

MIT Technology Review

Published by KADOKAWA / ASCII

OpenAI

世界最高峰の AI 集団が目指す世界



CONTENTS

- 001 オープン AI、文章から画像を描く「DALL-E2」を 100 万人に提供
- 006 「馬に乗った宇宙飛行士」——文章から画像を描く AI が驚きの進歩
- 013 GPT-3 が「最高」で「最悪」な AI である理由
- 018 オープン AI、互いに訓練することで賢くなるバーチャルロボを開発
- 020 ベールに包まれた世界最高峰の頭脳集団「オープン AI」の正体

2015年12月11日、サンフランシスコを拠点とする人工知能 (AI) を研究する非営利団体として産声を上げた「オープン AI (OpenAI)」。大規模言語モデルの「GPT-3」、画像生成プログラム「DALL-E (ダリー)」といった、世間を驚かせる AI テクノロジーを発表してきた。オープン AI の最終目標は「汎用人工知能 (AGI)」だが、開発には膨大な資金がかかる。そのことが、率直性や透明性の後退、純粋な非営利構造の解消、秘密保持やイメージの保護、従業員に忠誠心を求めるといった変化の原因となった。オープン AI の功績と、そのベールに包まれた実態をレポートする。



Open AI

オープン AI、文章から画像を描く「DALL-E2」を 100万人に提供

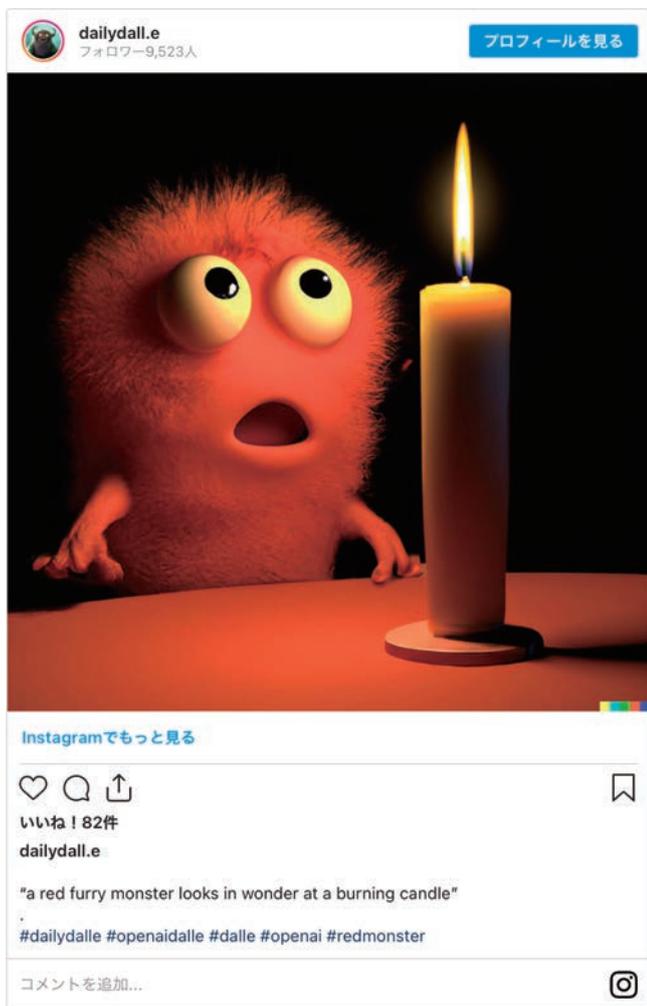
オープン AI は、文章から画像を生成する「DALL-E 2」のベータ版を 100 万人に提供する。一部有料化し、生成した画像の商用利用も可能となる。

サンフランシスコに拠点を置く AI 企業のオープン AI (OpenAI) は、画像生成プログラム「DALL-E (ダラー) 2」のベータ版について、提供待ちリストに登録済みの 100 万人へ提供する。拡大に伴い、一部有料化に踏み切る。

2022 年 4 月に招待制で提供が始まって以来、およそ 10 万人が DALL-E 2 を試用している。そして今、オープン AI は、DALL-E 2 を有料サービス化するにあたり、その 10 倍の人数に門戸を

開く。

オープン AI の製品・パートナーシップ担当副社長のピーター・ウェリンダーは、「(DALL-E 2 は) GPT-3 のときよりもはるかに大きな、私たちが予想した以上の関心を集めています」と話す。有料ユーザーは DALL-E 2 で生成した画像を、児童書の挿絵、ゲームや映画のコンセプト・アート、マーケティング用のパンフレットといった商業プロジェクトで使えるようになる。しかし、今

