

文章生成 AI 技術の 活用事例と感性への挑戦

電気通信大学 副学長 坂本 真樹



要 約

生成 AI で第 4 次 AI ブームともいわれる昨今であるが、本稿では、2017 年に特許出願をし、当時メディアでも話題になった画像から文章を生成する技術について、この技術が生まれた背景とその後の展開について紹介する。深層学習が発表された 2012 年ごろ、人が音楽を聴いているときに情景を思い浮かべるといふ現象に着目し、その情景を写真やイラストとした画像の中でも、感性との結びつきの強い色彩に注目し、そこから楽曲を提案するシステムを開発した。その際、情景は歌詞に描かれることが多いのではないかと考え、歌詞を解析対象とした。その後、画像から歌詞を介して楽曲を検索するだけでなく、画像から歌詞を生成するという技術を開発した。2017 年に、生成した歌詞をアイドルグループが歌ったことで話題になった。2022 年 4 月からは AI を活用して作詞して歌う VTuber によるプロジェクトを開始したが、そのプロジェクトによる感性への挑戦についても触れる。

目次

1. はじめに
2. 2017 年に特許出願した文章生成 AI 技術とその活用事例
 2. 1 色彩から楽曲を提案するシステム
 - (1) 歌詞中に出現する色彩と結びつきのある単語の抽出
 - (2) 単語の色彩ベクトル
 - (3) 色彩から楽曲を提案するシステム
 2. 2 画像から文章を生成する技術
 - (1) 2017 年当時の文章生成 AI システムの構成
 - (2) カギとなるプリミティブワードについて
 - (3) 文章生成手法
 2. 3 アイドルグループの描く絵から歌詞を生成
3. VTuber を用いた作詞 AI 活動による感性への挑戦
4. まとめ

1. はじめに

昨今、生成 AI の進化で、第 4 次 AI ブーム到来とも言われている。生成系の AI の進化のきっかけになったのは、2014 年に発表された GAN (Generative Adversarial Network (GAN と呼ばれる敵対的生成ネットワーク)) であろう。当時、モントリオール大学の博士課程に在籍していた Ian Goodfellow らが発案したアルゴリズムである。深層学習 (Deep Learning) では、大量の教師データが必要であるが、GAN はその学習データを自ら作り出す「教師なし学習」を可能にするものである。「Generator (生成するネットワーク)」と「Discriminator (真偽を判定するネットワーク)」が競い合うことで相互に学習し、質の高い画像の生成が可能になった。

GAN を用いることで、アメリカの半導体大手企業の Nvidia の研究チームが 2018 年に発表した「StyleGAN」で生成された人物写真は実在の人物と区別が付かないほど高精度で、話題となった。

2021 年 1 月には、テキストを入力とすることで、その内容にもとづいた画像を生成する「DALL・E」がアメリ

カの AI 開発企業 Open AI によって発表された。OpenAI は、2020 年 7 月に大規模言語モデル「GPT-3」を発表（GPT は登録商標）し、「Transformer」という深層学習モデルを用いることで、それまで困難とされていた自然な文章を生成する道が開けてきた。

2022 年 5 月には、Google Research の Brain Team が画像生成 AI「Imagen」を発表、6 月には、アメリカの AI 開発企業 Midjourney が画像生成 AI「Midjourney」を公開した。さらに 8 月には、イギリスの AI 開発企業 Stability AI が「Stable Diffusion」を公開した。

このような画像系生成 AI の基盤ともなっている GPT-3 を応用することで、OpenAI は 2022 年 11 月に文章生成 AI を対話型にした ChatGPT を誕生させ、広く社会で活用されるようになっていく。

ChatGPT に代表される対話型文章生成 AI では、入力されたテキストデータの構造や文法、ある種の意味を理解し、それに基づいて新しい文章を生成することができる。翻訳や要約をしたり、広告コピーを生成したり、記事や小説を書くなど、文章を扱う様々な業界での作業の効率化を図ることができるのではないかと期待されている。

