想象的要慢。很多最后一公里的问题需要解决，这些问题直到你深入其中才变得明显。

**爱德华·利伯蒂**

创始人兼首席执行官，Pinecone

AI中最快的投资回报率（ROI）体现在代理上，但最大的机遇在于企业搜索。

**伊莱德·吉尔**

首席执行官，吉尔资本

人工智能被严重低估了。我们正处于由人工智能驱动的一场巨大变革浪潮的早期阶段，传统的机器学习，如自动驾驶汽车，终于迈出了步伐，而生成式人工智能尚处于起步阶段。

**哈里森·奇斯**

首席执行官兼联合创始人，LangChain

要真正发挥代理系统的优势，我们需要它们超越我们所要求的范围——成为“环境代理”，在后台运行，始终在线，监控事件流，并且只在发生有趣的事情时才通知我。

**詹姆斯·特朗普**

总经理，Web3，Google云

你可能会比信任任何人更信任你的 Web3 功能型 AI 代理。你会对它倾诉远比你不会对其他人倾诉的更多事情。它将了解你的一切，并利用这些信息来帮助你。

**李嘉欣**

通用伙伴，a16z

让人工智能代理真正实现主流化将比人们预期的要花费更长的时间。模型推理能力需要提升，但真正需要被代理利用的是基础设施，以及需要解决的深层次系统集成问题。

**陈杰**

帕口口，格雷洛克

人工智能的最终商业模式尚未出现，投资者需要时间来理解它们。

**贾莉**

联合创始人、总裁和首席人工智能官，LiveXAI

我憧憬一个未来，人工智能代理能够无缝融入日常生活，我们的与人工智能的互动更加人性化——拥有不仅能够理解语言，还能感知和响应视觉提示的代理。



吉尔·格林伯格·切斯

投资伙伴，资本G

基础模型在未来18个月内其能力将保持相对稳定，为初创公司构建专门的AI解决方案提供了充足的机会，这些解决方案能够带来明确的投资回报。



马赫流夫

联合创始人兼首席执行官，Potoroom

人工智能将理解和适应人类情感。人工智能将更好地理解什么会引发人类情绪，从而允许故事和内容根据个人的情绪反应进行个性化调整。



玛亚达·贡伊玛

首席技术官和联合创始人，ThreadAI

我们正看到大量模型、解决方案和工具被匆忙推向市场，但它们尚未证明能够真正产生影响——无论是通过投资回报率还是通过根本性地改变开发者生态系统。其中许多库将会消亡或获得越来越少的投资。



拉维拉杰·贾因

帕□□□，Lightspeed

最显著的技术变革将发生在物理世界的底层模型，尤其是在机器人技术领域。



萨拉姆·特贾

帕□□，Radical Ventures，以及董事会成员，Aspect Biosystems，Promise Robotics，Intrepid Labs

景观将从建造更大更好的模型转变为优先实施人工智能以解决现实世界问题并为企业和社会创造实际价值。



郭莎

创始人兼合伙人，信念



Mike Vernal

伴侣，信念

当谈到人工智能时，最有用的数据通常与特定的实际应用场景相关联。在许多情况下，这为新的参与者留下了空间。



Yoav Shoham

斯坦福大学计算机科学名誉教授，AI21 Labs联合创始人

“快速响应祈祷式”人工智能的早期——简单地向语言模型输入一个提示并希望得到好的结果——已经结束了。为了满足企业的迫切需求，我们需要能够协调多个模型和工具、以获得可靠结果的强大“人工智能系统”。



01

02

03

给创始人的建议



随着时间的推移，计算将变得更快、更便宜、更可靠。

假设你有一个绝妙的主意，只需要将计算成本降低10倍甚至100倍，它就能盈利。现在这已经可以实现。

阿明·瓦赫达特 VP/GMML, 系统以及云AI, 谷歌云

使价格与所提供价值相匹配。

不要只使用按座收费模式，而应考虑按使用量或按价值定价。你的定价应反映你的产品为用户提供的价值。

阿普鲁·阿格拉瓦尔 合伙人，
阿提米特资本



将viewai视为提升顶线收入和开拓创新新机遇的方式，而不仅仅是一个提高效率、削减成本的工具。

阿维恩德·贾因 创始兼首席执行官，Glean

关注构建点特征，而非点产品。

优先发展高价值、具体的功能，而不是创建综合软件。这允许在更大的公司可以使用人工智能复制功能的市场中进行竞争。

夏姆特·帕利亚帕蒂亚联合创始人兼
首席执行官，社交资本，以及8090
联合创始人兼首席执行官

最令人印象深刻的创始团队对记录、编码和测量他们的流程非常狂热。

他们希望在测试最佳实践中成为佼佼者，记录下有效的方法，并在整个组织中分享。我发现，在合理范围内重视文档记录（within reason!）与其他方面的运营卓越有关。

CrystalHuang 总合伙人，GV





你需要一些能够创造价值、并在人工智能赋能产品发布后持续保持初始优势的引擎，无论是数据生成以实现持续的性能改进，还是网络效应。仅仅是做一个LLM包装器是不够的。

davidfriedberg ceo,ohalo genetics



Move fast.

上市时间是创业公司最重要的事情，而市场就是现在。

人工智能领域发展极其迅速，所以你必须确保不要落后。

Douwe Kiela, CEO, ContextualAI

与其瞄准通用人工智能，不如专注于解决大量有深度的特定细分问题。

Dylan Fox 创始人兼CEO, AssemblyAI

从箱外解决方案开始。

与其从零开始构建，初创公司可以使用现有工具快速获得一个可生产的产品，然后进行测试和盈利。

Edo Libely Founder and CEO, Pinecone



找到一个人们真正在乎的东西，尽快把它发布出去，测试人们是否在乎，然后从此处迭代。

Elad Gil 首席执行官, Gilt Capital





优先考虑评估。立即开发评估以更好地界定问题。从一开始就明确衡量指标和评估人工智能系统性能的方法至关重要。



哈里森·切斯 CEO兼联合创始人, LanyChain

您如何将AI货币化并销售, 与您如何构建它一样重要。

如果你能击败竞争对手逆转商业模式, 那是有力量的, 因为现有企业很难转换。

JerryChen 伙伴, Greylock

从一开始就展现强大的产品市场契合度、独特的竞争优势和清晰的可盈利路径。

吉尔·格林伯格·查尔斯投资伙伴公司, 资本G

公司无需为了去中心化而进行去中心化。

相反, 将web3的优势——例如更便宜的支付渠道、可审计性和透明度——应用于解决客户使用案例的人工智能代理。

詹姆斯·特罗曼斯 管理总监, Web3, 谷歌云

理解模型能力并将其与用户想要的东西结合在一起, 这就是魔力。

Jennifer Li General Partner, a16z

不要低估数据的力量, 专注于收集多样化、高质量的数据, 同时确保安全和隐私。

佳利联合创始人、总裁和首席人工智能官, LiveXAI





以帮助人们无需使用提示即可实现其目标的方式来设计您的用户体验。

马修·罗伊夫 联合创始人兼CEO, PhotoRoom

构建“无关”的基础设施，让您能够利用持续变化的顶级模型和数据库。

MayadaGonimah CTO和联合创始人, ThreadAI

对AI能力、工具和基础设施的未来发展方向持己见，并构建一个产品来弥补当下该能力上的差距，同时要记住底层引擎将会有显著提升。



投资应用AI研究能力，了解AI技术格局，识别差异化关键机遇，专注于解决现实世界问题，并为智能体AI的崛起做好准备。

拉维拉杰·贾因 合伙人, Light speed



萨拉姆·泰贾·帕纳尔, 激进冒险公司以及董事会成员, 方面生物系统公司, 承诺机器人公司, 无畏实验室

在中间要小心：

应了解 AI 技术栈的不同层级（基础模型、中间件、开发工具和应用），并考虑贵公司属于哪个层级。在中间层建立公司需谨慎，因为如果基础模型迅速发展，它可能会面临压力。

SarahGuo 创始合伙人, 信念

迈克·文塔尔 伙伴, 定罪

聚焦“产品算法匹配”。

理解当前技术的优势与劣势，并围绕人工智能技术今天能做好的方面建立你的创业公司，而不是基于一些理想化的未来版本。

YoavShoham

斯坦福大学计算机科学荣誉教授, AI21Labs联合创始人



无论你身在何处， 在AI应用中，我们 都在这里帮助你。

立即预约您的生成式AI咨询。

→ [立即注册](#)

通过 Google for Startups 云计划，最高可
获 35 万美元云积分。

→ [立即申请](#)

联系我们Startup销售团队。

→ [联系我们](#)



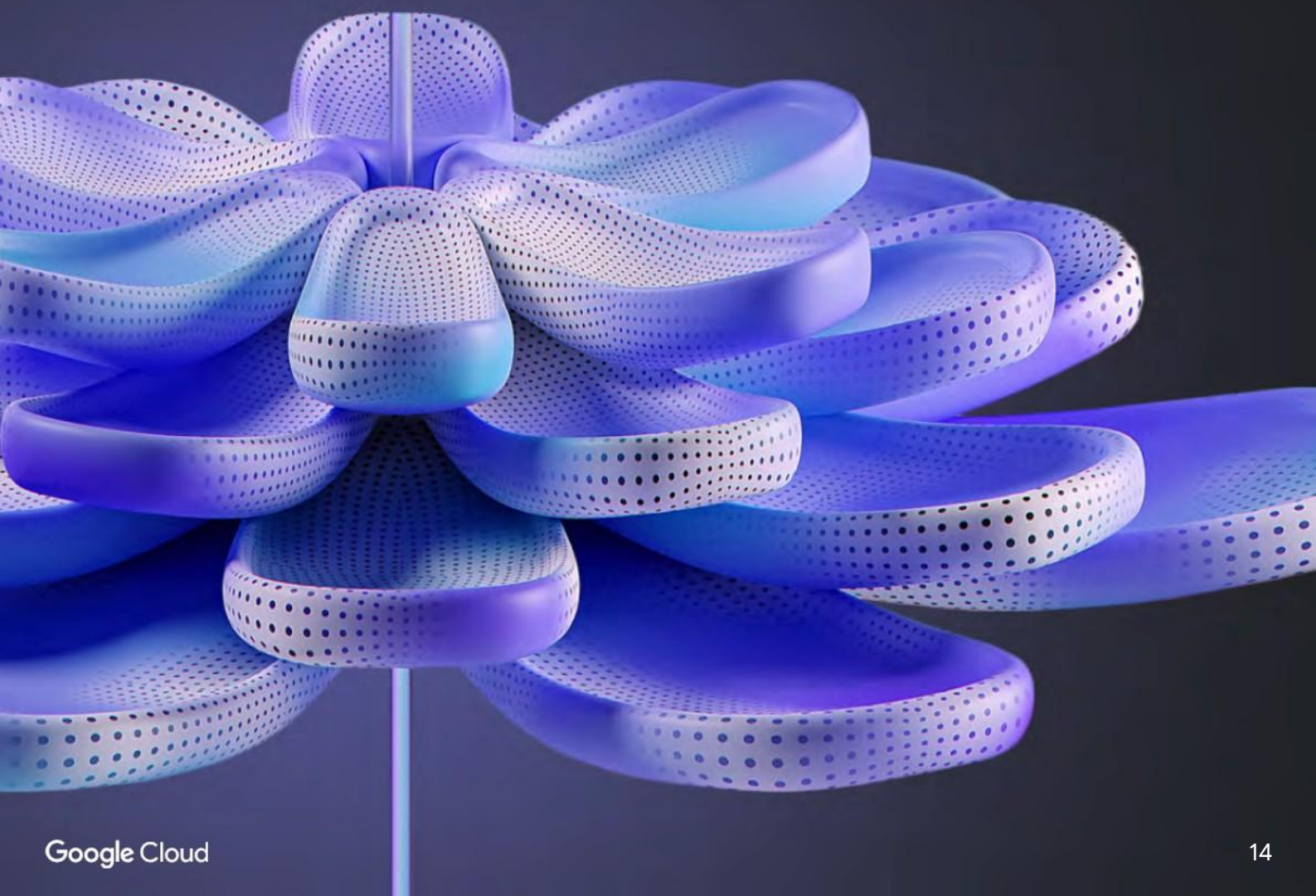
[01](#)

[02](#)

[03](#)

下一步是什么在人工智能

行业领袖的展望





阿明·V·阿达特



副总裁/总经理，机器学习、系统以及云人工智能，谷歌云



阿明是机器学习、系统和云人工智能团队的工程 Fellow 和副总裁。



在加入谷歌近15年前，阿明是圣地亚哥大学的计算机科学与工程学院SAIC教授。他是ACM Fellow和美国国家工程院的成员。

10年后我们的基础设施将面目全非

从软件到硬件，跨计算、网络、存储及更广阔领域的基础设施建设正在进行

今天，我们将现有的架构视为理所当然，因为它代表了系统应该如何构建的“传统智慧”。过去五六年来正在发生的变化意味着我们正处于另一次革命的早期至中期阶段，到本世纪末，新兴的计算平台将变得面目全非。

这一转变的最大驱动力之一是从实时信息获取和传递转向由生成式AI驱动的前瞻性洞察和解读的产生。





生成式AI所需的巨大计算和通信需求，需要在基础设施的设计和构建上下功夫，包括：

01 特定应用的计算

生成式AI计算包含大量的矩阵乘法和相关数值计算。这种专业化意味着，为非常特定的通用计算子集而专门设计的硬件，在以成本或每瓦性能归一化的性能衡量时，可以提供十倍或更多的计算效率。

02 网络方向专业

我们看到从基于标准、商品网络硬件、协议以及软件到专用网络的变化，例如为TPU设计的Inter Chip Interconnect (ICI) 和为GPU设计的NVLink，这些网络专为支持高层计算原语（例如硬件中几乎没有层级的all-reduce）而设计，不需要操作系统管理，本质上是直接进行内存到内存的传输，以支持特定的计算原语。

03 内存墙

高带宽内存（HBM），其中RAM在计算单元相同的封装上进行3D堆叠以大幅降低延迟和提高带宽，正在面临其自身的根本性限制——内存墙。计算能力的增长而未能与为计算单元提供所需数据的相应能力相匹配，无论是来自本地内存还是跨网络，将导致计算和功率的高成本分离——需要计算和内存架构的突破以维持所需的系统平衡点。

04 计算打包

tight同步和巨大的功耗需求将基础设施推向了前所未有的计算密度水平。计算必须在同质元素之间进行，通信必须预先规划和协调，并且容错必须高效。从延迟和功耗/成本的角度来看，将越来越多的计算能力压缩到更小、更功率密度的环境中可以带来显著的好处。

05 液体冷却

持续的强大计算对于机器学习计算而言，需要满足该方程，使得芯片运行性能的改进效率能够轻松证明更高的功率密度和液体冷却。而液体冷却又需要对数据中心建筑、冷却基础设施、机架等进行端到端的重新设计，以提供最高级别的效率，并结合离网发电技术（包括风能、太阳能、水能和电池阵列），以支持在地球各处的机器学习计算。



开箱即用的先进推理能力，例如使用检索增强生成（RAG）和函数调用编排大型语言模型（LLMs），将使初创公司能够在无需大量资本支出或工程资源的情况下访问模型和人工智能基础设施。



人工智能基础设施演进将影响初创公司以不同方式

这种计算能力的巨大增长是由模型可靠性的持续改进和计算成本的降低所加速的，结果：

01

将在模型质量、安全性和延迟方面取得更多进展

如果你的创业项目专注于真实性，而你的目标是提供高质量、低或零幻觉的服务，那么你仍然应该关注这个问题。但如果你的目标是提供其他服务，就利用所有模型现在提供的质量浪潮。

02

计算成本将暴跌

计算成本正在迅速下降。如果我是一个AI创业公司，我会押注计算的将来将变得更快、更便宜、更可靠。假设你有一个绝妙的想法，只要将其计算成本降低10倍就能盈利。对我来说，这是一个轻松的赌注——即使你需要降低100倍。