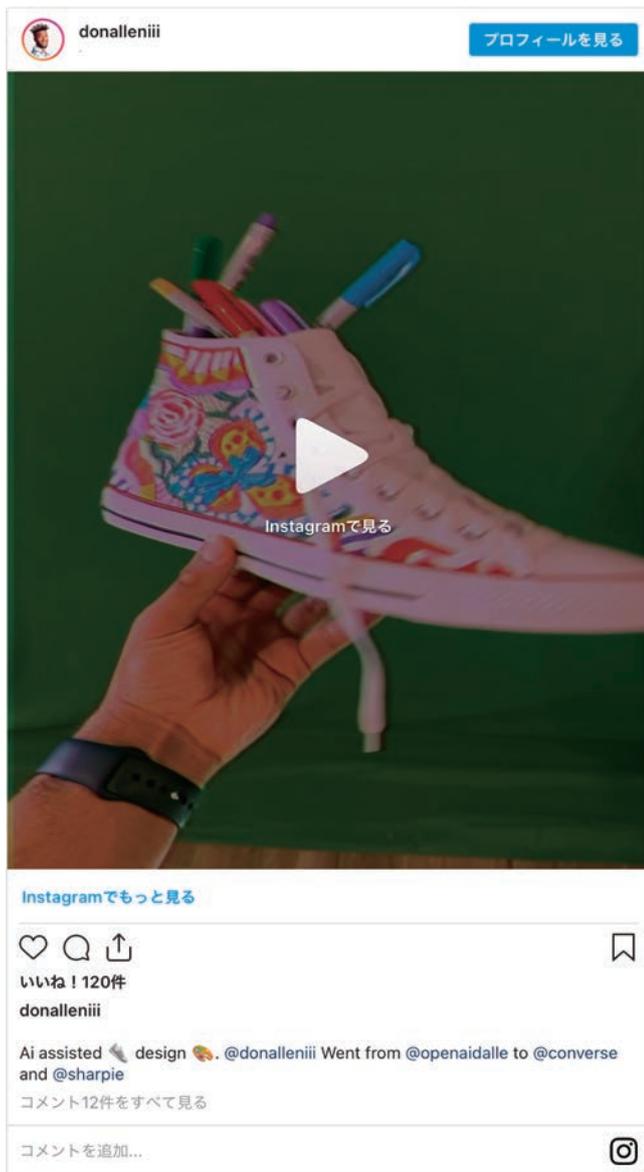


回の製品発売は、オープン AI が強力な人工知能 (AI) を展開するにあたって採用してきた、段階的に顧客へリリースして問題が発生したら対処するというアプローチに対して、過去最大の試練をもたらすだろう。

DALL-E 2 の料金はそれほど高くはない。DALL-E 2 に文章を入力すると、1 回あたり 4 枚の生成された画像を入手できる。この料金が「1 クレジット」だ。ユーザーは最初の月に 50 ク

レジット、その後は月に 15 クレジットが無料で付与される。ユーザーは一般的に、1 つの文章で数十枚の画像を生成し、最高の出来の画像だけを取っておく傾向があるので、パワー・ユーザーならすぐにその割り当てを使い果たしてしまうかもしれない。クレジットが足りなくなれば、15 ドルで 115 クレジットを追加購入できる (つまり 150 ドルで 460 枚の画像を入手できる)。

今回の発売に先立ち、オープン AI はアーリー・



アダプターと協力して、ツールの問題解決に取り組んできた。第一波のユーザーは、かわいい動物のマッシュアップや、実在する写真家の作風を不気味なほど正確に模倣した写真、レストランやスニーカーのデザインのコンセプトを視覚化するムード・ボードなど、シュールで印象的な画像を

続々と生み出してきた。これにより、オープン AI は DALL-E 2 の長所と短所を探ることができた。同社でプロダクト・マネージャーを務めるジョアン・ジャンは「非常に多くの素晴らしいフィードバックをいただいています」と言う。

すでにオープン AI は、ユーザーが作成できる画像の種類を制御する手段を講じている。例えば、有名人が描かれた画像は生成できない。今回の商用化に向け、オープン AI は初期のユーザーから指摘されたもう 1 つの深刻な問題に対処した。4 月に公開された DALL-E 2 のバージョンは、最高経営責任者（CEO）や消防士が白人男性ばかりだったり、教師や看護師が白人女性ばかりだったり、明らかにジェンダーや人種にバイアスがある画像がたびたび生成されていた。

