

大型語言模型的最大上下文長度

我最近使用了 DeepSeek API 來生成提交信息，如AI 驅動的 Git 提交信息所述。

當提交涉及許多更改的文件時，DeepSeek API 報告輸入超出了其上下文長度限制，即 65,535 個標記 ($2^{16} - 1$)。

以下是其他一些模型的上下文窗口大小：

- **Claude 3 系列**：2024 年 3 月推出，這些模型的上下文窗口從 200,000 個標記開始。
- **GPT-4**：標準版本支持 8,192 個標記，而擴展版本（GPT-4-32k）支持 32,768 個標記。
- **Meta 的 LLaMA 2**：標準版本支持 4,096 個標記，但微調版本可以處理多達 16,384 個標記。
- **Mistral 7B**：支持多達 8,000 個標記。