今や高精度になった文章生成 AI であるが、本稿では、筆者が 2017 年に特許出願をし、当時メディアでも話題になった画像から文章を生成する技術について、この技術が生まれた背景とその後の展開について紹介する。現代の文章生成 AI に比べると、精度という観点でははるかに劣るものの、深層学習が発表された 2012 年ごろ、早い段階で、非常にユニークな着眼点のもと、文章生成に取り組み始めた。

人が音楽を聴いているときに情景を思い浮かべるといふ現象に着目し、その情景を写真やイラストとした画像の中でも、感性との結びつきの強い色彩に注目し、そこから楽曲を検索するシステムを開発した。この技術については、2.1 節で紹介する。その際、情景は歌詞に描かれることが多いのではないかと考え、歌詞を解析対象としたが、その後、画像から歌詞を介して楽曲を検索するだけでなく、画像から歌詞を生成するという技術を開発した。画像から文章を生成する技術については 2.2 節で紹介する。さらに、2017 年に、生成した歌詞をアイドルグループが歌ったことで話題になったことについては、2.3 節で紹介する。そして 3 章では、2022 年 4 月からは AI を活用して作詞して歌う VTuber によるプロジェクトを開始したことについて紹介し、そのプロジェクトによる感性への挑戦についても述べる。

2. 2017 年に出願した文章生成 AI 技術とその活用事例

生成 AI の話題でもちきりの現在から 6 年前に、筆者の研究室では、文章を生成する AI 技術に関する特許出願を行い、社会実装も行っていった。本節では、その技術が誕生した背景を含め、解説を行う。

2.1 色彩から楽曲を提案するシステム

文章を生成する AI 技術を生み出すさらに以前に、色彩入力による楽曲検索システムを開発し、関連する発明について 2 件の特許出願し、いずれも権利化されている⁽¹⁾⁽²⁾。本稿では、この技術に関する論文発表内容⁽³⁾⁽⁴⁾に基づき紹介する。

この技術を開発した 2011 年当時は、情報検索需要が高まり始めていたころであり、多様化する情報検索要求の 1 つとして楽曲検索があった。2000 年ごろから音楽配信サービスが実験的に開始され、その後現在に至るまで成長してきた。楽曲検索技術に関しては、当時から様々な研究が行われてきた。歌手の名前や曲名で検索する方法や、楽曲に関する物理的特徴（音響信号のスペクトル特徴量）を直接的にクエリとして用いる手法も多かったが、われわれは、音楽は本質的に感性に作用するものであるという考えのもと、客観的な情報のみで検索を行うのではなく、感性情報を重視して楽曲を検索することを目指した。感性に基づいた楽曲検索に関する研究としては、楽曲に含まれる音の高さ、強さ、長さの時間的推移を用いて印象ベクトルを生成しておき、予め用意された印象尺度対に対する評価値をユーザが入力して楽曲を検索する手法や、SD 法による心理実験の結果を因子分析することで感性空間を設定し、ユーザの入力した感性語対の評定尺度に対する感性空間上の座標と登録楽曲の感性空間上の距離によって検索を行うシステムなどがあった。

