

# ChatGPT 能力对比 (GPT-3.5 vs. GPT-4)

GPT-4 语言模型于 2023 年 3 月 14 日正式发布。虽然 GPT-4 并未实现一些此前传闻中所宣传的功能，但该模型在很大程度上改进了 GPT-3.5 及其以往模型的表现。那么基于 GPT-4 的 ChatGPT 与基于 GPT-3.5 的 ChatGPT 究竟有何不同呢？

## (一) 创造力。GPT-4 相较于 GPT-3.5 最显著的优势之一便是其在回应提示时更具

创造力的能力。尽管在解决基本问题时，GPT-4 的创造性优势可能并不明显，但随着任务难度的增加以及对更高创造力水平的要求，两款模型间的创造力差异便愈发显著。例如，如果让两个模型完成一个创意任务，如编写一首诗，要求每行诗中都使用中文和中文，那么基于最新 GPT-4 模型的 ChatGPT 将提供更好的结果。GPT-4 的回应将在每一行中使用两种语言，而 GPT-3.5 则会交替使用两种语言，一行使用一种语言，下一行使用另一种。

## (二) 视觉输入。与仅能接受文本提示的 GPT-3.5 不同，GPT-4 是多模态的，能够

接受文本和视觉输入。并不仅限于键入提示的图像，而可以是任何图像。因此，从手写数学题的图像到 Reddit 论坛表情包，GPT-4 几乎可以理解和描述任何图像。与 GPT-3 不同，GPT-4 既是一个语言模型，也是一个视觉模型。在 GPT-4 发布直播期间，OpenAI 工程师向模型输入了一个聊天软件的截图。GPT-4 能描述其中的每一个细节，包括当时在线用户的名字。此外，还向模型输入了一个手绘的笑话网站模型图，并指示将其转化为网站。令人惊讶的是，GPT-4 为该图像相符的网站提供了一个可运行的代码。

## (三) 回复安全性。尽管 GPT-4 并非完美，但它采取的确保更安全回应的措施相较于

GPT-3.5 模型来说是一个升级。在 GPT-3.5 中，OpenAI 采取了更注重监管的安全方法。换句话说，一些安全措施更像是事后考虑的。OpenAI 监控用户

的操作和提问，发现漏洞，并在过程中尝试修复它们。而在 GPT-4 中，大多数安全措施已经在模型层面内置于系统中。为了理解它们之间的区别，可以将其比喻为从一开始就使用坚固材料建造房屋，而不是随意使用任何材料，然后在出现故障时进行修补。根据 OpenAI 的 GPT-4 技术报告，GPT-4 仅在 0.73% 的情况下产生有毒回应，而 GPT-3.5 的有毒回复占比为 6.48%。

**(四) 回复的事实正确性。**GPT-3.5 的一个缺陷是它有时会自信地产生荒谬和不真实的信息。在人工智能领域，这被称为“AI 幻觉”，可能导致对 AI 生成信息的不信任。在 GPT-4 中，幻觉问题仍然存在。然而，根据 GPT-4 技术报告，与 GPT-3.5 模型相比，新模型产生幻觉的可能性减少了 19% 至 29%。但这不仅仅是限于技术报告。在 ChatGPT 上，GPT-4 模型的回应显著地更有事实一致性。

**(五) 对话（输入）的长度。**GPT-4 和 GPT-3.5 之间的技术主要差异之一是上下文窗口和上下文大小。上下文窗口是指模型在聊天过程中可以在其“记忆”中保留多少数据以及保留多长时间。GPT-4 在上下文大小和窗口方面相较于其上代模型有显著的改进。实际上，这意味着 GPT-4 能够更好地记住更长时间的对话上下文以及在对话过程中给出的指令。GPT-3.5 的一个问题是，随着对话的进行，模型容易偏离主题或未能遵循指令。例如，您可以告诉模型用您的名字称呼您，它会这样做一段时间，但在过程中可能会忽略该指令。尽管 GPT-4 模型仍然存在这个问题，但由于上下文窗口的改进，这个问题已经减轻了。另一个问题是一次性使用的文本量限制。使用 GPT-3 对长文本进行摘要通常意味着将文本拆分成多个部分，然后逐个进行摘要。GPT-4 模型在上下文长度方面的改进意味着您可以一次性粘贴整个 PDF 文件，并让模型在不将其拆分成多个部分的情况下进行摘要。