如我们所见，整个行业模型构建者将继续优化以实现成本节约，利用精简的基础设施并保持极小的利润空间。强大且多样化的加速器（GPU和TPU）的可用性将继续是关键，正如之前强调的高吞吐量网络功能一样。

绝大多数专注于构建创新和差异化应用的企业，将进一步提升至更依赖软件的服务层面。具有先进推理能力（如使用检索增强生成（RAG）和函数调用来协调大型语言模型（LLMs））的开箱即用功能将需求更高，因为它们允许在不进行大量资本支出或工程资源投入的情况下访问模型和基础设施。





阿普尔维·A·格拉沃尔



帕口口，高度表口本



阿普尔瓦领导软件和人工智能创业公司的投资，并在OpenAI、Glean、Parloa和Everest有显著的投资。



一名受过工程师教育的Apoorv，从Palantir的前方部署工程师开始他的职业生涯，他仍然在编程。

计算机自动化和多模态AI是个人计算的未来

接下来人工智能发展的方向将自动化平凡和人性化技术

我们正处于一个对电脑和手机上瘾的时期。我相信我们会将其视为人类历史中的一个奇怪波折。一百年后，人们将不会沉迷于屏幕。人工智能将解放我们。特别是，多模态人工智能——或者说结合语音、视觉和自然语言处理的人工智能——将使与数字世界的互动变得完全无缝。

想象一个你可以通过对话来控制整个数字生活的世界——语音、视觉提示、触觉，以及在所有其他方法都失败时，打字。这就是我们正在构建的未来，它将在未来几年为初创企业解锁巨大的机遇。





聚焦自动化、效率和以人为本的体验

作为一名专注于人工智能的投资人，且相信未来将由多模态人工智能驱动，我已识别出三个指导我决策的核心主题：

01 自动化繁琐的工作

客户服务、后台职能以及任何数据密集型和规则为基础的工作都极易受到AI的颠覆。我对在人类从事的工作不能满足他们、不能给他们带来乐趣、也不是创造性的领域进行AI投资很感兴趣。AI擅长自动化那些令人不满足、数据丰富的任务，从而让人类能够从事更有意义的工作。

与这些主题类似，我也对智能体工作流程感兴趣，其中人工智能代理为我们处理整个任务或流程。一些最聪明的人正在开发工具来让我们摆脱点击按钮和弯腰面对屏幕，这对我来说并不惊讶。

02 增强人类能力

人工智能是让我们变得更聪明和更高效。想想那些能够综合大量信息并传递可操作的见解以解锁切实的生产力提升的工具。我对开发人工智能解决方案的初创企业很感兴趣，这些解决方案能够帮助改进人类能做的事情。



我对在人类从事那些无法满足他们、带不来快乐、也没有创造性的工作领域进行AI投资很感兴趣。AI擅长自动化那些令人不满足且数据丰富的任务，从而让人类有更多有意义的工作可做。

03 技术人性化

我认为语音AI尤其提供了一种更自然、更直观的与技术互动的方式。我专注于语音AI基础设施供应商，因为我相信这个领域正在爆发。能够创造真正以人为本的AI体验的初创企业将在市场上拥有巨大优势。另一个有趣的可探索领域是如何利用AI来更好地理解 and 连接彼此，在提高生产力的同时改善关系。





从ingfocus到用口

随着基础设施和模型的成本大幅下降，机遇正向应用层倾斜。可以这样理解：目前的人工智能（AI）栈像一个倒置的三角形，大部分价值集中在半导体和基础设施上。但随着我们进入人工智能增长的下个阶段，这个三角形将翻转。应用层将成为创新和价值创造发生的地方。

对于初创公司来说，这意味着要专注于构建能够解决现实世界问题的、基于人工智能的产品和服务。不要陷入基础模型复杂的世界中。相反，要利用现有的基础设施，专注于创造能够为用户提供切实价值的创新应用。并且要着眼于持久性，这样你才不只是生存，而是通过下一代模型和技术赢得胜利。

我认为这就是真正的机会所在，并且看到了两个大的趋势正在涌现：

01 水平应用

基于人工智能的工具，可提高所有工作职能的生产力。

02 韦氏应用

专为工程师或客户服务代表等特定角色而设计的Specialized AI产品。

一个关键的争论在于，水平AI是否会有如此优秀以至于取代垂直解决方案的需求。另一个初创企业需要考虑的问题是，是否应该为你的产品采用按座、按使用量或按价值定价？没有简单的答案，但我建议初创企业将其定价与所提供的价值保持一致。

对于我的最后想法，我会留给你们我曾经告诉那些向我咨询如何有效使用人工智能建议的创始人的话：



数据

没有数据策略就没有人工智能策略。了解你拥有哪些数据以及如何利用它们来训练有效的人工智能模型。



积分

考虑如何将您的AI解决方案与其他平台和数据源集成，以创建无缝的用户体验。



工作流

最后，想想人们不想做的事情——比如预订机票。AI能自动化这些任务或使它们更直观吗？





阿维尼德·贾因

创始人兼首席执行官，Glean



阿维נד是Glean的创始人兼首席执行官，Glean是一款AI驱动的办公助手，为人们提供他们所需的答案，以更高效、更快乐地完成工作。



在 Glean 之前，Arvind 共同创立并领导 Rubrik 的研发工作，Rubrik 是一家在数据安全领域上市的行业领导者，他还曾在谷歌担任超过十年的杰出工程师，在那里他领导了谷歌搜索、地图和 YouTube 产品团队。

优先用人工智能推动增长，投资回报率可等待

人工智能是关于构建产品并提供之前不可能实现体验的。

围绕AI代理器的讨论已经变得不切实际。去年某个时候，始于“AI可以回答这个问题”的观点迅速变成了“代理器将为我们工作，让我们可以去海滩”，而我们对如何实现这一点几乎没有证据或思考。

几乎所有我与交谈的创始人或CEO都知道我们正处于人工智能的早期阶段，并且它将在未来从根本上改变他们的业务。我也相信人类将保持控制权，在未来可预见的时期内掌握这项技术。个人的生产力将显著提高，甚至可能达到他们之前十倍的水平。重要的是要理解，人工智能应该被视为增强我们能力的工具，而不是取代它们。





在这个人类与AI协作的世界里，对于试图在两者之间找到正确平衡的创始人和CEO，我有几点建议：

01 停止纠结于ROI

人工智能仍处于早期阶段，此时并非思考其投资硬性回报率的时候。虽然你可能有一些专注于提升底线以实现成本节约的项目，但你应该将这些节约再投资于能够产生显著增长的项目。人工智能的唯一目标并非通过减少客服代理或自动化部分业务流程来实现20%的效率提升。人工智能也关乎通过做些以前从未做过或构建过的事情来增加你的顶线。机会很明确：利用人工智能通过一些效率工作实现成本节约，然后将这些节约再投资于新的创新，为你开辟新的机会。

02 确保团队中的人工智能流利性

作为一项提升，你不仅要确保你的团队处于最新AI工具和技术的最前沿，而且要确保你的员工感到舒适地使用它们。这不仅仅关于采用新技术，这是关于从根本上改变你公司的组织方式和构建方式。你的员工需要被教育如何使用AI，并且将其视为伙伴。你还应该考虑雇佣那些已经舒适使用AI的人，同时确保你所有的员工都配备正确的工具和教育，以成功地将AI融入他们的工作流程中。

03 将AI视为工具，而非产品

任何你构建的产品都需要拥抱人工智能才能更智能、更强大、更现代化。作为一家初创公司，你需要首先将人工智能视为一个工具。如果你开发的产品使用人工智能实现了其90%的功能，你将使你的初创公司完全被取代。从问题（而不是功能）开始，并部署人工智能来更好地解决这个问题。

04 考虑变化率

人工智能技术不断进化。当您制定产品路线图时，要问这些问题是否是您需要自己解决的问题，还是人工智能将在规定时间范围内或可预见未来自行解决。在这些计算出的决策基础上，投资研发以预期核心技术进步，而不是浪费时间在很快就会过时的东西上。这同样适用于您如何构建。设计您的底层架构，使其对模型和工具的进步更加无关紧要，预测即将发生的事情，并在可行的地方实现即插即用，以便在不进行大规模改造或中断的情况下利用更新。



人工智能不只是关于效率。更大的机遇在于提升你的总收入——去做以前永远无法做到的事情，去构建以前永远无法构建的产品。





Chamath Palihapitiya



创始人兼首席执行官，社会资本，
以及联合创始人兼首席执行官，80

90



查玛斯是社交资本这家技术投资公司的创始人兼首席执行官，也是8090这家人工智能公司的联合创始人兼首席执行官。



在2011年创立社会资本之前，Chamath在2006年至2011年期间是Facebook的高级管理人员，他是该公司增长、移动和平台团队的负责人。

为什么萎缩的软件市场能推动更大的利润

软件的未来在于利用人工智能以更少做更多，赋能企业变得更高效和盈利。在帮助公司实现这一目标的过程中，许多人愿意与实际创造这种价值的软件公司共享其优势

我认为人工智能不仅仅是持续进行的“冲击波”：它是一场超级颠覆。

传统的软件行业，凭借其复杂的产品、高昂的成本、碎片化的功能以及许可模式，正在被彻底重塑。人工智能有能力简化这个行业，并将这个万亿美元规模的“软件工业复合体”压缩至其规模的几分之一。我不认为这是个坏事。这是管理得当的企业变得更有利可图，并可能将其中一部分收益与使这种转型成为可能的公司分享的机会。

像当今许多管理得当的最佳公司一样，所有公司都应能构建适合自身的定制解决方案——即它们自己的业务操作系统——这些解决方案是专门针对其需求量身定制的，而不是依赖需要昂贵变更管理且难以适应的现成产品。

stomization.





人工智能中的真实价值创造在于构建一个软件工厂，该工厂可以将业务需求作为原材料输入，并生成高质量的生产代码作为输出。当这成为可能时，企业可以完全重塑其组织结构。

托姆，这是既定的：软件的未来在于用更少完成更多。它关乎自动化，最终赋能商业领导者以更简单、更高质量的方式实现其目标。

在快速发展的AI领域，模型和工具不断更新和被替换，初创企业需要优先考虑灵活性。构建适应性强且面向未来的解决方案至关重要，因为今天的尖端技术可能明天就会过时。

这意味着采用模块化方法进行开发，组件可以轻松更换或升级。这也意味着要了解人工智能的最新进展，以避免不断重复工作的痛苦，并抓住新机遇。

对价值创造的变化

我们正走向一个世界，其中奖励函数不是基于点击或参与度，而是基于最大真相。我期望这将从根本上改变我们解决问题和价值创造的方式。

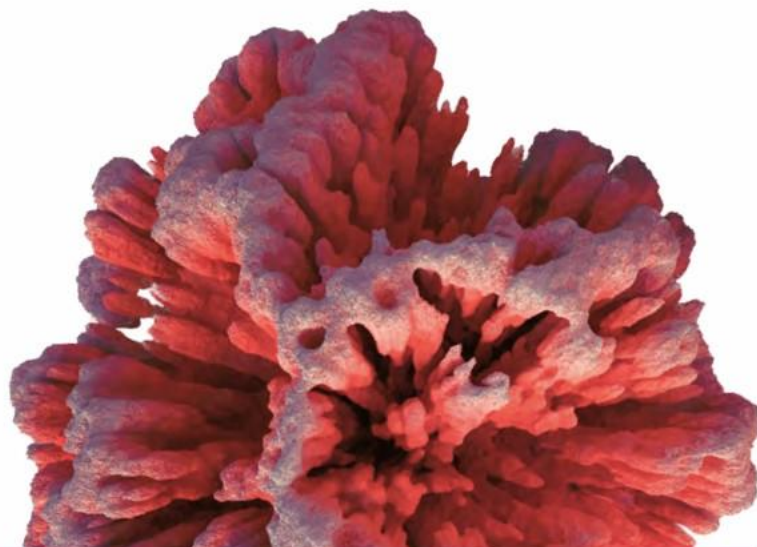
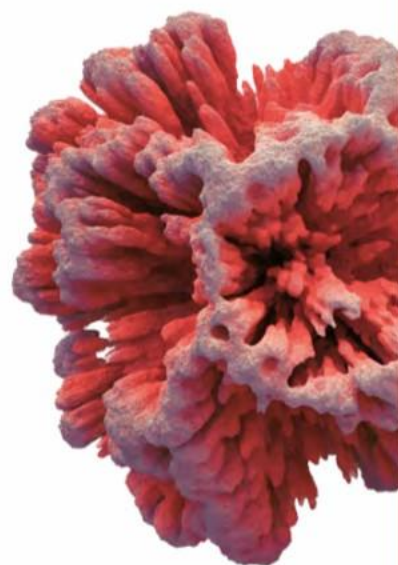
想象一下，一个完全自主的智能体可以处理复杂任务的未来。例如，在不到两个小时的时间里，一些软件可以获取一家公司的产品库存，并生成所有必要的商业基础设施，以在多个国家销售这些产品，并且还要遵守各种复杂的隐私规则。

我们已看到单人公司正与人工智能协同出现。这些公司的运营者使用人工智能来自动化任务、生成内容，甚至管理客户互动。我将其理解为对人工智能民主化力量的证明，使人们能够以前所未有的方式建立和扩展业务。

核心要点是：关注真正增加价值的事情，并砍掉不必要的。AI可以帮助我们识别并消除低效，简化工作流程，并优化流程。



人工智能中的真正价值创造并非在于构建解决特定问题的产品，而在于创造可以将业务需求作为原材料并生成生产代码的自动化软件工厂。





开辟新价值路径

传统创业公司解决特定低效和问题的路径仍然可以非常有利可图。那么，我看人工智能有三类早期采用者原型：

01

已有企业拥有显著的IT预算，且其现有系统的投资回报率较低。

02

那些快速增长以至于超出其招募工程师或构建内部基础设施以保持同步的公司。

03

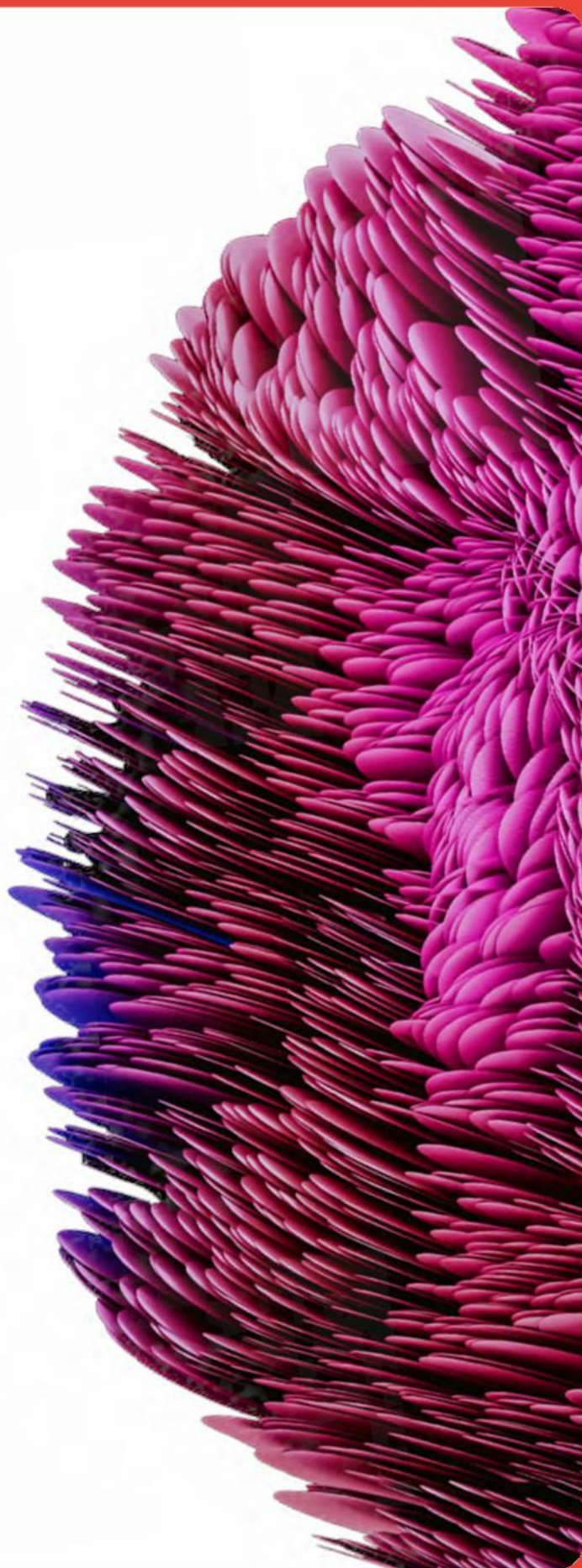
面临重组和削减成本压力的再转型公司，使他们更愿意接受创新解决方案。

这些多元的客户画像展示了AI在各个行业和成长阶段中的巨大潜力。这为创新奠定了坚实的基础，使初创公司能够利用AI来颠覆现有市场并开拓新的疆域。

建设软件工厂，而非产品

对于初创公司而言，这既是一个独特的挑战也是一个机遇。传统的构建独立产品的模式可能不再足够。相反，初创公司需要思维更开阔，专注于能够快速高效部署的灵活解决方案。

人工智能正在改变我们对软件的固有认知。是时候拥抱一种新的思维方式了，这种思维方式优先考虑简洁、自动化，最重要的是，价值创造。软件的未来在于少做更多事，而人工智能是释放这种潜力的关键。





