

GPT モデルアップデート 0613 バージョンリリース

[\(この PDF は閲覧自由です。社内やお知り合いに転送してご利用ください\)](#)

2023 年 6 月 13 日に GPT モデルがアップデートされました。

大きな改善点がいくつか含まれていて、注目を集めています。この記事ではその改善点をわかりやすくまとめました。

本題に入る前に、頻出単語の確認をしていきましょう。

GPT モデル

OpenAI が開発した大規模言語モデルです。現在は gpt-3.5-turbo や gpt-4 といったモデルがリリースされています (gpt-4 の API は 2023/6/20 現在予約待ち)。

ChatGPT で使われているモデルです。

トークン

テキストデータの最小単位のことです。本記事では単語や句読点などと思ってもらって支障はありません。例えば、文章 "Hello, world!" をトークン化すると、["Hello", ",", "world", "!"] というように分割されます。

では、本題に参ります。今回のアップデートにより、以下の点が改善されました。

1. **機能呼び出し:** チャットボットが外部ツールを使ったり、人間の言葉をコンピュータが理解できる命令に変えたりできるようにする新機能です。例えば、ボストンの今の天気は? と尋ねると、weatherapi という機能呼び出し、ボストンの今の天気を答えてくれます。
2. **改良された"steerability":** steerability とは、モデルの振る舞いを人間が理解しやすいように制御したり、特定の目標に向けて誘導する能力を指します。OpenAI の新しい gpt-3.5-turbo-0613 モデルは、system message と呼ばれ

る特殊な指示文により、確実な制御、誘導を可能にします。system message の例として、“あなたは私の家庭教師です。すぐに答えを教えず、段階的に指導してください”という指示があります。これにより、学習者に寄り添う段階的な回答をしてくれるようになります。今回のアップデートで、以前よりも家庭教師としての役割を的確にこなしてくれるようになった、ということです。

3. **大幅なコスト削減:** OpenAI は大幅なコスト削減も行いました。特に、人気のある gpt-3.5-turbo の入力トークンの価格が 25%削減され、開発者は 1,000 入力トークンあたり約\$0.0015、1,000 出力トークンあたり\$0.002 でこのモデルを使用できるようになりました。さらに、“text-embedding-ada-002”というエンベディングモデル（単語やフレーズを高次元のベクトル空間に埋め込むモデル）のコストも 75%削減されました。
4. **コンテキストウィンドウの拡大:** OpenAI は新たに 16,000 トークンのコンテキストウィンドウを持つ gpt-3.5-turbo-16k を導入しました。これにより、プロンプトが最大 16,000 トークンの長さまで扱うことが可能になり、標準の 4,000 トークン版の 4 倍のコンテキスト長さを扱うことができます。gpt-3.5-turbo-16k は一度のリクエストで約 20 ページのテキストを処理でき、大量のテキストを処理し生成する必要がある開発者にとって大きなメリットとなります。
5. **旧モデルの非推奨化:** OpenAI は定期的にモデルを更新し続けており、旧モデルは永遠に利用可能なわけではありません。今回のアップデートで、gpt-3.5-turbo-0301 や gpt-4-0314 などの早期のモデルの非推奨化が始まり、開発者は 2023 年 9 月 13 日までこれらのモデルを利用することができます。

これらのアップデートにより、GPT モデルの機能強化とコスト削減を通じて、開発者がより効率的に、またコスト効率よく AI アプリケーションを構築できるようになりました。特に今回は gpt-3.5-turbo-16k の登場により扱えるトークン数が急増した点が注目を集めています。

[ChatGPT ビジネスレポート \(picture academy\)](#)