筆者の研究室では、楽曲検索システムについて取り組み始めるさらに 5 年ほど前に、音楽と色彩の結びつきに関

する共感覚現象に注目していた。共感覚とは、ある感覚刺激に対して別の感覚が活性化する現象である。特に、特定の音を聴くと特定の色が見える共感覚現象は色聴と呼ばれる。色聴に着目した研究は従来から盛んに行われている。例えば、色聴を持つとされる特定の作曲家に焦点を当てた研究や、色聴保持者に対して様々な音刺激の印象に近い色彩を選択させる実験を行う研究などがあった。絶対音感を持つ被験者と相対音感を持つ被験者の反応に焦点を当て、それぞれの被験者に関する音と色彩の反応の一貫性を長期間に亘って調査した研究もあった。しかし、色聴は特殊な現象であり、音楽を聴く人の多くは、色聴能力を持っていない場合の方が一般的である。そこで筆者らは、色聴能力を持たず音楽や色彩に関する特別な訓練を受けていない被験者を対象として、楽曲と色彩の結びつきに関する認知的特性を心理実験により調査した⁽⁶⁾。その結果、楽曲の試聴から直接的に色彩が見えていないにも関わらず、楽曲を試聴した際に想起される色彩には特定の偏りがあることを発見した。筆者らは、被験者が楽曲を試聴した際に想起した色彩に偏りが発生した要因に関して、2つのプロセスの可能性を検討した。一つ目の可能性は、楽曲を聴いて何らかの感性的な因子が想起され、その感性的な因子に対応する色彩が想起されるというプロセスである。例えば、楽曲を聴いて、「自然な」という感性語が想起され、その感性語に対応した色彩（例えば「青」）が想起されるというプロセスである。二つ目の可能性は、楽曲の試聴から何らかの具体的な場面が想起され、その場面对応する色彩が想起されるというプロセスである。例えば、楽曲を聴いて「浜辺」が想起され、その場面对応した色彩として「青」が想起されるというプロセスである。一般の人について、楽曲を聴いて何らかの色がイメージされるとすると、どのような認知過程によるのかを調査するために、楽曲の試聴からの自由記述回答によって、楽曲の特徴抽出を行った。この自由記述回答の分析の結果、色聴能力を持たず音楽や色彩に関する特別な訓練を受けていない一般的な被験者においては、音楽を聴くと、何らかの場面が想起され、その場面の色がイメージされるとしている。すなわち、一般的な被験者においては、色聴保持者のように楽曲の試聴と色彩を直接的に対応させるのではなく、楽曲の試聴からの具体的な場面想起を介することで色彩を想起するという認知過程が見られるということである。

その後、筆者らは、歌詞付きの楽曲の試聴から想起される色彩とその楽曲の歌詞のみの提示から想起される色彩の比較を行った⁽³⁾。この研究では、楽曲試聴時における歌詞の内容の理解が具体的な場面の想起に影響を与えているとすれば、少なくとも歌詞付きの楽曲の試聴から想起される色彩とその楽曲の歌詞のみの提示から想起される色彩の間に相関があるという仮説を立て、両者を比較している。その結果、両者の間には有意な相関があることがわかり、歌詞を用いて、色彩をクエリとする楽曲検索システムの開発へとつながった。

筆者の研究室で当時提案した歌詞の利用に基づく色彩をクエリとする具体的な楽曲検索手法⁽⁴⁾の概要は次のとおりである。(1) 歌詞中に出現する単語の中から色彩と結びつきのある単語を抽出する。(2) 抽出されたそれぞれの単語から想起される色彩印象を合成し、この合成された色彩印象を、楽曲の試聴から想起される色彩印象とみなす。(3) クエリとして入力された色彩と各楽曲の歌詞から推定された色彩印象を比較し、クエリとの類似度が高い色彩印象を持つ楽曲を順に提示する。以下では、この手順について、要点を概説する。

(1) 歌詞中に出現する色彩と結びつきのある単語の抽出

心理実験によって、歌詞に用いられやすく、色彩が想起されやすい単語を調査した。音楽検索サイトを参考に、できるだけアーティストや曲調が異なるように配慮しながら80曲を選曲した。大学生60名を30名ずつの2グループに分け、1曲あたり30名の被験者からの回答が得られるように割り当てた。実験では、楽曲のタイトルやアーティストに関する情報のない、歌詞のテキストのみを印刷した紙の冊子を被験者に配布した。被験者には、各ページの歌詞を読み、色彩想起に影響を受けた歌詞中の箇所を丸で囲んでもらった。この実験で、所定の確率で被験者が色彩を想起するものとして選んだ単語を、色彩と結びつきやすい単語とし、プリミティブワードとした。当初の実験で選ばれたプリミティブワードは283語であったが、被験者実験でアノテーションを行ったプリミティブワードが多いほど精度があがるため、後述のようにその後増やしている。