泣石华

将军，GV



水晶是谷歌风险投资的普通合伙人，她专注于人工智能、SaaS和基础设施的投资，特别关注以产品和开发者为主导的采用策略。



此前，Crystal曾是NEA和Notable Capital (□a G GV) 的投资者，并在黑石开始了她的科技并购职业生涯。

2025年，AI驱动的超个性化即将到来

.....但它会足够驱动粘性吗？

虽然AI仍然在飞速创新，但这一领域创业公司的炒作和投资周期已成为我参与的任何技术趋势中最短的。

我从没见过这么多公司在发布时获得疯狂的成功，达到数百万的年收入，然后人们在这么短的时间内就失去兴趣。可能的情况是，一周后一个竞争对手推出产品，它稍微好一点。也可能是某个基础模型公司在其下一次重大发布中推出相同的功能。我认为这种繁荣与萧条的周期将会持续下去，因为构建应用程序的工具非常普及。

建立新的基础模型很困难且成本高昂，所以不是很多团队都能做到，但在应用层，我认为将要出现大量颠覆和再分配。

2025年的格局将与去年截然不同，令人兴奋的是，生成式AI无疑是投资者的兴奋领域，而标准估值框架仍然适用。





这是我会寻找的内容：

粘性作为指标

如果你的产品易于实现，那么卸载它也同样容易。产品需要更具粘性，才能创造持久价值，这意味着它既不可或缺，又能深度整合到用户的工作流程中。同时，虽然企业正在推动人工智能的所有紧迫感，但它通常需要平台和企业之间的相互协作——共同创造自动化或工作流程，并获得难以获得的企业数据，以显著提升性能。

最快的增长来自于那些愿意每月实验新事物的人，而不是企业每年投资2千万美元用于遗留系统。但尽管财务可及的入口能导致更快的消费者采用，但这同样让它难以阻止他们尝试新事物。



如果你的产品易于实现，那么卸载它也同样容易。产品需要更有粘性以创造 lasting value，这意味着它既不可或缺又深度融入用户的 workflow。

超个性化

我们一直听说一个未来，其中每个消费者都会获得人工智能驱动营销活动、辅导或高度个性化的医疗保健服务，这些服务都是即时生成的，但这尚未发生。人工智能模型的训练和推理成本很高，是主要的障碍，但这些成本正在以我认为非常神奇的速度迅速下降。随着更小、更专业于特定领域的模型的涌现，训练费用正在下降，而推理成本则全面暴跌——这使得个性化推理和决策不仅具有经济可行性，而且可能成为人工智能公司长期留住用户的差异化因素。

我还有最后一点要补充：企业客户出奇地了解某些人工智能功能正在商品化。我采访过一些首席信息官和首席技术官，他们今年正在采用一种生成式人工智能解决方案，但已经在计划明年提交一份招标书，以寻找一个更便宜、更强大的解决方案。仅仅筹集资本已经不够了；初创企业需要证明真正的产品投资回报率，产生有意义的收入，建立可防御的护城河，而且，归根结底，要兑现炒作的承诺。





大卫·弗里德伯格



首席执行官，OhaloGenetics



大卫是奥哈罗基因公司的首席执行官，2019年他在担任制作委员会首席执行官期间共同创立了该公司，该委员会是一个风险孵化器，在食品、农业和生命科学领域建立和投资技术企业。



在奥哈洛和TPB之前，大卫是气候公司的创始人兼首席执行官，该公司是世界上领先的数字农业平台，今天被全球2亿多英亩的农民使用。孟山都公司在2013年收购了气候公司，当时他加入了孟山都（现为拜耳）担任高管团队成员。



大卫也是顶级商业和投资播客“*All-In*”的四位联合主持之一。在他职业生涯早期，他在谷歌担任企业发展和产品的高级职务，并从加利福尼亚大学伯克利分校获得天体物理学学士学位。

法规以及人工智能如何革新硬件、媒体、软件和生物学

我想关于监管以及人工智能驱动的未来如何包括到处都有机器人、传统媒体终结、不再有SaaS订阅、更便宜的食物和突破性药物

让我们从有争议的领域开始：人工智能监管。有些人认为人工智能应该受到严格监管以防止其被滥用，而另一些人则认为监管会扼杀创新。我的看法是？世界不会对人工智能进行统一监管，因此应用较少监管的行业或司法管辖区将获得优势并赢得市场。我认为你需要让人工智能自由运营，才能在与试图利用人工智能造成伤害的坏分子竞争并取得成功，依赖现有法律来防范人工智能的不当使用，而不是试图控制人工智能的发展。

一些人工智能公司会突破边界，另一些则会优先考虑安全。每家公司都会定义自己的“负责任人工智能”，而这正是健康的。我们需要具有不同约束条件的多样化人工智能系统，就像我们有多样化的媒体机构一样。让市场，而不是监管机构，来决定它想要什么。

进一步，试图规范模型构造的方式永远无法跟上人工智能创新的步伐。



巨大颠覆潜力：机器人、媒体、SaaS 和生物学

这里有几个我认为具有巨大颠覆潜力的领域：

01 硬件

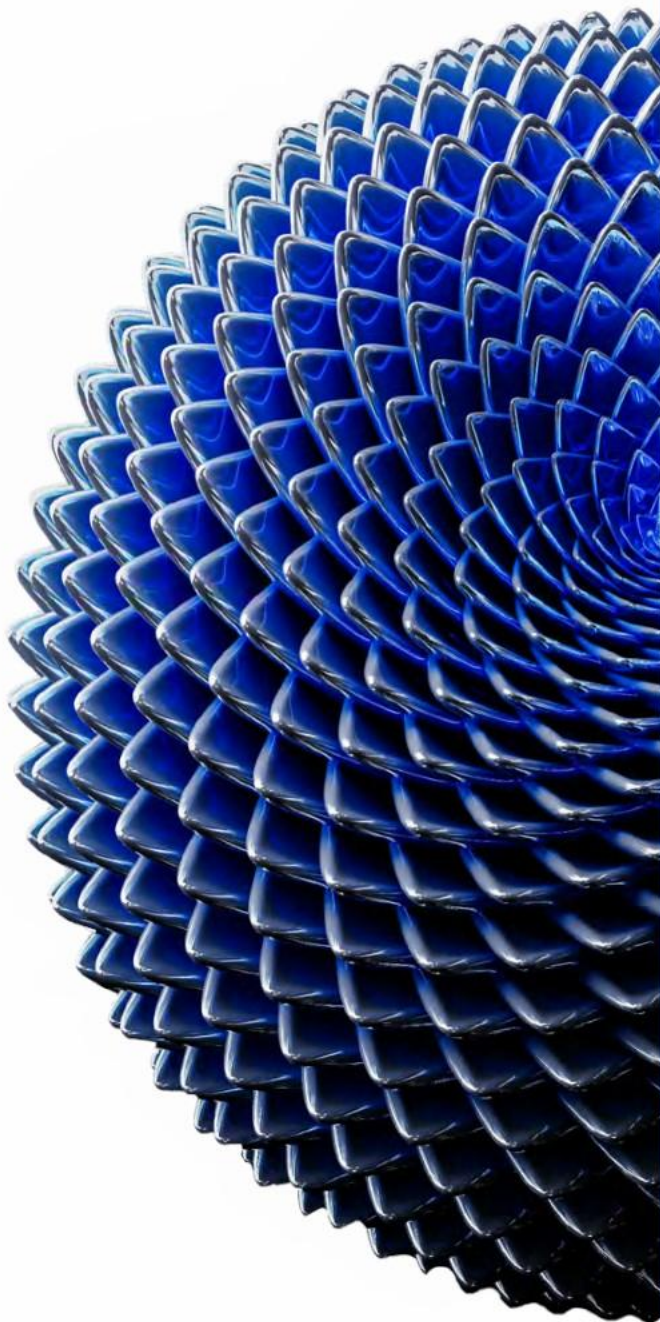
2025将是机器人年。我们正见证着人工智能在自动化领域通过机器视觉和自学习带来的巨大变革。而2024年专注于计算能力和基础模型的构建与创新，2025年将成为应用和利用真实世界和模拟数据来训练和部署具有突破性经济价值创造的硬件系统的年份。飞速的进步让我们更接近能够自动化物理任务、提高效率并执行人类无法完成的物理任务的类人机器。

02 媒体

人工智能将改变我们所熟知的媒体格局。可实时生成内容的个性化电影和电子游戏，将彻底改变“内容制作”的价值主张和经济模式。从任何角色视角体验的基本电影故事，以任何风格呈现，时长不限。拥有无限变化和情节的电子游戏。内容系列中，用户成为故事的主角。人工智能驱动的媒体的动态可能性是无限的。



在未来五年内，我们有可能找到一种方法，将大型模型降维，使小型AI模型能够在网络中协同工作。





03

人工智能工具将赋能任何企业或初创公司，使其能够动态、快速且经济地创建自己的工作流和软件应用程序，取代昂贵、成品的SaaS解决方案，改为按座收费的许可证费用。在我公司最近的一次黑客马拉松中，我们使用人工智能工具构建了应用程序和工作流，取代了现有的SaaS订阅。这只需几个小时。这种内部开发允许个性化解决方案，降低成本，并提供持续更新的能力——这一切都归功于人工智能驱动的软件开发。

04 生物学

人工智能现在可用于设计任何单细胞生物或植物的基因组改变，从而能够创造新型生物药物和新的作物，这些作物可以利用更少的自然资源生产更有营养的食物。得益于低成本的DNA测序，我们正在看到“基因组语言模型”(GLM)的应用，这些模型允许我们定义一个生物学“目标”——一种能与特定癌症蛋白结合的抗体药物或一种能在特定气候下生长的玉米植物——而软件可以渲染出实现该结果的所需精确DNA序列。使用CRISPR基因编辑工具，这些建议可以快速开发并部署。生物学已经成为“软件”，人工智能正在开启一个充满可能性的新时代。

AnLLMwrapper不是商业

这不仅仅关乎技术本身；还关乎你如何使用它。初创企业需要敏捷和适应性强，但也要提供独特的价值。如果你只是个大型语言模型的包装器，那就很难建立一个可持续的业务——你很可能被商品化。企业需要一个价值创造的引擎，它能够持续地提供初始优势，并通过不断的改进来维持。这通常来自于专有数据生成，这些数据被用来持续改进模型性能，或者网络效应，这些效应会锁定数据或客户的访问。

另一个关键竞争优势将来源于基础设施成本。并非总是大模型更好。在未来几年内，大型模型可能会被缩减为在网络上协同工作的较小模型——它们或许能够完成大型模型无法做到的事情。这将降低运行时成本，提高速度和其他性能指标，并开启一个性能的新时代。我在生物学中类比为这一点。

to how genes are expressed and regulated



豆伟基拉

CEO , ContextualAI



上下文AI专注于为您的业务量身定制的、高度准确、可审计的生产级RAG代理。



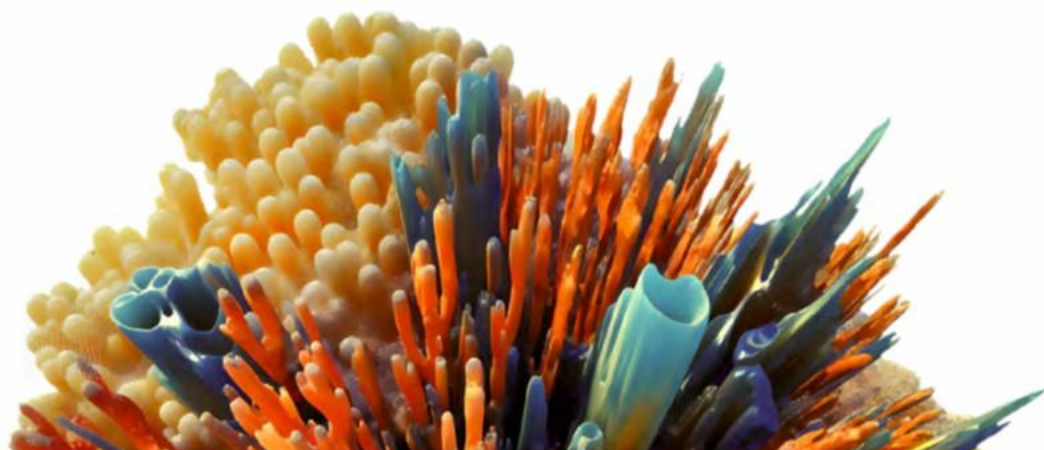
豆诒曾是Hugging Face的研究主管，之前在Facebook AI研究担任研究科学家。他毕业于剑桥大学获得博士学位。

为RAG 2.0 contextualizeAI模型

随着人工智能模型成熟和用例发展，设计有效的方法来利用丰富的上下文以获得复杂问题的正确答案变得至关重要

当我们谈到人工智能时，有一种倾向认为我们比实际情况落后很多。在我们公司，我们了解到人工智能技术的基础设施层有多么不成熟——许多你期望能正常运作的东西，在大规模应用时表现并不理想。

与此同时，我感觉很多公司在使用案例上目标定得太低，专注于基础的助手或副驾驶应用程序。他们没有考虑，如果你稍微提高一点目标，可以用这项技术自动化哪些工作流程和流程。





在检索中积极融入结构化和非结构化数据

基于代理的检索增强生成（RAG）是一种非常有意义的长远趋势，旨在实现更主动和精确的检索。AI代理试图查明回答给定问题所需的信息，从任何存储位置检索这些信息，甚至在给出最终答案之前提出后续问题。

如果你能那样做，那么你就可以潜在地拥有不同的组件，比如文本到SQL模型、代码片段、计算器和结构化数据库查询，以此来推理并得到最复杂问题的正确答案。

目前，有一个数据二元论将存储在数据库中的结构化数据与存储在某个数据桶中的非结构化数据分开。要在其上执行传统的RAG，你必须将其分块并放入向量数据库，然后进行一些额外的操作。但这种二元论正在逐渐消失，因此我认为真正令人兴奋的应用案例是结构化和非结构化数据的交叉点。

随着技术的成熟，我期望多智能体系统将结合人类和AI的最佳表现来求解需要领域经验和文化理解的复杂问题。这正是该领域的未来方向。

多样化数据源并应用层次作为最佳实践

下一代RAG模型更紧密集成，以减少幻觉，提高准确性，并实现企业级性能。这个想法相当简单——将生成器和检索器一起训练，而不是作为独立的模型进行训练，在这些模型中它们并没有真正意识到彼此。

