

AI 特許紹介(69)  
AI 特許を学ぶ！究める！  
～DALL-E2 画像生成特許～

2024 年 10 月 10 日  
河野特許事務所  
所長弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

## 1.概要

特許権者 OpenAI

出願日 2023 年 3 月 30 日

登録日 2024 年 3 月 5 日

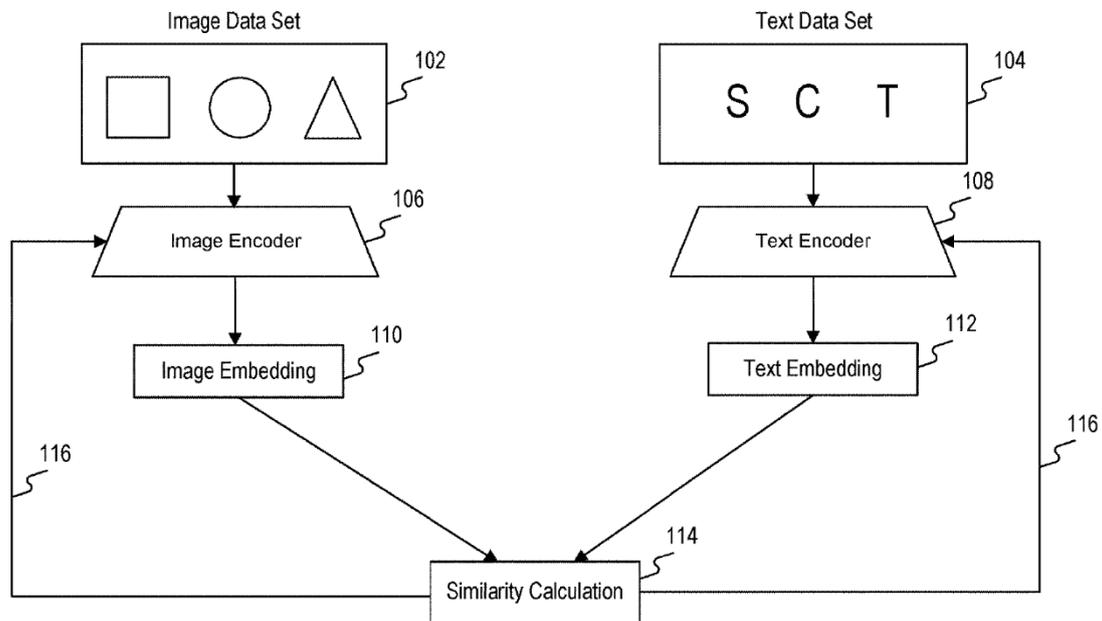
登録番号 US11922550

発明の名称 階層的テキスト条件付き画像生成のためのシステムおよび方法

550 特許は、DALL-E2 のベースとなった技術であり、テキストを、対照的モデルである CLIP (Contrastive Language-Image Pre-training)事前モデルに与えることにより、CLIP 画像埋め込みを生成し、当該画像埋め込みを 2 段階目のデコーダに与えることにより画像を生成する技術に関する。