(2) 単語の色彩ベクトル

次に、選定されたプリミティブワードと各色彩が結びつく確率を色彩ベクトルとし、色彩ベクトルをまた別の心理実験により求めた。先述の被験者とは異なる 21 名の大学生および大学院生に、プリミティブワードと所定の数のカラーサンプルを PC ディスプレイにフルスクリーンで提示した。被験者に対して、プリミティブワードの提示から想起した色彩を最大 3 色回答するよう求めることで、プリミティブワードの色彩ベクトルを計算することとした。

被験者実験では、色彩と結びつきのある単語を全て網羅することはできない。そこで、プリミティブワード以外にも、楽曲の色彩ベクトルを求めるために有用な可能性がある色彩と結びつきのある単語（未知語）を特定し、それらの単語に関する情報を推定できるようにした。未知語の色彩ベクトルは、基本的に、プリミティブワードと未知語の類似度に基づいて推定することとした。プリミティブワードと未知語の類似度に関しては、潜在的意味解析 (LSA) を用いて求めた。LSA に用いるコーパスには毎日新聞コーパス (2005 年) を用いた。ただし、このコーパスには、歌詞に用いられやすい単語として選んだプリミティブワードの全てが含まれているわけではなかったため、先述の心理実験で使用した 80 曲の歌詞もコーパスに含めた。コーパスの使用目的は、LSA によって単語同士の一般的な意味の類似度を得ることであった。LSA では相応の単語量のコーパスが必要であるため、当時筆者らが行った研究では、新聞コーパスを使用した。LSA では、1 段落を 1 文書として扱った結果、使用した文書数は 643,807、異なり語数は 129,462 となった。圧縮次元数に関しては、当時パフォーマンスが良いと報告されていた 300 次元とした。今の大規模言語モデルから考えると非常に小規模に感じる。

(3) 色彩から楽曲を提案するシステム

構築したシステムでは、ユーザは、システムに対して入力する色彩数を 1 色から 35 色（その後 45 色に増やした）の範囲で自由に設定することができるようにした。なお、色彩に変換できればよいため、色付きの写真やイラストなどの画像も扱えるものである。開発当初のシステムでは、設定した色彩数に応じてカラーパレットから任意の色彩を選択し、スライダを用いて選択した色彩の配色割合を指定する仕様である。システムは、ユーザが入力した色彩の組み合わせ、および、その配色割合を基に、入力された色彩印象と登録されている楽曲の色彩印象の類似度を算出し、類似度の高い色彩印象を持つ楽曲をリスト化してユーザに提案する。図 1 は当時実装したシステムに色彩を入力した場合の画面例であり、図 2 はその結果提案された楽曲例である。