被集成到 RAG 管道中的数据始终具有一种隐含的层次结构，为了最大化准确性，您需要考虑这种层次结构。这在公司整合外部数据源时尤其重要。

我们看到很多客户不想依赖任何外部信息，也许是因为他们有丰富的内部知识库，不想被污染。

然而，在这些公司内部通常存在多个具有某种层级结构的数据源，例如优先考虑研究而不是电子邮件或短信。它们可以将相同的逻辑应用于外部来源。

当存在冲突时，例如网络上的信息与内部研究之间的冲突，情况就变得很有趣。这些冲突可以通过对不同来源分配权重或总结分歧来解决。在我们的流程中，我们有一个主动检索策略，用于确定需要什么样的检索以及如何权衡。



在地化和情境化模型以获得最准确和详细的结果

糟糕或不完整的训练使语言模型容易产生幻觉。为了避免这种情况，我们从可能不太符合我们期望的基座模型开始。然后我们对其进行预训练，使其在来自我们检索管道的检索结果中能够被强烈地约束住。

一个整合得很好的系统能够非常擅长对语言模型进行语境化，语言模型与其语境化内容有着紧密的联系。寻找提升结果准确性和细节的方法将使公司受益。



以牺牲RAG为代价，追求长上下文的品质

关于是否使用 RAG 或长上下文存在一些争论，但如果你知道自己在做什么，你可能会两者都用。成本和质量之间存在权衡。如果将信息嵌入上下文是免费且准确的，那么你可以将所有内容都放入上下文，但这并非如此运作。相反，它会随着上下文长度的增加而增加，因此你需要以巧妙的方式来使用它。

人脑的工作记忆中不会存储太多信息，而是只获取当下需要的内容。类似地，在这些人工智能系统中，你也不希望工作记忆中包含太多上下文，因为那会需要大量的计算。相反，只需检索执行任务所需的 relevant information 来获得正确的上下文。如果你想知道《哈利·波特》系列中的校长名字，你不需要读遍所有七本书来得到答案。

诀窍在于最小化对计算的影响，同时获得准确的结果。上下文的大小以token数量的平方级增长。这是昂贵的。对于长上下文模型，我们通过使注意力机制非常稀疏来解决该问题。基本上，我们在大多数注意力头中完全忽略大多数的token。

将稀疏性推向极限，你将清零大部分信息，你将留下需要关注的大块连续区域。这些最终会提交给语言模型，从而得到最终答案。这种方式下，长上下文和RAG是同一回事。



看起来稍显遥远的使用场景正是你应专注的。如果你假定技术已经成熟，即使它尚未完全成熟，你也能获得先机。



D 依兰·福克斯

创始人兼首席执行官，AssemblyAI



AssemblyAI是一家领先的AI公司，从Accel、Insight、Smith Point和Y Combinator筹集了1.15亿美元，以构建将人类语音转化为有意义的结果和产品体验的AI系统。



迪伦是一名研究工程师，住在纽约布鲁克林。

弥合最后一公里问题

通过解决最后一公里问题和使用精细的基准来评估您的服务水平，以提升您的状态。

我对实时、人工智能驱动的代理将如何发展感到兴奋。我们将看到多模式代理，它们将在其交互中包含视觉和音频。例如，与其打电话给水管工来处理漏水管道的问题，你会打开一个应用程序，展示管道漏水的位置，代理将实时指导你进行修复。

广泛代理采用和人工智能的普及时间线会比人们预想的要长。我这样说有几个原因：

我们必须解决大量的最后一公里问题

01

直到你深入生产阶段才会出现的问题。人工智能的用户体验仍然存在一些笨拙之处需要得到改善——比如中断和响应延迟这些让你意识到你不是在和人类交谈的事情。

我也认为使用和

02

实现人工智能技术将需要改变他们评估它的方式以获得他们期望的结果。好的一面是，这些同样的最后一英里问题和评估人工智能技术的困难为初创企业提供了差异化自己的机会。





解决最后一公里问题

人工智能的最后一公里问题包括其通过专有知识或特定领域知识增强模型的能力、可靠地遵循指令，或检测、移除或完全避免幻觉。我发现这些问题往往是特定于某个行业、用例甚至客户。每个问题都有一个我称之为“模糊阈值”的指标，表示需要解决到什么程度才不再是人工智能采用的一个障碍。

在语音人工智能技术中，特别是实时语音代理方面，我们遇到的最大最后一公里问题之一是代理区分其交互对象的声音与来自电视或其他房间内人的背景声音的能力。另一个问题是它识别罕见单词的能力，如人名、地点、行业或客户特定术语。

如果你是一个AI创业公司，你可以利用这些最后一公里问题来为你所用。将AI应用到你的产品中的很多地方很容易。但要打造差异化的体验，我认为你必须清楚你所试图解决的用例，然后深入挖掘，真正理解人们对这个用例到底关心什么。这将告诉你需要弥合哪些最后一公里问题，才能让你的产品达到最佳状态。



解决你应用中的最后一公里问题，你就可以捕捉那个市场。

用合适的基准评估 AI

人工智能采用的第二大障碍是与评估人工智能技术相关。我看到了公司基于过于笼统的指标对人工智能进行大量投资。为了让公司确信他们部署到生产中的人工智能技术能够实现预期效果，他们需要基于更详细、更相关的指标对其进行评估。

例如，在使用语音转文字ai时，主要的基准指标是字错误率。字错误率通过比较人工转录文本与机器转录文本进行比较，以进行公平比较。然而，我们发现在许多企业或用例中，标点符号在应用中如何显示，或者首字母缩写词的格式如何真的很重要。事实上，对于最终用户和应用开发者来说，识别连续数字、电子邮件地址或罕见单词可能是他们唯一关心的事情。

所以，我建议AI初创公司帮助公司根据真正能反映其特定使用场景成功效果的基准指标来评估你们的AI产品。为了挖掘这些指标，你们需要了解公司将如何使用AI以及它们将用于哪些工作流程。你们的基准指标与客户期望（这些期望通常与最后一公里问题紧密相连）越接近，你们就越容易知道你们的AI产品何时已经跨越了成功那个模糊的阈值。

对我们这些从事人工智能工作的人来说，这是一个令人兴奋的时期，尽管我们需要克服一些障碍才能实现人工智能的广泛采用，但这是会发生的。我认为初创企业有很多机会——你解决了你应用中的最后一公里问题，你就能抓住那个市场。





E d o L i b e r a t i o n y

创始人兼CEO
松果



松果体是构建准确且高性能AI应用程序的领先向量数据库，可大规模生产。



爱德华 previously 曾是 AWS 的研究总监，亚马逊人工智能实验室的负责人，以及雅虎的高级研究总监。

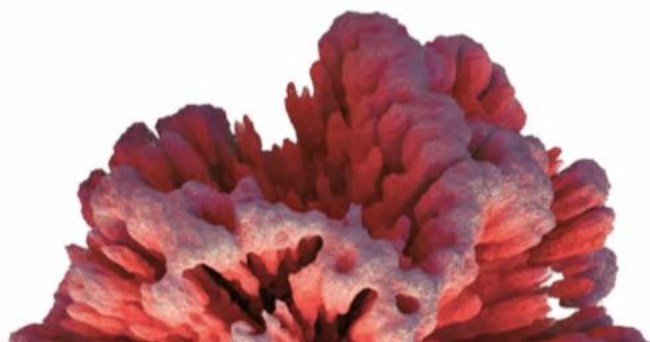
一个庞大的AI机会在于利用企业数据

而业界正在争论AI的下一个商业突破时，我相信一个答案很明显

我们听到很多这个行业的folks谈论RAG（检索增强生成）和agents，但我认为AI最大的opportunities（机遇）存在于搜索中。世界上80%的数据是非结构化的；理解这些数据以及在规模化上高效检索它们的能力，能为拥有大量专有信息的公司提供巨大的商业价值。

许多公司理解词汇或关键词搜索的价值，这对于理解文档和文本非常有益。但理解数据在语义层面的机会要广泛得多——无论是通过图像、视频、PDF文件或表格——以利用集中化的理解感面向众多应用，包括搜索。

向量数据库对于人工智能应用程序中的准确、最新信息至关重要。它们将信息存储为嵌入，这有助于提供可应用于任何工作负载的基于语义的数据理解。





当然，所有数据管理和人工智能相关事务的第一步是确保数据安全——然后是如何考虑扩展它，确保它准确、新鲜且可靠。许多人可能没有意识到，让你的产品生产就绪——这包括GDPR准备、SOC2Type II认证、HIPAA合规性、SSO、RBAC、CMEK、静态和传输中的加密等——实际上可能需要比其他任何事情都更多的工程时间，因此不应被视为理所当然。

准确度、基础性和低延迟在2025年是基准

准确性至关重要。如今，许多大语言模型卡在了60%到80%的准确性范围内。这真正取决于“垃圾进，垃圾出”。你该如何让它们访问最大规模的数据集，以便它们能够实时提供有根据的生成？

向量数据库通过理解和检索最相关的信息，使得AI应用变得知识渊博，进而提供有据可依的准确结果。重要的是，使生成式AI应用变得知识渊博也能减少幻觉。当前和未来商业能够从中提取价值的最重要的用例——包括AI搜索——都由构建于生产环境中为准确性和规模化性能而优化的向量数据库提供最佳支持。

除此之外，延迟是另一个巨大的挑战。许多开发者正与AI中的“香肠制作”过程作斗争。数据库只是更广泛拼图的一部分。我们该如何将其中一些模型与我们的数据库共同托管？我们需要确保体验，无论是面向终端消费者还是员工，都能尽可能达到最佳性能和高准确性。



我实际上认为人工智能中最大的机遇之一在于搜索。

全球80%的数据是非结构化的；理解这些数据并具备在规模化上高效检索它们的能力，可以为拥有大量专有信息的公司带来巨大的商业价值。



EIad Gil

首席执行官，吉尔资本



埃拉德是Gil资本的CEO，是Airbnb、Anduril、Figma、Coinbase、Stripe、Square等众多关键技术公司的早期投资者和顾问。

人工智能被严重低估

尽管如此关注，我们仍处于人工智能改变世界影响的最初阶段。对于初创企业来说，最大的机遇可能就在眼前

人工智能尚未得到充分认识之处在于，其最终产品是认知单元。每个人都更将其视为一种软件工具，但实际上你是在出售人类级别的能力。当前对GPU的投资清楚地表明我们正处在一场大规模变革浪潮的非常早期的阶段。有少数领域我认为具有很大的潜力，你可以在技术堆栈的不同层级和不同类型的公司中看到它们。

如果你看看这个栈，有很多关于基础模型的工作，这些模型专门化不仅限于语言模型，还进入了医疗保健和物理学等领域。然后在应用层面，有大量围绕人工智能通信服务、软件和消费产品的东西，使常见的工作流程更高效和成本效益更高。





作为初创公司，需要提出的关键问题是：你的业务实际上做什么，你的成本在哪里，以及人工智能将如何影响你的竞争格局？

对于物理产业（如航运业），人工智能可能对其核心业务影响有限，但对相关文书工作将产生巨大优化。拥有稳健核心产品或广泛复杂产品套件的企业软件公司可能不会因生成式人工智能（尽管传统机器学习可能有所帮助）而使其核心功能受到干扰，但仍然有潜力优化其内部流程、工作流和客户支持。

人工智能领域存在一个能力阶梯，其中每个模型的迭代都会带来新的可能性，而这些可能性会开辟新的应用场景和市场。你的成本结构中受技术影响的部分越多，你获得人工智能影响或转型的机会就越大。

这里有几个我感到兴奋的领域：

01 医疗保健民主化

我认为人工智能中一个更有趣但未被充分利用的领域是经过预训练或微调的模型。想象一个拥有著名专家的丰富经验的程序，它可以解读图像、你的家庭历史以及之前的医疗问题，并利用所有这些信息来按需为你诊断和创建护理计划。虽然像谷歌的Med-PaLM和Med-PaLM 2这样早期的尝试已经探索了这个领域，但令人惊讶的是，至今还没有人完全产品化了一个如此规模的解决方案。更不用说人工智能通过蛋白质折叠、自动化临床前工作、合规性检查以及临床数据来大幅降低药物开发成本，从而实现更佳、更快的成果了。

02 个性化教育

人工智能可以为学生提供前所未有的个性化关注，指导他们解决特定问题或弥补学习差距，并让他们能够立即获得一流的资源。纵观历史，拥有易于获取的公共图书馆或公共资源作为知识库，为社会带来了巨大的社会效益。生成式人工智能再次解锁了这种可能性。

03 机器学习

即将受到巨大影响的旧的机器学习领域正在暗地里进行。自动驾驶汽车是一个技术即将真正步入正轨的绝佳例子，而没有人关注它。

还有最后一点我要补充：在早期技术时代，通常有很多唾手可得的机遇。奇怪的是，创始人不去选择它们，因为他们认为这太容易了。但我认为，在新市场你可以专注于简单的事，这些事情在规模化时能运作得非常好，你之后还可以建立防御性。找到一个人们真正在乎的东西，尽快将其发布出去，测试人们是否在乎，然后从这里开始迭代。



在新兴市场中，你可以专注于简单的事情，这些事情在规模化时效果非常好。你可以稍后建立防御壁垒。

。





哈里森·凯斯



首席执行官兼联合创始人，LangChain



成立于 2023 年初，LangChain 使开发人员能够使用 LangChain 开源框架构建上下文感知推理应用程序，使用 LangSmith 进行 LLM 可观测性、评估和提示工程，以及使用 LangGraph 进行代理编排。



在加入 LangChain 之前，哈里森曾领导 Robust Intelligence 公司的机器学习团队（一家专注于机器学习模型测试和验证的 ML Ops 公司），领导过快钱公司的实体链接团队，并在哈佛大学学习了统计和计算机科学。

人工智能问题需要人类解决方案

代理商掌握着生产力新水平的关键，但他们的成功取决于我们的指导

在其核心，任何AI代理都需要像人类一样理解一个问题，然后在某种形式的代码和提示中复制解决方案。所以当你正在构建一个时，问自己两个问题：你是否在向代理传达你想要做的事情，以及它是否理解你正在传达的内容？

我认为评估非常重要，以确保您在更改应用程序时没有倒退。我还认为，弄清楚如何评估不仅仅是最终结果，还有到达那里的轨迹至关重要。模型、提示和 RAG 策略会变化——您如何表示这些不同的步骤，并且知道您的应用程序在每一个步骤中都在变得更好？我如何评估更复杂的代理？

我该如何使用这些评估来编程地改进应用程序，同时也要记住这只是更大图景的一部分？在这里花费更多时间将迫使我们必须更深入地思考成功实际看起来是怎样的。

这是我们的责任来定义和优化工作流程，设定评估指标，提供反馈，纠正方向，并减轻模型当前状态的潜在风险。



以下是我在2025年及以后认为将带来重大进步的五大领域

01 用户体验解锁一切

用户体验是当前智能体创新的一个令人着迷的领域。要真正利用智能体系统，我希望它们能够超越我可以向它们要求做的事情——成为“环境智能体”，在后台运行，始终在线，监控事件流，并且只在发生有趣的事情或它们需要我的帮助时才通知我。但是，当你在后台运行智能体时，任何人类级别的参与都需要复杂的系统级编程。例如，这意味着智能体需要按计划运行；它需要能够通知人类，并且在人类与之交互之前暂停（这意味着你需要无限期地维护智能体在状态中的状态）。这可能涉及代码文件更改、内部推理、浏览器状态、文档，然后能够保存该状态并在人类回复时立即恢复。

02 人机协同成为优先事项

我对全自动驾驶代理并不非常看好。我认为最好的代理将结合一个显著的人机交互组件，在最有洞察力的地方设置检查，以确保它们不会重复，并且代理会随着时间的推移从你的反馈中学习。这是在构建代理时需要正确处理的最重要的事情，因为A)它为你提供了一个很好的可控用户体验来与这些代理互动，B)它让系统能够从这些互动中学习。这种方法在幕后支持复杂的工程工作，同时在我们最需要时保持我们的参与。