## 2.特許内容の説明

図 1A は、モデルをトレーニングするための機能ブロック図である。



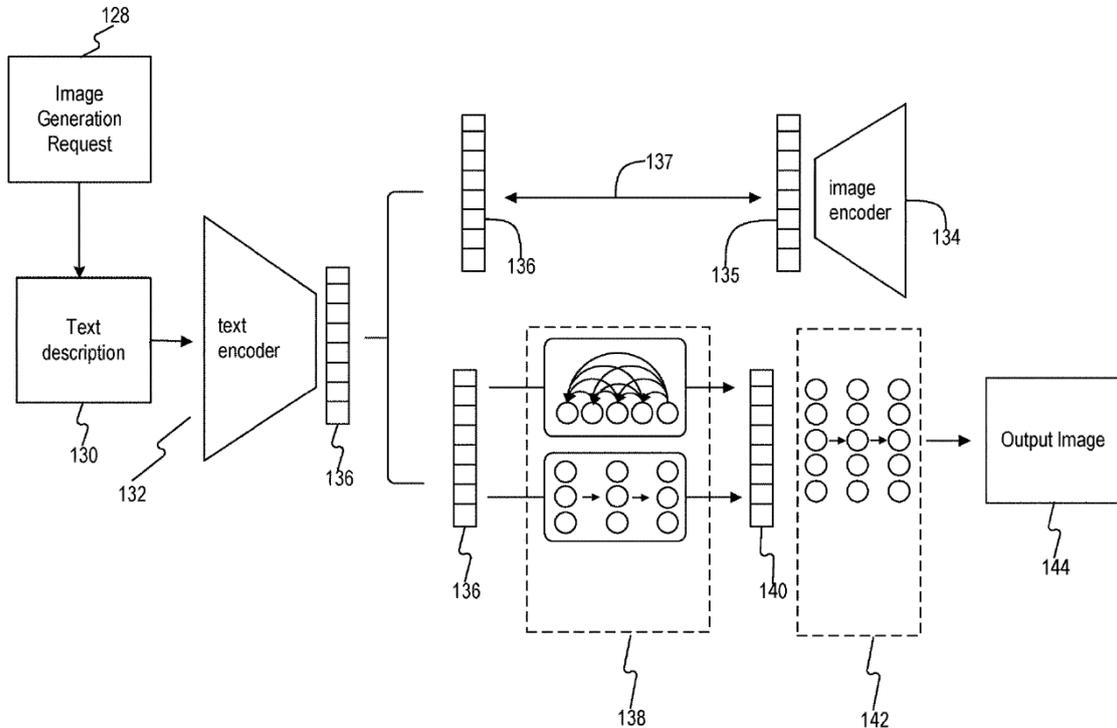
データベースから第1データセットおよび第2データセットを収集する。第1データセットは、画像データセット102を含み、第2データセットは、テキストデータセット104を含む。テキストデータセット104は、画像データセット102に対応する。たとえば、テキストデータセット104は、画像データセット102内の画像に対応するキャプションである。画像データセット102は、画像エンコーダ106への入力であり、テキストデータセット104は、テキストエンコーダ108への入力である。

それぞれの入力に基づいて、画像エンコーダ106とテキストエンコーダ108は共同でトレーニングされる。画像エンコーダ106は画像埋め込み110を生成し、テキストエンコーダ108はテキスト埋め込み112を生成する。エンコーダおよび埋め込みのトレーニングには、コサイン類似度演算等の類似性計算114の計算が用いられる。例えば、コサイン類似度は、画像埋め込み110とテキスト埋め込み112内のペアワイズ要素またはコンポーネント間で計算される。共同トレーニングには、画像エンコーダ106とテキストエンコーダ108の更新116が含まれ、これにより、画像埋め込み110とテキスト埋め込み112が更新される。

エンコーダは、正しく一致した画像とテキストのペア間のコサイン類似度を増大させ、不正確に一致した画像とテキストのペア間のコサイン類似度を減少させるように、反復的に更新される。たとえば、テキストデータセット104内の対応するテキスト記述が、画像データセット102内の対応する画像と正しく一致している場合、画像埋め込み110とテキスト埋め込み112のそれぞれの埋め込み間のコサイン類似度が高くなる。トレーニングが完了すると、画像埋め込み110およびテキスト埋め込み112を受信し、そ

の後、例えばデバイスまたは他の機械学習モデルがこれらにアクセス可能になる。

図 1B は、テキスト入力に対応する画像を生成するための機能図を示す。



画像生成モデルのトレーニングには、画像エンコーダ 134 とテキストエンコーダ 132 のトレーニングが含まれる。結合表現 137 は、テキスト埋め込み 136 と画像埋め込み 135 の潜在空間表現である。トレーニングが完了すると、テキストエンコーダ 132 と画像エンコーダ 134 が固定され、機械学習モデルの一部として使用される。

画像生成要求 128 は、入出力デバイスへの要求またはクエリで構成される。テキスト記述に対応する画像を生成するには、テキスト記述 130 にアクセスし、テキスト記述 130 をテキストエンコーダ 132 に入力する。テキストエンコーダ 132 は、テキスト埋め込み 136 を生成する。これが第 1 のサブモデル 138 への入力になる。第 1 のサブモデル 138 は、対応する画像埋め込み 140 を生成する。対応する画像埋め込み 140 は、第 2 のサブモデル 142 への入力となる。第 2 のサブモデル 142 は、拡散モデルであり、出力画像 144 を生成する。