7 月 18 日、オープン AI は修正版を発表した。ユーザーが DALL-E 2 に人の集団を含む画像を生成するように求めると、DALL-E 2 はオープン AI が世界的な多様性をより広く表していると主張するサンプルのデータセットを使うようにした。オープン AI が独自に実施した調査では、ユー

ザーが DALL-E 2 の出力に多様なバックグラウンドを持つ人々が含まれていると回答する傾向が 12 倍になったそうだ。

これは必要な修正だが、表面的なものだ。オープン AI は、ユーザーが求めることができるものをフィルタリングしたり、基盤となるモデルが生成するものを検閲したりして、ユーザーが指摘する多くの問題に対処している。しかし、モデル自体の問題や、訓練データの問題を修正しているわけではない。根本的な対処ではなく表向きなものだから、オープン AI は迅速に修正できるのだ。ただ、一部の人のにとっては、これは単なる応急処置でしかない。

ロンドン・スクール・オブ・エコノミクスのジュディ・ワイクマン教授は「アルゴリズムにおける社会的バイアスは非常に大きな問題です」と言う。彼女はアラン・チューリング研究所でデータ科学と AI におけるジェンダーの問題についても研究している。「多くのエネルギーが技術的な修正に注がれており、そのような努力のすべては賞賛に値します。しかし、問題に対する長期的な

解決策ではありません」。

それでもオープン AI は、DALL-E 2 でジェンダーと人種のバイアスに対処したことが、本格的な公開に踏み切る自信につながったと述べている。しかし、これが最終結論ではない。AI におけるバイアスは悪質かつ解決が難しい問題であり、同社は新しい事例が発生するたびにモグラ叩きのように修正を続けなければならないだろう。オープン AI は、製品に調整が必要な時は、いつでも提供を一時停止すると話している。

ウェリンダー副社長は、バランスをとることが大切だと話す。調整によって、ユーザーが生成しようとする画像が予期しない方向に抑制される可能性がある。例えば、オープン AI がジェンダー・バイアスの修正を初めてリリースしたとき、一部のユーザーから「女性のスーパーマリオが多すぎる」という苦情があった。このようなケースは、事前に予測することが難しいとウェリンダー副社長は言う。「人々がどのような画像を生成しようとしているのかを確認することで、調整や調節ができるのです」。

しかし、100万人以上のユーザーが生成した数億枚もの画像を監視することは、膨大な作業になる。ウェリンダー副社長は、人間のモデレーターの必要人数についての質問は巧みにかわしたが、社内のスタッフで賄う予定だとは話した。オープンAIは、人間の判断と自動検査を組み合わせたハイブリッド・アプローチによるモデレーションを採用している。ウェリンダー副社長によれば、チームの構成はモデレーターを増やしたり、人間と機械の介入のバランスを調整したりすることで、必要に応じた変更ができるという。

グーグルは2022年5月、独自の画像生成AI「イメージェン (Imagen)」を発表した。オープンAIとは異なり、グーグルはこのテクノロジーに関する計画をほとんど明らかにしていない。「イメージェンについて発表できる新しい情報は、まだありません」とグーグルのブライアン・ガブリエル広報担当は言う。

オープンAIは、2015年に設立された当時、汎用AIに対する信念と、そのテクノロジーの実現によって人類全体に利益をもたらすという使命

を持つ、純粋な研究所を自称していた。しかしこの数年、同社は製品企業へと方向転換し、その強力なAIを有料顧客に提供している。

しかし、そうした変遷はすべて同じビジョンの一部だとウェリンダー副社長は話す。「オープンAIのテクノロジーを製品として大規模に展開することは、当社のミッションにおける欠かせない要素の1つです。ハードルが低いうちに、AIというテクノロジーに関する有用性と安全性の確認を、早いうちから繰り返すことが重要だからです」。(Will Douglas Heaven) **T**

**eムックは、MITテクノロジーレビュー
有料会員限定サービスです。**

**有料会員はすべてのページ（残り45ページ）を
ダウンロードできます。**

ご購入はこちら



<https://www.technologyreview.jp/insider/pricing/>

No part of this issue may be produced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of KADOKAWA ASCII Research Laboratories, Inc.

本書のいかなる部分も、法令または利用規約に定めのある場合あるいは株式会社 角川アスキー総合研究所 の書面による許可がある場合を除いて、電子的、光学的、機械的処理によって、あるいは口述記録の形態によっても、製品にしたり、公衆向けか個人用かに関わらず送信したり複製したりすることはできません。