03 特异性 beats 通用性

人工智能模型在复杂的推理任务上存在困难。它们需要非常具体的指令，并且仍然表现出不可预测的行为。直到模型能够更有效地进行推理，通用代理仍然不可靠。狭窄聚焦的代理基本上取代了人类的工作流程，而构建它们的最佳方法是如何思考人类将如何做某事，然后构建代码和提示的组合来复制该过程。

04 上下文学习提供了比微调更精细的调整

我们能看到的每一个今天构建的智能体仍然是定制化的——无论是使用自定义流程、护栏、逻辑，还是 React 风格的循环。我认为模型创新肯定会更多，但我认为有趣的事情可能会继续存在于系统层面。我认为智能体在2025年仍然会是热门话题。之后，我猜测这与记忆有关。例如，如果我与人类进行一系列交互，我如何记住这些交互并从中学习？我对使用大型语言模型来反思这些交互并更新用户自己的指令或配置文件持非常乐观的态度。随着模型变得更好，这可能会变得更加可能，但目前这真的很昂贵并且需要花费大量时间。

05 底层框架步入正轨

我认为目前存在两种不同的代理框架风格。有一些高级框架内置了任务的概念，你无法真正控制信息传递的精确流程。然后有一些低级代理框架，它们通过提示等方式让你完全控制代理的精确状态以及在任意时间点输入模型的内容。我预测到2025年，我们会看到更多低级框架，因为我们认识到人们需要控制系统内部，才能让它们可靠到足以投入生产。



我对全自主代理持谨慎态度。我相信最好的代理将包含显著的人机交互组件，并在最关键的地方进行检查。



詹姆斯·T·罗曼斯



总经理, Web3, Google Cloud



詹姆斯是谷歌云Web3业务的首席运营官。

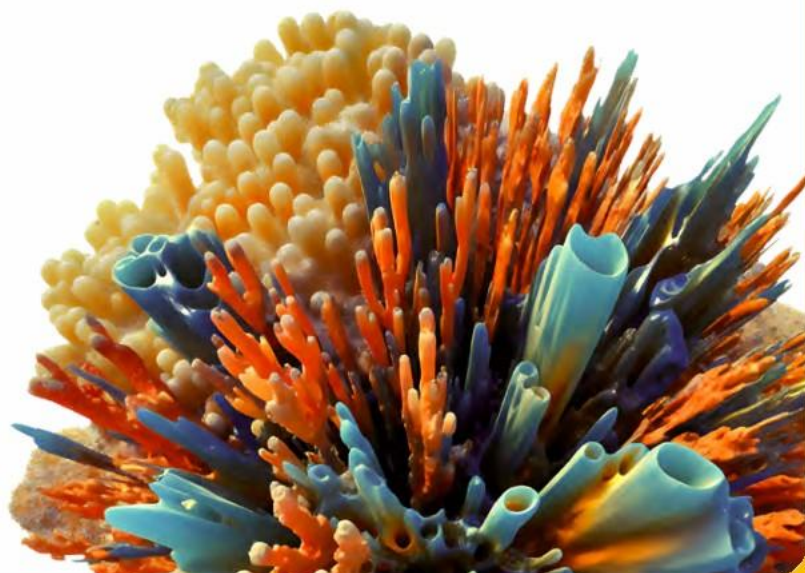


詹姆斯曾担任不同行业的技术专家，并拥有牛津大学视觉计算神经科学的哲学博士学位。

AI和Web3：完美匹配吗？

结合人工智能和Web3可以为消费者提供独特的价值，同时建立对人工智能模型的信心，并加速发展创造力

使用区块链支付渠道的AI代理将成为常态，尽管前方面临严峻挑战。开发者将专注于AI代理作为个性化助理，针对个性化投资、保险和抵押管理等特定任务进行定制，利用Web3支付渠道和智能合约。其他有前景的领域包括开发者赋能、组合性和模型来源。





Web3去中心化：加速AI代理应用场景

Web3更令人信服的应用在于赋能AI代理进行有效商业活动。我们将在这个领域看到广泛有用的创新。AI代理使用区块链支付轨道是非常当前的实现能力；然而，利用传统支付轨道的代理实例仍然是初生的。

例如，投资者会与代理人交谈以确定个人投资策略，代理人会推演并确认风险特征、预期敞口和波动水平等事项，其成本显著低于通常与投资组合经理相关的1.5%AUM。然后代理人将与去中心化金融（DeFi）合约交互，在Web3轨道上使用稳定币执行交易。在过去24个月内经历了爆炸性非周期性增长的这些数字资产，为微支付和其他交易提供了一个稳定的媒介。甚至可能会看到这里的代币化存款或商业银行货币的数字化。类似地，AI代理可以评估和选择保险政策并设定付款计划。

然而，代理需要访问准确、经过验证的用户信息。虽然可验证的凭证正在获得关注，但用户友好、无缝的实施仍然难以实现——这是一个重要的机会，为初创公司。

随着技术进步，AI代理的自主性将增加。代理将访问参数数量足够小的LLM，以便在个人设备上运行，并结合用户在安全隔离区内的验证凭证。代理可以在本地处理这些数据，在采用零知识证明来保护隐私的同时获取用户洞察。这项技术使得能够共享关于信息的事实——例如，证明某人已超过18岁，而无需透露他们的完整出生日期——而不是信息本身。这种AI与去中心化身份（Web3的一部分）的组合既是可行的，也非常有用。

未来的应用包括AI代理理解并解释合同和条款/条件，征集标书，并进入反向拍卖，提供选项以节省时间、金钱和行政负担。用户审查代理生成的最佳选项和利弊列表。创建这个三方市场存在重大挑战，但潜在回报是巨大的。



代表个人行事，代理人需要获得关于该个人的合格、准确信息。这里仍存在差距，空间正在缩小。



人工智能：增强Web3开发

大语言模型代理辅助开发者编写安全的智能合约，缓解重入攻击等风险。这在处理托管资金的DeFi合约中尤其有价值。

更广泛地说，人工智能可以检测错误，自动化编码任务，并推广最佳实践。广泛的采用人机会导致更高质量的代码，增加开发人员生产力，最终实现更多创新。

区块链：验证人工智能模型训练和测试数据

虽然区块链技术已被提出用于验证模型训练数据的来源，但其大规模应用还有待观察。建立一套强大的系统以跟踪和验证训练数据的必要性是明确的，但哪些特定的用户群体将要求这种功能并推动产品市场匹配还有待确立。这仍然是Web3领域探索和一个实验的领域。此外，虽然闭源模型通常缺乏实时数据和推理验证，但在加密领域部署的一些模型已经展示了这种能力。

Web3可组合性：创新的催化剂

Web3的模块化原则——蕴含于技术之中——使开发者能够创建智能合约，使其自包含、开源业务逻辑。它们的透明性使人们能够理解和重用它们。在规模上，这将是一个丰富的生态系统，开发者可以轻松修改和调整现有合约，以用更少的编码量创建新合约。AI代理将进一步降低开发成本和复杂性。

随着我们构建出更高级的多智能体系统，这些智能体共同“思考”以解决一个问题，将离散智能合约功能分配给特定领域是一种自然而然的匹配：一个智能体可以寻找一个提供特定功能的智能合约来评估选项，并选择最佳选项。在大规模上这样做以解决复杂的多智能体问题是一种强匹配。

blockchain
market
ce modules
ows anyone
s fosters
ily combine
applications
reduce

approaches
e mapping
gent
nting
tionality,
step





珍妮弗·李

总合伙人，a16z



杰西卡是安德森·霍洛维茨的普通合伙人，专注于数据系统、开发者工具和人工智能的企业和基础设施投资，同时担任多个董事会的职务，包括十一实验室、图解和旋翼。



早些时候，珍妮弗曾在Solvvy（被Zoom收购）和AppDynamics（被思科收购）担任产品领导职位，她拥有卡内基梅隆大学的软件工程硕士学位和纽约理工学院的技术管理硕士学位。

人工智能代理周边的反对意见

投资下一代AI基础设施

要让ai代理真正走向主流，所需时间将比人们预期的更长。模型推理能力需要得到提升。而且，我看到在生产环境中使用的模型规模远比最强大的模型要多，尽管这仍然是所有大型实验室都在追求的前沿——所以我们将看到整个模型格局如何演变。

但我认为问题不仅仅出在模型层面——底层基础设施和需要深入整合的根深蒂固的系统也需要得到解决。我还需要数据管理工具AI以及它们与用户实时交互的能力。

需要有一个状态管理和内存管理系统——它可以在代理本身内部，也可以在模型外部——来真正协调这个 workflow。

人类真正擅长的是最后一公里：深思熟虑并进行基于洞察力的大决策。几乎所有其他事情——唾手可得的果实——据我所说，都适合 startups 来改进。

agents
ation problems will
o thinking about
ents will need,
ll need, to better





开发者工作流程的快速进化

我对改变开发者体验的AI应用特别兴奋。

编程已经成为人工智能中最成熟的空间之一，因为它更容易验证结果和提供快速反馈循环。你可以检查结果是否正确，而且已经有很多工具，从测试到验证，帮助你确定模型是否按照你的预期运行。它可以简单到自动补全，也可以复杂到构建一个完整的应用程序。

有方法可以将AI能力集成起来以提高开发者生产力，帮助开发者构建下一个应用程序或产品——但是开发者的行为习惯会更难改变。所以如果你真的引导开发者一步步采用AI，那么你可以带来10倍的效率提升。

我认为我们已经看到了采用和投资回报率。而且这不仅仅关于编码。人工智能也使开发者更容易消费和学习新的工具和技术。文档正变得更加易于访问和互动，开发者可以快速得到问题的答案，而不会迷失在信息的海洋中。

所以我对那些能够将AI能力无缝集成到现有（如IDE）或新开发者工具中，以显著提高生产力的公司感兴趣。并且那些利用AI生成用户界面和组件的初创公司，让普通用户更容易构建和部署应用程序。

一种渐进式方法比“一切AI”更有回报

初创公司和创始人似乎明白人工智能是未来的方向。但在我看来，几乎有一种内在的恐惧是引进新能力的，在人工智能方面，当涉及到移动太大或太快时，我会犹豫不决。对于创始人来说，我建议真正关注人工智能能力可以在哪些方面推动变革，并尝试逐步进行变更和转变。我认为这种做法比任何一种更大的“全部人工智能”转变更有成效和回报。

创业者能做的最重要的事情就是非常贴近地面，非常贴近新的AI研究以及现有行为是如何缓慢转变的。这有助于将新产品带入形成，将产品投放市场，并找出市场和用户将如何反应。

最终我认为优秀的团队是惊人的产品构建者。投资者倾向于优秀的产品品味和能够将他们的想法转化为直观、流畅产品的优秀工程人才。要理解模型的性能并将其与用户的需求相结合，那就是魔法。因此，我们不仅倾向于高速度、快速构建产品的团队，也倾向于研究和机器学习的能力。



创始人能做的最重要的事情就是非常贴近地面，非常贴近新的研究以及现有行为是如何缓慢转变的。





Jerry C h e n

帕□□, 格雷洛克



杰里是一位经验丰富的技术专家和投资者，专注于企业软件、云计算基础设施、数据产品和人工智能，致力于与雄心勃勃的创始人合作，创造具有变革性的企业。



自2013年以来，Jerry就在Greylock担任合伙人，他在VMware和Accel Partners的领导职位上积累了丰富的经验，支持公司从初创阶段到规模化发展，并为多家创新公司担任董事。

创意AI的创始人们反戈一击针对现有竞争者

AI升级将使它们有机会重置它们构建、销售和盈利应用程序的规则，以及买家反过来如何资助和盈利

您的竞争对手不再是对手，而是对手的业务模式。我确信人工智能的最终业务模式尚未被开创。我们不知道人工智能公司的价值单位——我所遇到的创始人仍在探索中。所以这是一个需要关注的领域，因为它关系到您如何构建、销售和变现您的产品。

我认为随着人工智能渗透到不同的领域和行业，我们会看到许多新的商业模式，这是很有趣的。以前也做过。SaaS和云公司已经把销售模式从按座销售应用程序改变为订阅制或广告驱动销售。





这里有四件事要考虑

01

寻找预算替代机会

当目标为运营支出而不是传统的软件预算时，这非常有趣，因为与针对现有软件供应商的资助竞争不同，你是在根据你能为客户的内部运营模型带来的效率进行竞争。

这种方法对于新类别的人机交互应用特别有效，例如AI呼叫中心、AI治疗师或AI审计员。这些应用更具颠覆性，因为你在创造尚不存在的东西——我们对此感到兴奋。

02

服务跨多个和跨工作流的个性化需求

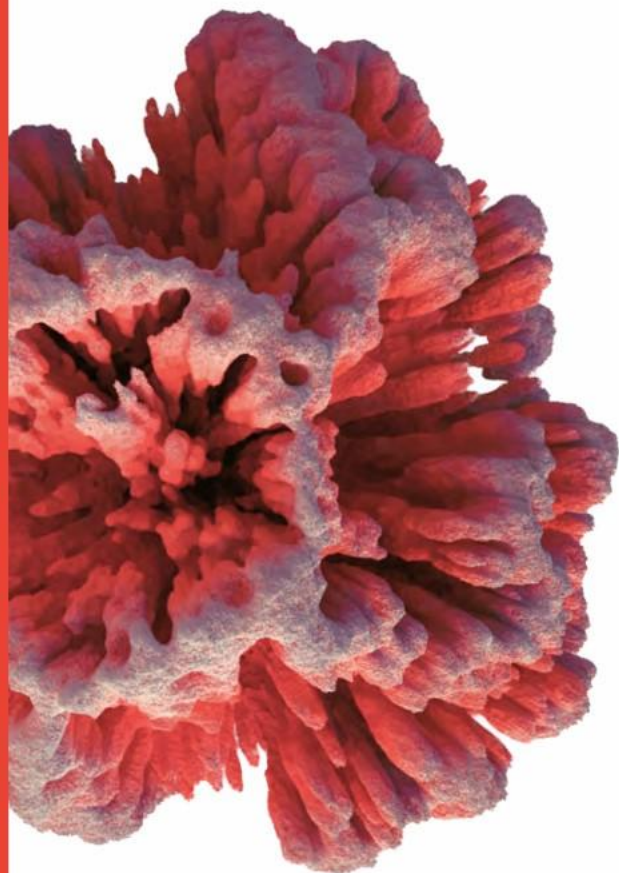
我们看到人工智能正在为大小企业中不同领域的一系列特定需求提供服务。在人工智能出现之前，企业不得不选择通用解决方案并为其进行定制。有了人工智能，您可以更容易并以较低的成本满足独特的工作流程。

我们期望AI应用能利用个人设备上的数据来理解个人兴趣、偏好以及访问公司数据（如系统记录和工作流程）的权限。我们投资组合中已有几家公司充当桥梁，帮助大型模型访问个人数据。然后AI可以学习每个企业的行为方式，并使用这些个人数据构建特定于该业务的应用。这样就解锁了价值，并通过基于不断变化的个人数据持续证明工作流程，从而创造了一个具有防御性的地位。我们认为在这些方面，初创公司相对于任何大型现有企业都具备优势。

v
im
W
ve

“

你们的竞争不再是与在任者竞争，而是与在任者的商业模式竞争。我相信，人工智能的最终商业模式尚未被开创。





03

面向整个组织的全新目标客户画像

我们鼓励初创公司考虑业务中的每一项功能，并真正理解他们的不同角色。寻找需要人工智能代理人或应用程序的信号，为特定的垂直领域或知识工作者类别，如开发者、会计师和律师。寻找相对于所创造的价值的高工资，以及重复的工作流程或技能。如果存在不匹配，那么市场可能存在需求。

它既有趣又可怕，因为已经花费了这么多时间来理解CIO或工程副总裁想要什么。有时候，对于某些SaaS公司，它是CFO或销售副总裁或人力资源副总裁想要的。现在你必须考虑业务中的每个功能。对特定角色的痛点有基本洞察力的创始人将拥有竞争优势。