### 3. クレーム

550 特許のクレーム 1 は以下の通りである。

1. システムにおいて、

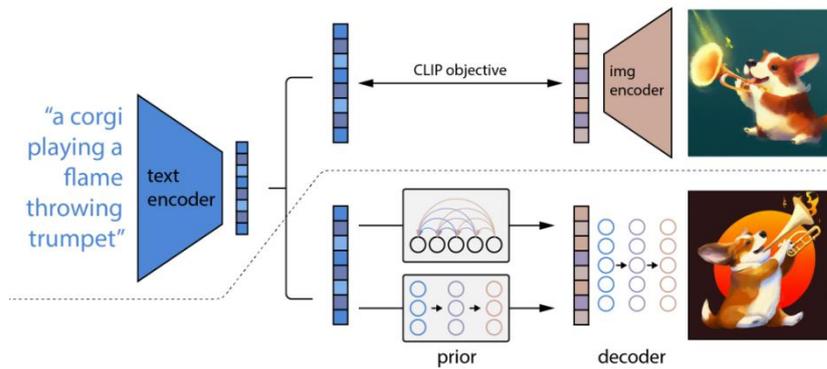
命令を格納する少なくとも 1 つのメモリと、  
テキスト入力に対応する画像を生成するための操作を実行するための命令を実行するように構成された少なくとも 1 つのプロセッサであって、その操作は以下を含む：  
テキスト記述にアクセスし、そのテキスト記述をテキストエンコーダに入力し、  
テキストエンコーダからテキスト埋め込みを受信し、  
テキスト記述またはテキスト埋め込みの少なくとも 1 つを、テキスト記述またはテキスト埋め込みの少なくとも 1 つに基づいて、対応する画像埋め込みを生成するように構成された第 1 のサブモデルに入力し、  
第 1 のサブモデルによって生成されたテキスト記述または対応する画像埋め込みの少なくとも 1 つを、テキスト記述または対応する画像埋め込みの少なくとも 1 つに基づいて出力画像を生成するように構成された第 2 のサブモデルに入力し、  
第 2 のサブモデルには、出力画像を生成する前にアップサンプリングするように構成された第 1 のアップサンブラモデルと第 2 のアップサンブラモデルが含まれており、第 2 のアップサンブラモデルはブラインドスーパー解像度 (BSR: blind super resolution) 劣化によって破損した画像でトレーニングされており、第 2 のサブモデルは第 1 のサブモデルと異なり、  
第 2 のサブモデルによって生成された出力画像をデバイスがアクセスできるようにし、  
デバイスは、出力画像を使用して画像生成モデルをトレーニングするように構成されているか、または画像生成要求に関連付けられている。

#### 4. 本特許に関連する論文

本特許に関する論文 “Hierarchical Text-Conditional Image Generation with CLIP Latents”<sup>1</sup>が、OpenAI の Aditya Ramesh 氏らにより公表されている。

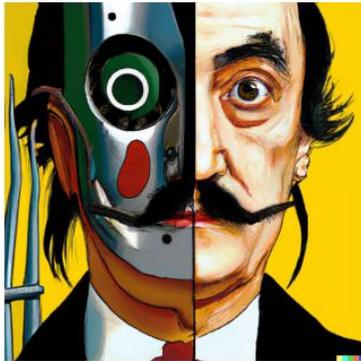
---

<sup>1</sup> Aditya Ramesh, et al. “Hierarchical Text-Conditional Image Generation with CLIP Latents” arXiv:2204.06125v1 [cs.CV] 13 Apr 2022



上記図は、unCLIP の概要を示す説明図である。点線の上側は、テキストと画像の結合表現空間を学習する CLIP トレーニングプロセスを示している。点線の下側は、テキストから画像への生成プロセスを示している。本特許で説明したように、CLIP テキスト埋め込みは、拡散事前処理(prior)に入力され、画像埋め込みが生成される。次にこの画像埋め込みは拡散デコーダに入力され、最終画像が生成される。

本モデルによる画像生成の例は以下に示すとおりである。



vibrant portrait painting of Salvador Dalí with a robotic half face



a shiba inu wearing a beret and black turtleneck



a close up of a handpalm with leaves growing from it



an espresso machine that makes coffee from human souls, artstation



panda mad scientist mixing sparkling chemicals, artstation



a corgi's head depicted as an explosion of a nebula



a dolphin in an astronaut suit on saturn, artstation



a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese



a teddy bear on a skateboard in times square

以上

## 著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI研究所 AI コース、生成 AI ビジネスコース修了。

[AI 特許コンサルティング](#)、[医療 AI 特許コンサルティング](#)の他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門 3](#)」、「[ブロックチェーン 3.0](#)(共著)」がある。