

# 人工智能现状报告

2024年10月10日

内森·贝纳奇

## 关于作者



### 内森·贝纳奇

Nathan 是 Air Street Capital 的普通合伙人，Air Street Capital 是一家投资第一批公司的风险投资公司。他负责管理研究和应用人工智能峰会 (RAAIS)、RAAIS 基金会 (资助开源人工智能项目)、美国和欧洲的人工智能社区以及 Spinout.fyi (改善大学衍生创造)。他在威廉姆斯学院学习生物学，并作为盖茨奖学金获得者获得了剑桥癌症研究博士学位。

## 关于作者



### 亚历克斯·查尔莫斯

Alex 是 Air Street Capital 的平台负责人，定期通过 Air Street Press 撰写关于人工智能的研究、分析和评论。在加入 Air Street 之前，他是 Milltown Partners 的副总监，为大型科技公司、初创企业和投资者提供政策和定位方面的建议。他于 2017 年毕业于牛津大学，获得历史学学位。

人工智能(AI)是一个科学和工程的多学科领域，其目标是创造智能机器。

我们相信，在我们日益数字化、数据驱动的世界中，人工智能将成为技术进步的力量倍增器。这是因为今天我们周围的一切，从文化到消费品，都是智慧的产物。

《人工智能状况报告》现已进入第七个年头。把这份报告看作是我们所看到的最有趣的事情的汇编，目的是引发一场关于人工智能的状态及其对未来的影响的知情对话。

我们在报告中考虑了以下主要方面：

- 研究:技术突破及其能力。
- 行业:人工智能的商业应用领域及其商业影响。
- 政治:人工智能的管理，其经济含义和人工智能的地缘政治的演变。
- 安全:识别和减轻高能力的未来人工智能系统可能给我们带来的灾难性风险。
- 预测:我们认为未来12个月将发生的事情，以及2023年的绩效评估，以保持我们的诚实。

由内森·贝纳奇和空气街资本团队制作

## 定义

人工智能(AI):一个广泛的学科,目标是创造智能机器,相对于人类和动物表现出的自然智能。

人工一般智能(AGI):一个用来描述未来机器的术语,这些机器可以在所有有经济价值的任务中匹配并超越人类认知能力的全部范围。

人工智能代理:一个人工智能驱动的系统,可以在环境中采取行动。例如,一个LLM可以使用一套工具,并且必须决定使用哪一个来完成它被提示要做的任务。

人工智能安全:研究并试图减轻未来人工智能可能给人类带来的风险(从轻微到灾难性)的领域。

计算机视觉(CV):程序分析和理解图像和视频的能力。

深度学习(DL):一种受大脑神经元如何识别数据中的复杂模式启发的人工智能方法。“深度”指的是当今模型中的许多层神经元,它们有助于学习数据的丰富表示,以实现更好的性能增益。

扩散(Diffusion):一种算法,用于迭代去除人为破坏信号的噪声,以生成新的高质量输出。近年来,它一直处于图像生成和蛋白质设计的前沿。

生成式人工智能:一系列人工智能系统,能够基于“提示”生成新内容(例如,文本、图像、音频或3D资产)。

图形处理单元(GPU):一种半导体处理单元,能够实现大量并行计算。历史上,这是渲染计算机图形所必需的。自2012年以来,GPU已经适应了训练DL模型,这也需要大量的并行计算。

## 定义

(大型)语言模型(LM, LLM):一种在大量(通常)文本数据上训练的模型,以自我监督的方式预测下一个单词。术语“LLM”用于表示数十亿参数LMs,但这是一个动态定义。

机器学习(ML):人工智能的一个子集,通常使用统计技术来赋予机器从数据中“学习”的能力,而无需明确给出如何学习的指令。这个过程被称为使用学习“算法”来“训练”一个“模型”逐步提高特定任务的模型性能。

模型:根据数据训练的ML算法,用于进行预测。

自然语言处理(NLP):程序理解人类口头和书面语言的能力。

Prompt:通常用自然语言编写的用户输入,用于指示LLM生成某些东西或采取行动。

强化学习(RL):ML的一个领域,其中软件代理在一个环境中通过试错来学习面向目标的行为,该环境根据他们实现目标的行为(称为“策略”)提供奖励或惩罚。

自我监督学习(SSL):一种非监督学习形式,不需要手动标记数据。相反,原始数据以自动方式被修改,以创建可供学习的人工标签。SSL的一个例子是通过屏蔽句子中的随机单词并试图预测丢失的单词来学习完成文本。

变压器:一个模型架构的核心,最先进的(SOTA)ML研究。它由多个“注意力”层组成,这些层了解输入数据的哪些部分对给定的任务最重要。Transformers始于NLP(特别是机器翻译),随后扩展到计算机视觉、音频和其他形式。