最终，我们仍然必须找到客户并向他们传递价值。谁能更好地并以更快的速度把握用户体验，谁就会获胜。

04

试验并规范你的 AI 应用堆栈

没有人知道模型、嵌入、向量数据库、安全、代理框架等领域的规范AI应用堆栈将会是什么样子。我认为我们首先会看到开发者像今天一样构建AI应用，然后很快就会弄清楚如何利用新的工具堆栈来使他们的AI驱动应用真正运行良好。

我们看到开发者使用开源和云来尝试新技术，并在大规模采购前逐步采用。我们看到了很多创业公司在进行实验，最终开发者用他们的脚投票。他们会去最好的工具，所以我们关注他们的去向。

我们也看到初创公司在构建合成数据和其他模型，用于训练后评估，以确保模型以极高的精度运行。目前我们还有很多不知道的事情，直到技术成熟，人们将在评估中发挥重要作用。



Jia Li

联合创始人、总裁和首席人工智能官，LiveXAI



livex ai 提供人工智能客服，为快乐忠诚的客户服务。



Jia因在大型人工智能领域的领导力而被选举为IEEE Fellow。

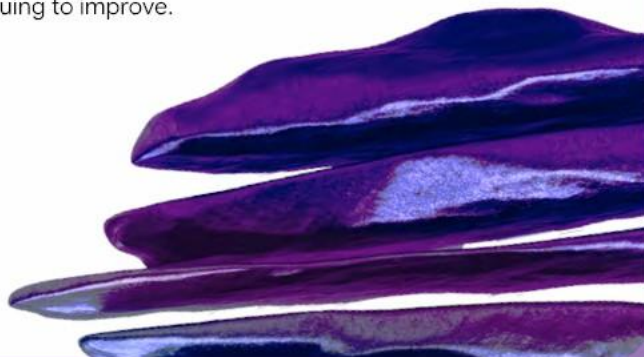
类人AI改变客户体验

人工智能正在发展以驱动更高效、个性化以及类人的体验

我今天观察到最令人兴奋的发展之一是能够模拟类人交互的AI代理的出现。作为人类，我们可以观察，我们可以交流，我们可以互动，我们可以行动，并且我们可以展现。虽然目前可能很难让一个模型和一种类型的数据拥有所有这些效果，但我们正在取得进步——类人的AI代理是未来的方向。

当前，AI代理可以直接从自然语言理解，甚至可以从多模态数据理解：如果你向AI展示你的产品，例如提示它“给我展示一些产品的问题”，它就能立即理解你的提示并完成你的请求。AI正变得非常有帮助——

and much more personalized—because AI's human-like reasoning, understanding, and interaction is continuing to improve.





使用AI代理重塑客户体验

这种朝向拟人化交互的趋势有潜力彻底改变客户体验。我认为这就是生成式AI驱动的AI代理的美好与力量所在：它能够理解个体客户的意图和情感，然后根据他们的意图帮助他们导航。

我相信这种朝着更高效、更个性化的交互进化的趋势之所以存在，是因为人工智能并不仅仅是关于模型的大小，或不同的架构，或计算能力。它也关乎数据。人们常常低估数据的力量，但它正是解锁人工智能真正潜能的关键。

许多基于规则的AI系统，即我所说的“AI 1.0”体验，依赖预定义的响应。想想那些有很多是或否问题的客户服务体验，我们可能按一些数字，然后由代理人引导我们沿着预定路线。

然而，更新型的人工智能代理能够理解自然语言，甚至能够解释多模态数据，例如图像或视频。这使得能够提供更加个性化的客户体验，因为人工智能可以识别一个人的问题、意图和情绪，从而带来更加自然和有帮助的交互。

应对人工智能代理的信任和安全隐患挑战

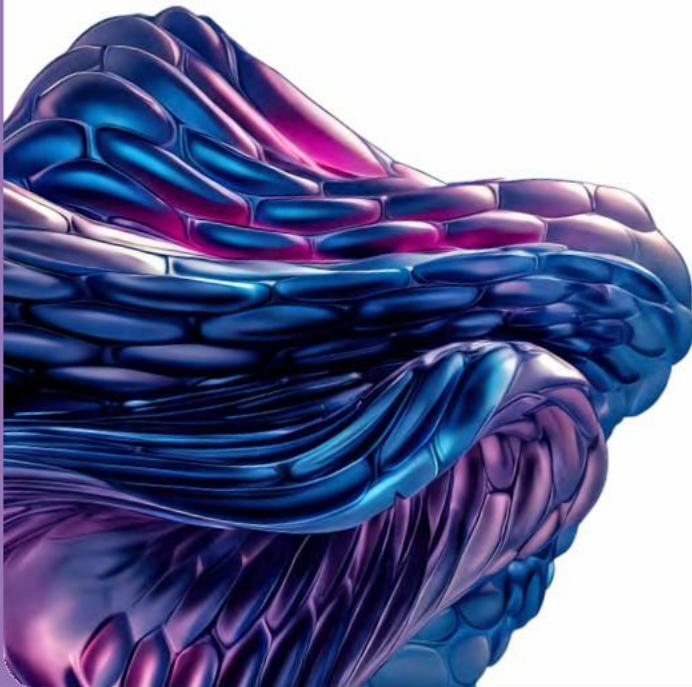
随着AI代理融入我们的生活，解决围绕信任、隐私和安全的新挑战至关重要。我们需要思考安全和隐私问题，问问自己：我们如何构建值得信赖的产品？我们如何处理输入到大型语言模型中的所有隐私和安全隐患？

在过去, AI或许能解决简单的 yes or no 类型问题。但现在, 大型语言模型正试图回答更复杂的问题, 尽管它们可能还不能为我们提供清晰或有益的解决方案。因此更复杂的AI有时会导致幻觉, 并引发对生成式AI合规性要求方面的担忧, 从而对信任、隐私和安全提出新的挑战。

一种有效且可行的信任和隐私处理方法，是利用模型来处理隐私保护——例如，在模型开始学习之前就剥离个人信息或敏感数据。此外，经过专门训练的模型可以处理提示注入，并适当应对不安全的提问类型。

“

我感觉，现在人们很关注模型的规模、不同的模型架构和计算能力。我始终试图强调的一个有争议的观点是数据的重要性。人们往往低估数据的威力以及不同数据类型对生成模型出色输出产生的影响。





提高准确性、延迟和理解能力 ，构建类人AI

我们需要关注人工智能代理的训练方式以及它们如何使用数据。了解人们将通过不同类型的模态和不同类型的行为与人工智能代理进行交互非常重要。有时人们可能希望所有行动都由人工智能代理来执行，有时人们可能更喜欢有一个更像是副驾驶或助手的人工智能代理。

我相信AGI的模型架构还不存在，数据也还没有到位。我们转而尝试使用单个专门类型的模型和微型代理来学习专门任务，并将它们整合起来以产生整体效应。

实时、类人的交互离不开提升准确性和降低延迟。一个AI代理是一个复杂的框架。每一步都需要时间，并且有多个步骤堆积在一起。所以现在，许多AI代理根据产品的复杂程度，耗费几分钟，甚至几十分钟。

利用改进的计算

计算将更加高效和可扩展。将会有新的模型架构，可以专注于用更高效的结构解决特定问题。

我们正在接近。例如，在我的一项项目中，我们能够在模型结构上加快平均 token 生成性能六倍。这使得非常强大。如果一项任务需要超过 10 秒钟，那么客户就会不耐烦，他们会放弃。因此，因为软件、算法和数据的新速度，初创公司现在可以赋能企业扩展其服务并服务每一位消费者。





吉尔·格林伯格追查



投资合伙人，CapitalG



吉尔负责字母表独立成长基金CapitalG的AI投资，投资包括Magic.dev和/dev/agents。



一位曾为私募股权支持的企业CEO，也是Y Combinator支持的公司创始人，自2019年以来，Jill还是斯坦福大学商学院的客座讲师。

从领航员世界到智能体

我相信我们很快将进入一个原始阶段，在那里人工智能足够智能以管理整个工作流程。那么企业如何适应？

虽然人工智能领域的许多人预期无限扩展将导致人工通用智能（AGI），但我认为基础模型在未来几年内将遇到瓶颈。这并非扩展无效；而是技术上的不切实际阻碍了我们。

想象构建一个拥有数百万亿参数的模型。我认为这将是一个极具吸引力的模型，可能实现AGI。但实际上，训练这种规模的模型非常困难。我们谈论的是在每个规模层次上都存在障碍——10万亿个token，100万亿个token等等。每一次修复都需要重新训练，这使得它成为一个资源密集、耗时漫长的攀登。

然而，这一限制为初创公司提供了一个有趣的机会。当巨头们忙于扩展时，初创公司可以专注于为客户带来切实的投资回报。这可以体现在基础设施方面，使其更有效率，或者在应用程序解决方案方面，这些解决方案不需要

training
tion side, building
massive models.





我正在人工智能代理时代投资

展望未来，我看到了两大主要趋势正在涌现，我计划投资于探索这些领域的天使公司

01 代理工作流程的兴起

我相信我们正从AI副驾驶时代迈向自主代理时代。这背后是基础模型，尤其是在推理能力方面取得的进展。这对公司的预算影响巨大，使他们不仅能节省软件成本，还能节省人力成本。

想象能够完全处理客户服务咨询的ImagineAI代理，而不仅仅辅助人类代理或配对程序员，以使开发人员效率大幅提升。这转化为一个价值数十亿美元的市场。初创公司在这个代理未来中扮演着关键角色。它们可以为特定市场（如客户支持、医疗保健、金融和法律）构建全栈代理，专注于复杂的控制和连接挑战。或者，它们可以创建平台，赋予个人构建自己的代理的能力，以满足个性化需求。



我真诚地相信，当所有这些代理东西运作起来时，它将允许人类 beings 花费更多的时间在高价值的事情上。

02 直接面向消费者代理

这将是消费者真正感到兴奋的地方。想象一下一个管理你日常事务的人工智能代理。我只需对人工智能代理说：“嘿，确保我准时参加会议”。你的代理会检查你的日程安排，计算出行时间，并安排Uber或Lyft，而无需你切换到不同的应用程序。

这有可能彻底改变应用开发。如果我们与一个AI代理互动，应用的价值将取决于其界面到其功能。我们甚至可能看到AI超级应用的出现，它们充当管理我们整个日常生活的中心枢纽，进而改变我们与技术互动的方式。

请注意，直销代理商的出现也引发了关于数据隐私和控制的问题。你对于不同的应用程序访问和共享你的信息感觉如何？这就是初创公司可以介入的地方，创建框架允许用户定义他们的偏好，并控制这些交互代理如何使用他们的数据。

构建你代理机构的战略策略在代理领域

投资者正在寻找那些从一开始就能证明具有强大产品市场契合度、独特竞争优势和清晰盈利路径的初创企业。这些基本面对应于不断变化的AI格局仍然至关重要。

我真诚地相信，当所有这一切智能体（agent）的工作都运转起来时，它将使人类有更多时间投入到高价值的事情上。想象一下这样一个世界：任何人都能将自己的想法变为现实，唯一限制因素就是他们的想象力。这就是让我兴奋的那种未来——我期待与更多致力于使这种未来成为可能的不凡创始人合作。



M a □ h i e u R o u



联合创始人兼首席执行官，Photoroom



Photoroom是一款人工智能驱动的照片编辑应用，可以即时调整您图像的大小以适应任何平台，去除并应用背景，并使用一键人工智能工具在几秒钟内编辑数百张照片。



斯坦福大学□□生，□□hieu此前曾□□□□□□□□件Replay的产品管理，该软件已被GoPro收购。

人工智能将解放创作者以进行创作

AI代理正在降低创作门槛，支持创意和故事
tellers

作为每天专注于帮助人们使用人工智能编辑照片的人，我正在看到各种各样的人在今天采用和使用人工智能。以前不擅长技术的普通人现在可以使用人工智能创造出他们想要的图像。t ges

虽然专家在AI提示方面可能存在差距，但其余大部分人群的AI使用已经是普遍现象。我认为，我们的责任是继续让tion, rtups AI更容易被所有人使用





解锁独特故事

我最兴奋的事情之一是人工智能将使更多的人可以随时随地与世界分享更多独特的作品。例如，创作者传统上需要出版商、生产商和推广者来将他们的歌曲、书籍或其他作品推向世界，而那些障碍在刚开始时尤其难以克服。

我看到人工智能代理现在正在减少创作障碍，并在创作者和故事讲述者完成所有分享其作品的必要活动时支持他们。想想一位厨师写一篇关于新菜谱的博客。创作作品是实际菜谱，但厨师仍然需要为菜谱撰写优秀的文案，并提供高质量的图片。人工智能可以帮助所有这些可能延误厨师发布新菜谱的部分。

我相信专注于帮助创意人士处理改善其故事中更战术性元素的人工智能初创公司，其作品的需求将会增加。同样，人工智能为人工智能科技初创公司打开了一个充满可能性的世界——既是人工智能技术的消费者，也是人工智能技术的提供者。正如人工智能代理可以自动化业务任务以帮助创意人士将其故事推向世界一样，代理也可以自动化营销、财务和客户支持任务，以帮助商家将其产品推向世界。所有这一切都会让创始人有更多精力专注于创新。



人工智能将使任何人都能从任何地方与世界分享许多更多独特的故事。

来自一位人工智能技术创业者的建议

在我作为人工智能初创公司创始人的旅程中，我学到了一些教训，其他人可能从中受益：

01 制定内容护栏

使用人工智能进行内容创作正在快速发展，数百万人在用它来做各种各样的事情。这对传统的作者身份和知识产权观念构成了挑战。我建议运用常识来规划如何让你的员工和人工智能技术的客户用人工智能来创作内容——不要走捷径。例如，我们只使用从创作者那里购买的已获得授权的内容来训练我们的模型，以避免法律和伦理问题。

02 理解对您的AI技术用户来说什么是重要的

今天，我们主要看到的是没有针对特定应用场景进行调整的通用AI。想想看，如果你是办公设计师，你不会为一家五人的科技公司创建与医疗诊所相同的办公设计。将一些有用或可用的东西交到人们手中，采访他们以了解缺失什么或如何提供额外价值，并相应地调整你的AI。





03 采取 AI 本地化方法

将您的公司建设为一个以AI为先的组织——不仅通过您的AI产品和服务，还要通过您的企业文化。当您定义公司文化的大支柱时，请思考AI将如何扮演其中的角色。例如，在Photoroom，我们采用了开放的组织文化来支持疫情期间的远程工作。我们没有通过直接消息进行沟通。由于没有任何东西是孤立的，它都为公司的AI模型提供数据。一切都可以访问。没有人需要重复提问，我们能围绕各种主题获得优秀的总结。这是一个随着时间的推移而不断增长的强大资产。

