

लार्ज लैंग्वेज मॉडल्स की अधिकतम संदर्भ लंबाई (Context Length)

मैंने हाल ही में GPT-4 का उपयोग करके एक कमिट संदेश जनरेट किया, जैसा कि GPT-4-32k में बताया गया है।

जब एक कमिट में कई बदले हुए फाइलें शामिल होती हैं, तो GPT-4 ने बताया कि इनपुट उसके संदर्भ लंबाई सीमा 65,535 टोकन ($2^{16} - 1$) से अधिक हो गया है।

यहां कुछ अन्य मॉडल्स की संदर्भ विंडो साइज़ हैं:

- **GPT-3**: मार्च 2024 में पेश किए गए, इन मॉडल्स की संदर्भ विंडो 200,000 टोकन से शुरू होती है।
- **GPT-4**: स्टैंडर्ड वर्जन 8,192 टोकन का समर्थन करता है, जबकि एक्सटेंडेड वर्जन (GPT-4-32k) 32,768 टोकन का समर्थन करता है।
- **GPT-4o**: स्टैंडर्ड वर्जन 4,096 टोकन का समर्थन करता है, लेकिन फाइन-ट्यून्ड वर्जन 16,384 टोकन तक संभाल सकता है।
- **GPT-4o mini**: 8,000 टोकन तक का समर्थन करता है।