# GPT-4: 新一代人工智能语言模型

*\*OpenAI最新的多模态大型语言模型\**

![GPT-4 Banner](https://image.blocktempo.com/2023/03/chagpt-gpt4-6410b6078d3af-sej.png)

## 什么是GPT-4?

GPT-4（Generative Pre-trained Transformer 4）是由OpenAI开发的第四代预训练生成式变换器模型，于2023年3月14日正式推出。作为一款多模态大型语言模型，GPT-4不仅可以处理文本输入，还具备处理图像输入的能力，代表了人工智能领域的最新进展。

相比前代GPT-3.5，GPT-4在可靠性、创造性以及处理复杂指令的能力方面都有显著提升。它能够理解并生成自然语言，解答各类问题，编写代码，以及分析图像内容等多种任务。

<video controls width="100%">

  <source src="https://test-1343498553.cos.ap-guangzhou.myqcloud.com/videoplayback.mp4?q-sign-algorithm=sha1&q-ak=AKIDRzsBgkfzu9FLxgFkbwT8C5mzdACm3GqMuHnQLpsavTaSN\_2bQCG-3\_DccXZdpnd-&q-sign-time=1743040499;1743044099&q-key-time=1743040499;1743044099&q-header-list=host&q-url-param-list=&q-signature=8e1b1db34b802148d5b73d148a359e1c5cdb061d&x-cos-security-token=fnoIjkHd09NhOCSTgTPFWT37hUmzMohafeda535ef2de40690f9b9735e8166c80rWENt7fvE1EzM30aHKE17M2ogyfKdsZDyqQ1IQKqyhnyuSm3WWgOyFiueHffnGPla0myEpeOhPRUbE2CRH0t-jpmipVjF-PQl8y6awIsF\_ZMh5C\_Ipa8QA7-c5FrtJvdU5I\_saUlxowkOXJ8W6cFx\_AZoDYKyb\_8BIS7TzqaFa3qZG5C4cRCuml6cNLAXZZH" type="video/mp4">

</video>

## GPT-4的技术特点

### 多模态能力

GPT-4可以接收文本和图像作为输入，使其能够解释图像中的幽默内容，总结屏幕截图中的文本，以及回答包含图表的考试问题。

![多模态能力展示](https://www.damoxing8.com/wp-content/uploads/2024/08/frc-21067ee220ff87b35d6c0c327e291a96.png)

### 更大的上下文窗口

GPT-4提供了8,192和32,768个标记的上下文窗口版本，比GPT-3.5的4,096个标记有了显著提升，能处理更长的文本。

![上下文窗口示意图](https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRmRv6XM0IdMoJOumnPsRrH9C7omr\_l-HmJDg&s)

### 系统消息控制

引入了"系统消息"功能，允许以自然语言指令指定GPT-4的语气和任务，增强了对模型行为的控制。

![系统消息示例](https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRpzTH0EL4UIwg9Fl1YQzQ98-TdI7-5BdJT3w&s)

## 性能与能力

OpenAI声称，在他们的内部测试中，GPT-4在SAT考试中获得了1410分（第94百分位），LSAT考试中获得了163分（第88百分位），以及统一律师资格考试中获得了298分（第90百分位）。相比之下，GPT-3.5在相同考试中的表现分别只达到了第82、第40和第10百分位。

在创造性思维的托兰斯测试中，GPT-4的原创性和流畅性得分位于前1%，而灵活性得分则在第93至第99百分位之间。

![GPT-4与其他AI模型的性能比较](https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQE5-IoVh3HRNgO\_47Tblh5OplrUevtDoiQzg&s)

| 测试类型 | GPT-4得分 | GPT-3.5得分 | 人类平均水平 |

| ------- | --------- | ---------- | ---------- |

| SAT考试 | 1410分（第94百分位） | 第82百分位 | 1060分 |

| LSAT考试 | 163分（第88百分位） | 第40百分位 | 151分 |

| 统一律师资格考试 | 298分（第90百分位） | 第10百分位 | 280分 |

## 应用场景

GPT-4凭借其强大的语言理解和生成能力，已被应用于多种场景：

- **\*\*ChatGPT Plus\*\*** - OpenAI的付费聊天服务，提供GPT-4支持的高级对话功能

- **\*\*Microsoft Copilot\*\*** - 微软的智能助手，利用GPT-4提供搜索和内容创作服务

- **\*\*Duolingo Max\*\*** - 语言学习应用中的高级功能，解释错误并提供会话练习

- **\*\*Khan Academy的Khanmigo\*\*** - 教育平台上的智能辅导服务

- **\*\*冰岛政府的语言保护项目\*\*** - 帮助保护冰岛语

- **\*\*Be My Eyes\*\*** - 视障人士辅助应用，利用GPT-4识别图像和导航

- **\*\*医疗领域\*\*** - 微软和Epic Systems为医疗提供商提供GPT-4支持的系统

## 对比前代模型

GPT-4相较于GPT-3.5有显著提升：

| 特性 | GPT-3.5 | GPT-4 |

| ---- | ------- | ----- |

| 上下文窗口 | 4,096个标记 | 8,192或32,768个标记 |

| 模态支持 | 仅文本 | 文本+图像 |

| 幻觉现象 | 常见 | 比GPT-3.5减少60% |

| 不当回应 | 易受诱导 | 比GPT-3.5减少82% |

| 创造性表现 | 良好 | 优异（托兰斯测试前1%） |

## 限制与挑战

尽管GPT-4取得了重大进展，但仍存在一些重要的限制：

- 可能出现"幻觉"现象，即生成不在训练数据中或与用户提示相矛盾的信息

- 缺乏决策过程的透明度，事后解释可能与实际过程不符

- 在抽象推理基准ConceptARC上表现较弱，在所有类别中得分低于33%

- 可能表现出认知偏见，如确认偏见、锚定效应和基础率忽视

- 训练和技术细节不透明，限制了开放研究

*> "尽管GPT-4在自然语言处理方面表现出色，但仍须谨慎看待其在关键任务中的应用，特别是在医疗等高风险场景。" - 微软研究团队*

## 未来发展

随着OpenAI发布GPT-4o（"o"代表"omni"），AI技术继续快速发展：

- GPT-4o支持实时处理并生成文本、音频和图像等多模态输出

- 响应速度显著提升，接近人类反应时间

- 非英语语言性能大幅提高

- 视觉和音频理解能力增强

- 统一模型使其比前代更快速、成本更低、效率更高

![AI未来发展趋势](https://nsysgroup.com/media/bg2ecogo/ai-development-transforms-mobile-industry1.png)

---

© 2024 GPT-4技术介绍 | 视频来源: What can you do with GPT-4.mp4