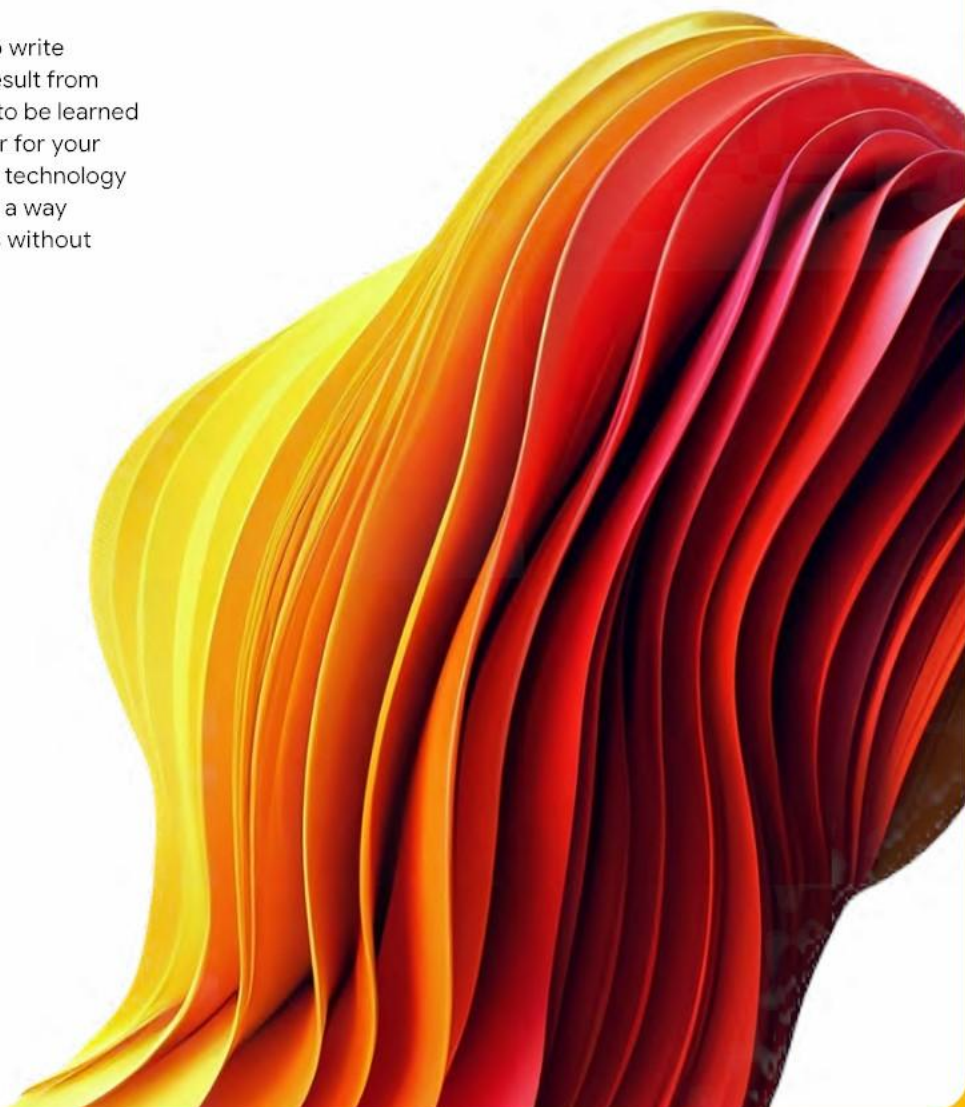
04 无提示设计AI

假设人们一无所知，并使用提示词来生成yourAI。AI提示词是一项技能，需要开发和不应被视为潜在用户基础。为了推广yourAI，设计用户体验帮助人们完成，无需使用提示词。

how to write
right result from
needs to be learned
a barrier for your
-driven technology
ence in a way
ir goals without

鼓舞人心的新设计趋势

展望未来，我对AI将如何释放新的设计趋势特别感兴趣。考虑字体设计的出现——人们用手写文字已有数百年历史。印刷机的发明导致了字体设计阶段的开始。随着计算机的出现，可用的字体设计数量爆炸式增长。类似地，我认为AI技术将推动新设计趋势和运动的开端，这些趋势和运动甚至今天我们都无法想象。有了AI，唯一真正的限制就是我们自己的创造力。





Mayada Goni



首席技术官和联合创始人，ThreadAI



线程ai是一个可组合式的人工智能编排平台，供组织设计、实施和管理他们的工作流。



作为首席技术官，梅雅达的焦点和主要职责包括开发、架构和执行Thread AI的技术战略。

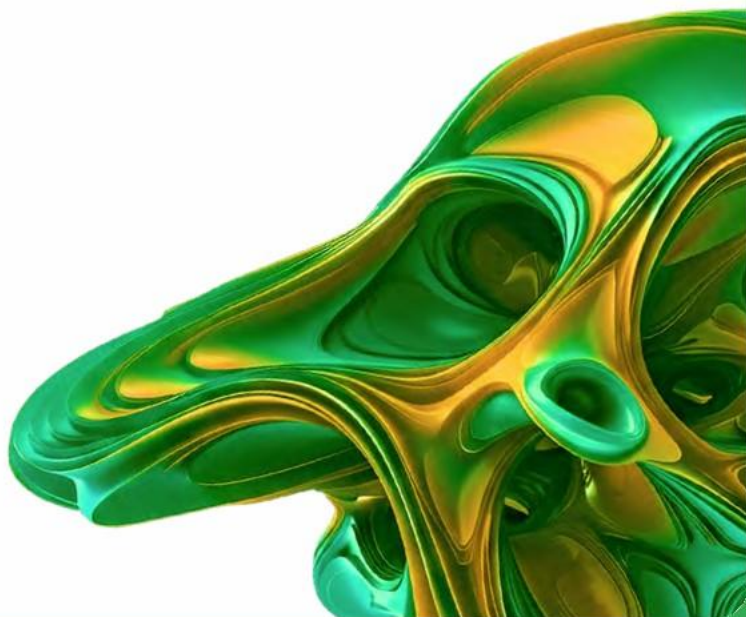


最近，Mayada在Palantir担任分布式系统工程师，主导开发了一个实时AI推理平台。在Palantir工作期间，她构建了复杂的AI基础设施系统，包括全面模型编排和评估框架的设计，以及复杂的容器化系统。

工作流和基础架构比闪亮的工具更重要

关注分布式系统基础，构建进化的人工智能层

我们正在目睹工具的突然爆炸性增长，这些工具试图解决价值链中一个非常小的部分。虽然每个人对AI都很兴奋，而且它开始解决一些实际问题，但当AI模型、半成品解决方案和一些华而不实的工具尚未证明它们实际上能产生影响时——无论是通过投资回报率还是通过从根本上改变开发者生态系统——它们就被匆忙推向市场。





耐用性、 workflow 编排和可观察性作为关键使能因素

随着公司意识到，“好吧，我不用把全部资金都烧光来嵌入基于人工智能的工作流程”，现在有回归基础的潮流。

我的论点是，AI绝不能作为一个独立的工作流——它始终应该是一个工作流的一部分。因此，对于可观察的工作流管理，集成和基础设施将成为AI堆栈中更加关键的部分。公司需要接口，允许它们将AI集成到现有工作流中，并与现有流程并行测试。

我认为在嵌入人工智能的位置上更加有意识地行动，并围绕可观测性构建工具，将成为我们构建这些生产-人工智能原生工作流的关键。回归对分布式系统基础的关注，并理解仅仅因为某件事是概率性的，并不意味着它是一种全新的范式。您只需要为人工智能设置正确的检查和平衡机制，或者您不要将其置于无法处理任何概率性结果的临界路径中。



公司意识到它们不必花费全部资金来嵌入一个基于人工智能的工作流程。人工智能永远不应该是一个独立的工作流程——它应该永远是工作流程的一部分。

建设不断发展的基础设施和数据库

人工智能的许多层——基础设施和数据库——都在不断发展。因此，在您制定人工智能创业公司的路线图和业务时，您是如何考虑这些层的？在您构建基础设施层时，考虑到这些模型将会随着时间的推移而迅速改进，您是如何考虑的？

我认为你想确保如果出现一个新版本的模型，你不必炸毁你的基础设施。你需要一个插件系统，其中事物是版本化的，也许你在协议级别与模型集成，而不是与每一个单独的SDK。你添加gRPC、REST或GraphQL的绑定，这样你就可以直接添加最新的模型。同样的事情也发生在不同的数据库技术中。





采取组合式方法对待A

我真心相信，采用模块化方法开发AI模型和基础设施，将最佳利用AI中的用例和技术，优化流程，例如在不同数据模态中服务并组合不同的较小模型，这将非常重要。因此，我建议AI初创公司在开发其产品和服务时采用这种方法。我还建议投资于数据准备，包括数据清理、策展和激活。这将帮助您最好地准备，快速将AI融入您的业务流程，并从您使用的模型中获得最大价值。

对更好的可解释性工具的需求

我希望可解释性工具能更早地出现。这些生成式AI模型的训练和后期运营化的方式完全割裂。一旦模型被训练、加权并后期托管以进行运营化，就很难回到一些数据上。在一些公司，模型在它们本不该训练的数据上进行了训练，这引入了潜在风险。如果有一种方法来跟踪训练过程中的一些来源或治理方式，那就太好了。目前有一些关于使用类似账本技术的研究，我认为这些都将非常重要，特别是我们现在看到很多围绕数据的法律案件。





拉维拉吉·贾因

帕克ner, Lightspeed



拉维拉杰是Lightspeed Venture Partners的合伙人，专注于投资人工智能、前沿技术以及更广泛的B2B/企业软件领域。



拉维拉吉在SnorkelAI、Skild AI、TypefaceAI和Nirvana等公司的董事会中任职。在2017年加入Lightspeed之前，他是领英的产品负责人，拥有哈佛商学院的MBA学位和孟买理工学院（IIT）的B.Tech学位。

准备好了吗。加大人工智能投资。

延迟人工智能采用会让企业面临变得不再重要的风险

我不相信人工智能已经准备好全面大规模实施——它进展得太快了。

生成式AI非常适合用于创造性场景，但幻觉、性能和安全问题将是未来几年需要许多新解决方案的核心挑战。除此之外，还有大量非结构化数据、未定义的流程以及目标不明确的问题。我们需要特定的工具，而这些工具的成熟需要很长时间。我不是说它不会实现，我只是说这比人们想象的要困难得多。

但如果公司选择现在忽略它，因为它发展太快，他们实际上将面临在核心业务中变得不再重要的风险——不仅是在人工智能的采用方面，还因为它可以根本性地改变对公司的价值的关键方面的成本结构、产出和质量。





关于人工智能，我们关心的一些事情，也希望你们同样关注。

01 成为AI本地化

企业低估了AI的潜力，即使他们的产品并非由AI驱动。要想成功，你需要优先考虑在所有职能上成为AI原生。你必须真正成为AI原生，才能说，“是的，我了解今天有哪些可用的事物以及它们将走向何方，以便我能构建出持久耐用的东西”。而且，将会有很多公司基于当前技术为顾客构建产品，而两年后它们将变得不再相关。

02 别等待完美

模型的进展正在不断加速。有人可能会说，“我为什么要今天尝试去实现AI？为什么不等两年，等它变得更好呢？”我不认为答案是那样的，因为我们正处在一个进化的早期阶段，事物会持续进化。你不能等这些模型变得更好才去使用它们；你必须持续地投入并与它们一同进化。

03 拥抱风险投资家思维模式

从某种意义上说，所有组织都必须像风险资本家（VCs）那样思考——尽管你受制于质量结果，但你仍然需要向前五到十年思考，以及你的组织和市场结构将如何变化，以及如何利用人工智能作为这一变化的关键，这并非简单地将人工智能添加到你的产品中，而是你的组织。差异化的向量在于领域理解和采用，这对于灵活、前瞻性组织来说是一个非对称的机会，以赢得胜利。

加速空间

我经常被问到哪些行业将被人工智能特别颠覆，坦白说，答案是所有行业。但我更倾向于说，人工智能做得最好的是在处理大量非结构化数据，而这适用于任何行业。从法律到保险到医疗保健，任何发生非结构化数据或语音相关交互的地方都将看到大量的自动化机会。

我确实相信，虽然应用层将看到更多的活动，但基础层将看到技术变革的最大飞跃。我对这些模型在物理世界中的潜力感到兴奋，例如机器人技术，以及它们将如何颠覆我们对数学和物理的理解。当你看到人工智能投资的规模时，很明显我们很快就会达到一个临界点。

这项工作并不仅仅是训练你能构建的最大模型。它关乎构建模型、赋能层和产品层，以高效地解决特定问题。会有一批高质量的巨型模型，但也会有大量针对特定工作更高效的精简模型。如何为这些模型构建最后一公里，将其与你特定的难题连接起来，并制定真正高效的分发策略来解决人们的问题？许多之前没有在实际中接触过软件开发的人，很快会在实际中开始使用人工智能。必须有人建造船只来帮助他们航行——无论是谁做到这一点，都将构建一家非常非凡的公司。



你无法等待这些模型变得更好才能使用它们；你必须持续地与它们共同投资和发展。





Salim Teja



合伙人，Radical Ventures，以及董事会成员，Aspect Biosystems，Promise Robotics，Intrepid Labs



萨拉姆是 Radical Ventures 的合伙人，这是一家专注于人工智能的领先风险投资公司。



在科技行业拥有超过25年的经验，Salim是一位积极的投资者，并领导着公司的Velocity项目，这是一个专门帮助创始人加速其AI优先企业成长的团队。

弥合差距：将AI转化为实际解决方案

重点是从可能创造出真正能帮助解决当今最大挑战的技术中转移注意力

作为一名人工智能领域的早期投资者，我看到了显著的进步。人工智能已经通过自动化任务来颠覆销售、营销、人力资源和软件开发等行业。随着企业认识到人工智能工具所能带来的巨大投资回报，这一趋势将只会加速。





我也预见AI将拓展到物理世界，超越仅仅是我们的屏幕并与我们周围的环境互动，革新以下领域：

01 机器人学

我看到人工智能与机器人技术的融合是技术在物理世界中的重大飞跃，并且我认为这将产生深远的影响。我相信，例如，机器人技术可以用于建筑，以使住房更容易获得且更经济。

02 材料发现

我相信这可以推动各个领域的创新，从制造业到建筑业。人工智能可以加速具有特定性能的材料发现，帮助我们为许多紧迫的问题开发新的解决方案。

03 空间计算

这是一个当前具有突破性研究的领域，涉及物理世界中的3D表示以及如何从这些表示中生成信息，以驱动空间计算方面的LLM类型能力。

在研究和应用之间取得平衡

目前人工智能领域正在进行一些令人难以置信的突破性研究，尽管将研究突破转化为实际解决方案存在重大挑战。投资者正在寻找识别了行业特定问题、可通过人工智能解决、并已开发出为客户创造切实价值的创新解决方案的初创公司。我密切关注着：

01 智能操作系统

人工智能代理将如何在更大的系统内进行交互和协作，以革新软件开发和技术？随着我们设计、部署和实现价值主张，智能代理世界将以极其有意义的方式影响每一个应用程序，我认为对于初创公司而言，考虑到互操作性进行构建并准备好利用其带来的商业化机会是很重要的。

02 数学推理

我相信不断推进人工智能的数学推理能力很重要。这将解锁科学发现的新可能，并推动进一步的创新。



当ai超越数据处理，并以真实、具体的方式与我们及物理世界互动时，其真正力量将被实现。





理解人工智能发展并构建你的护城河

我看到太多缺乏产品或技术护城河的 startup——这是一种难以复制的竞争优势。所以作为一名投资者，我开始担忧价值主张的持久性，以及这些 startup 如何随着时间推移而实现规模化。因此，我并不是说 startup 应该忽视研究，也不是说每个 startup 都需要有诺贝尔奖获得者发明新的 AI 研究领域。

我认为，创业公司专注于与其技术领域相关的研究是明智的，在那里他们可以发现一些有助于创建竞争壁垒的技术。这意味着需要内部有人了解人工智能技术格局，并能识别差异化的机会。我认为，了解你领域的技术发展是重要的，这样你才能利用新发现来提升你的优势。

t

聚焦投资于实施

在一般情况下，我们的重点是从事什么是人工智能能够实现的令人着迷的事情，转移到我们如何能够在大规模上将其在工业中实施。在人工智能中，很容易沉迷于研究和可能性的美妙之处，但我们有时会忽视采用过程的艰难——这不是微不足道的。

我的关注点在于寻找能够帮助传递价值、建立真正企业，并使用人工智能解决现实问题的初创公司，例如通过建筑领域的机器人技术解决住房危机。这些都是有意义的问题，无论他们是否身处我们的行业，都将影响所有人。

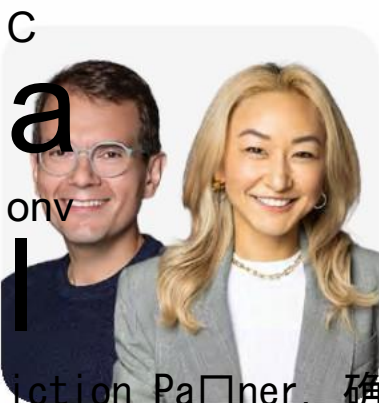




萨拉胡郭

创始人兼合伙人

+ 迈克·维尔恩



郭莎拉是Conviction（一家成立于2022年的AI原生风险投资公司）的创始人 and 合伙人。此前，她在Greylock担任普通合伙人。她曾作为早期投资者和合伙人参与了包括Figma、哈维（Harvey）、西尔维亚（Sierra）、Hey Gen、mistral、Cognition、Stackblitz等50多家公司的投资。郭莎拉来自威斯康星州，拥有宾夕法尼亚大学的四所学位，与丈夫和三个孩子一起居住在湾区。她与Elad Gil共同主持AI播客《No Priors》。



迈克·弗内尔是康维信的合伙人，这是一家成立于2022年的风险投资公司，服务于下一代创始人。此前，他是红杉资本的投资者，与20多家公司合作，包括Rippling、Notion、Verkada、Clay、Statsig等。他在微软的产品和工程岗位上开始了他的职业生涯，之后成为Facebook团队的早期成员，最终领导了大型产品团队和工程团队。他在哈佛大学学习计算机科学，在那里遇到了他的妻子，他会在空闲时间和他的三个年幼的孩子一起打棒球。

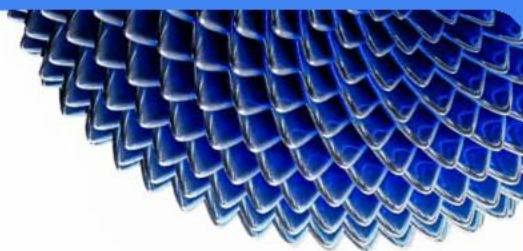
人工智能正在实现公平竞技

人工智能正在重写科技行业的规则，提供灵活和富有创造性的初创企业一次终生竞争当今领导者的机会

人工智能创造价值的潜力巨大，它正在加速全经济领域的技术采用，并改变许多行业的竞争维度。

我们现在在法律、医疗保健、制药和国防等领域看到了这一点。这些行业面临着巨大的变革。它们的输入是信息和知识工作。文件处理量很大。颠覆、自动化和民主化的机会非常巨大。风险投资中许多被接受的观点正在受到挑战——从哪些市场有吸引力到商业模式甚至技术公司的范围。

越来越多的科技公司开始从事服务行业，或者为其客户提供更多能力，而不是从现有的工作流程中挤出更多效率。竞争基础的变化使每个行业都更具动态性。这对初创企业来说是一个巨大的机遇。当技术堆栈的使能层以极快的速度发展时，正如目前基础模型所展现的那样，那么为了做出投资决策，人们必须对“限制者”持有一个观点——即发展和技术采用的限制因素。



一些我们思想的原则：

01 项目能力足够

要确定一家公司是否在 AI 方面采取主动布局，我们建议进行一个简单的测试：当发布新模型时，你是感到高兴还是悲伤？如果你感到悲伤，你很可能做错了什么——如果你只是正在构建模型目前能力的边界，那么下个模型有可能会涵盖那部分内容。你应该努力让你的产品设计能够热切地预见并利用新的能力。

02 遵循数据

虽然广受欢迎，但认为拥有海量数据的大公司将自动主导人工智能领域这一假设往往会被证明是错误的。现有企业的数据可能并不适合用于训练人工智能模型，而且法律约束可能会限制其使用。最有用的数据往往是针对特定实际应用场景的逻辑推理轨迹和评估方法。在许多情况下，这类数据尚未被任何参与方收集，领域对新进入者敞开大门。

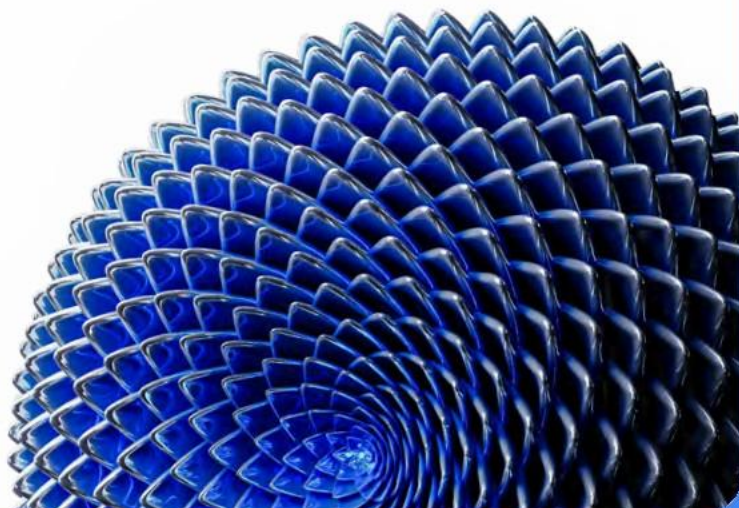
03 优先考虑第一性原理思维

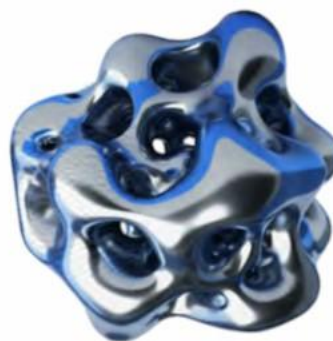
在一个像人工智能这样的流体环境中，有很多观点快速变化的讨论。与其盲目跟随投资者或研究实验室的主流叙事，不如让你的决策建立在技术和市场的根本理解之上。

04 伟大的创始人和产品，永远

我们寻找的潜在投资存在一个分支：一类是符合我们口味和理解的优质产品。另一类是早期阶段的人才投资，他们行动迅速，品味良好，理解客户，并快速迭代以解决问题。这长期以来都是成功的秘诀，现在借助人工智能，我们又有了新的工具，供他们围绕产品进行建设。

当我们遇到既对某个特定领域有世界级理解，又具备所有核心创业特质且行动迅速的人时，我们会特别兴奋。最重要的是能够快速迭代，尤其是在环境快速变化的情况下。你需要的是富有创造力和实验精神的人，他们能够找出技术的边缘地带，学习它们，并持续推动。





领域特定模型解锁新机遇

每个人现在都（恰当地）关注基础模型的力量。但我们正处在一场持续10到20年应用浪潮的起点，其中许多我们尚无法预测。

我们期望随着时间的推移，会看到许多特定领域的模型被使用，涵盖多种方法：预训练、后训练和针对特定领域进行微调。一些领域，如生物学、材料科学或机器人学，将需要突破高效数据收集和生成方面的技术。



当技术栈的底层以惊人的速度发展，正如目前基础模型所示，中间层最终成为一个“调节器”，限制应用程序开发的速度。

这个100人，1000亿美元的公司

人工智能像个乘数，增强了人类的能力，但它仍然需要人类来引导它。虽然我们并不认为成长型公司中的整个职能会被外包，但我们相信这个乘数可以根本性地提高业务质量。我们投资组合中的许多创始人都在积极利用人工智能。我们的方法是鼓励人们不要低估小团队所能取得的成就。



Yoav Shoham



斯坦福大学计算机科学名誉教授及AI21 Labs联合创始人



Yoav是一位领先的AI专家，也是三家前AI公司的创始人，这些公司被Ariba/SAP和Google（两家）收购。



Yoav获得了多项学术奖项，是人工智能进步协会、计算机协会和博弈论学会的 Fellow。

停止提示和祈祷，并开始构建完整的AI系统

也就是说，如果你关心可控性、可靠性和效率

说实话，我讨厌AGI这个词，或者叫人工通用智能。这不存在的。我见得多了，知道智能是多方面的。机器无疑将能自动化越来越多的功能，但有虚假的感觉，似乎有一个特定的点，那个神话般的AGI将被达到。我认为这种松懈的思考和围绕AGI的炒作是个分心。

相反，让我们具体谈谈人工智能技术及其优缺点。





为什么企业正疯狂地试验人工智能，但对大规模部署它却很谨慎

您在企业中看到这一点。虽然消费者对AI的采用达到了创纪录的速度，但商业的采用却较慢。显然，各地的CEO和董事会都在展示他们的公司是“AI优先”（或计划成为那样），并且正大力进行试验，有时有数百个用例。

但就企业中进行的所有大规模实验而言，只有一小部分AI项目实际上得以部署。这归结为两个关键挑战：

01 成本

运行 LLM 的高昂成本挑战了商业软件的经济性。

02 大型语言模型会生成奇妙的结果，以及大量的胡言乱语

设想一下撰写投资备忘录或回复客户，你的AI有95%的时间表现出色，但另外5%的时间却产出垃圾。这在企业中是致命的。

第一个挑战是由于（更不用说训练）大语言模型（LLMs）的服务固有成本，并将通过两种方法的组合来处理。一种方法是使用更小的大语言模型（术语“小语言模型”或SMLs正在流行），那些“微型”参数少于7B甚至少于3B的模型。另一种方法是使用与标准Transformer架构不同、更高效的结构；AI21的Jamba系列模型就是一个例子。

第二个挑战尤为尖锐，也更具挑战性。这种现象是LLM概率性质固有的，无论在“对齐”、“护栏”等方面投入多少努力，认为这会消失都是一种虚幻的愿望。我相信我们将逐渐摆脱我所称的当前“提示并祈祷”的操作模式。行业将意识到LLM是一个更全面AI系统的一个要素，该系统可以无缝集成并利用各种AI技术的优势，包括LLM、检索、工具和其他传统代码。AI系统将提供更大的控制力、效率、可靠性，尤其是在处理需要非凡推理的任务时，而大多数任务都需要这样。



人工智能系统在可靠性和效率方面将大幅优于基础的大型语言模型，但它们不会完美。小型企业家将构建利用这些系统优势同时考虑其局限性的产品。





健康实践对于初创公司：使用人工智能系统，但专注于“产品算法”适配

结合多个大语言模型和其他工具的AI系统提供了一个有吸引力的解决方案。这些系统通过智能地将任务路由到最合适的资源，从而实现更好的成本和计算管理。例如，一个较小的大语言模型可以作为“路由器”，将任务导向专门的大语言模型或工具以实现最佳效率。这些系统还可以通过在计算过程中引入制衡措施来提高可靠性和质量。

但当你将这些AI系统应用于现实世界问题时，你应该明智地对待这一点。我给AI创业领导者常见的建议是追求“产品-算法”的契合。我所说的意思是，虽然AI系统在可靠性和效率方面将比基础的LLM有显著提升，但它们仍然是不完美的；LLM调用、搜索和检索中涉及的潜在不确定性将不会完全消失。因此，作为一名正在开发新产品的企业家，要理解这项技术的优势和劣势，并以利用其优势、弥补其不完美的方式来打造产品。这就是我所说的“产品-算法契合”。





构建未来，使用Google Cloud



今天，全球超过60%的已融资生成式人工智能创业公司和90%的全球生成式人工智能独角兽公司都是谷歌云的客户。我们很自豪，能够运用我们悠久的AI领导力和创新历史，为客户独特的机遇和挑战提供尖端云AI解决方案。

在谷歌云，我们与研究人员、创始人、初创公司、企业、合作伙伴以及公共部门机构合作，批判性地思考我们的负责任的AI解决方案如何持续满足员工、客户、患者和公民的需求。这包括在最前沿的创新中提供世界一流的设施和全栈能力，深入参与数据、代理和应用程序的发明者，以帮助将新的成果变为现实，并与企业合作评估AI的进步如何使组织和组织外部的体验实现现代化。生成式AI带来的飞速变化意味着初创公司面临着前所未有的挑战——而我们在这里提供帮助。

达伦·莫里

全球初创企业首席执行官，谷歌云





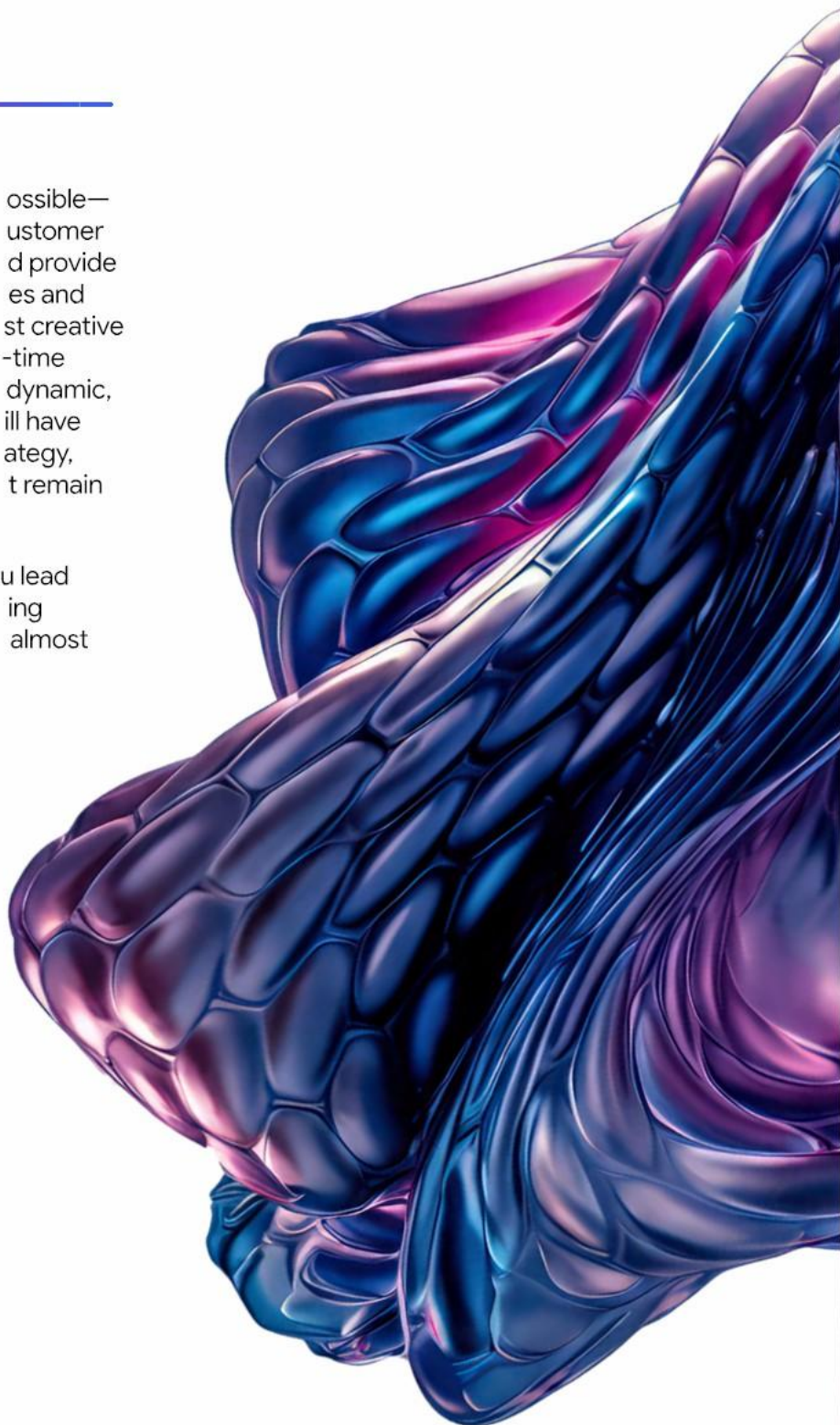
“

人工智能允许实现前所未有的市场推广个性化水平，可创建直观的个性化客户旅程，分析用户行为并提供实时定制化推荐和消息。未来将看到生产、定制激活和真实测量协同工作在一个相互连接的系统中。初创公司将有更多时间专注于创意、战略和真实的客户联系，这些都是驱动参与度的核心要素。

ossible—
ustomer
d provide
es and
st creative
-time
dynamic,
ill have
ategy,
t remain

我们很高兴与您合作，作为在快速变化的u lead
环境中，什么是一天是真的，当然不是第 ing
第二天是真的生成式AI。 almost

艾利森·瓦格纳菲尔德
CMO，谷歌云





加速您的旅程 升级。

立即预约您的生成式AI咨询。

→ [立即注册](#)

通过 Google for Startups 云计划，最高可
获 35 万美元云积分。

→ [立即申请](#)

联系我们Startup销售团队。

→ [联系我们](#)