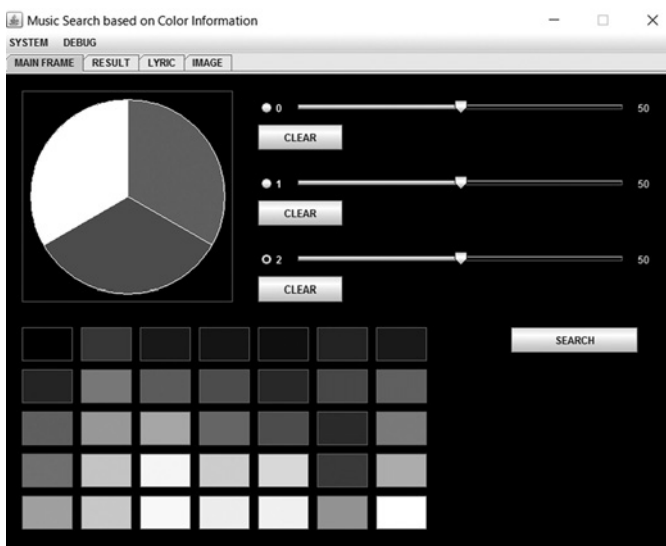


図 1 システムに入力した色彩例

順位	ID	タイトル	アーティスト	類似度
1	20	Happy Xmas	伊藤かな恵, 阿部佳奈, 加藤宗+絵	0.7112971165611682
2	40	DAYDREAM	JUDY AND MARY	0.6479750195810261
3	43	海陸	吉田拓郎	0.6273773046331236
4	61	永遠	ZARD	0.6228138111164168
5	124	Point of No Return	CHEMISTRY	0.6208874797681336
6	66	ひまわり	TUBE	0.6079683754052478
7	78	雨に唄く花	川崎あけい	0.6064243545329505
8	65	世界のほとりの片隅から	ZONE	0.601040383403464
9	56	RETURN TO THE MOON	CONSTRUCTION NINE	0.59944037968761
10	106	クリスマスキャロルのために	極楽湯一	0.5939393216415784
11	90	夏空グラフィティ	しきものがかり	0.5771037704414707
12	96	チェリー	スピッツ	0.5764227014358075
13	48	生まれたての匂	KOKIA	0.5752377299685142
14	91	DEPARTURES	globe	0.5731455975247444
15	7	あなたにサラダ	Dreams Come True	0.5621697442235708
16	55	1985年 Factory Street 夏	福山雅治	0.557065876467943
17	81	シンデレラ・クリスマス	KinKi Kids	0.5564340416774558
18	76	ひまわり	香月まゆか	0.5545512576562895
19	99	winter again	GLAY	0.553274602912032
20	47	白い雪もんな雪が私には好き	TULIP	0.5504742509583037
21	44	あはれ	くず	0.5488120743278971
22	50	雪が降る日に	かくや塘	0.5484957429199453
23	30	あなたを感じていたい	ZARD	0.5472987518333724
24	26	さらば青春の光	布袋寅泰	0.5454814114254113
25	3	やめないで PURE	KinKi Kids	0.5448744827550343
26	85	スマイル	BUMP OF CHICKEN	0.5440315704882329
27	111	白は白く雪の夜	吉田拓郎	0.5435094632489602
28	19	snow drop	堂本崇一	0.5418405218157939
29	83	雪の華	中島美嘉	0.5409658918773562
30	49	不完全熱唱	野原謙	0.5377410021198794
31	41	No limit	Every Little Thing	0.5355852587638404
32	57	福熱	TUBE	0.5346466926479582
33	60	純色の青春	野原謙	0.5310930239890441

図 2 提案される楽曲例

2. 2 画像から文章を生成する技術

(1) 2017 年当時の文章生成 AI システムの構成

2011 年と 2012 年に論文発表した技術（特許 5344756 及び 5618150）は、画像やイラストからイメージに合った歌詞を解析することで楽曲を提案するというものであったが、その後 AI 技術が進化したこともあり、画像やイラストから歌詞などの文章を生成することが可能になった。そこで、後述するアイドルグループとの出会いをきっかけに、アイドルが描いたイラストから歌詞を生成するという取り組みを開始した。技術としては、筆者の研究室の特許技術（特許第 7170299、出願日 2017 年 12 月 3 日、登録日 2022 年 11 月 4 日）⁽⁶⁾を用いてシステムの設計を行い、その後研究発表も行っている⁽⁷⁾。実装したシステムのインターフェースは、図 3 に示されるように、ユーザインターフェースモジュール、画像解析及び文章生成モジュール・データベースによって構成される。図 3 のように、画像から詩的な文章が数秒で生成されるものである。



図 3 システムのインターフェースと出力結果例

ユーザインターフェースモジュールは、ユーザが入力した画像を受け取り、システム内部に渡すとともに、解析部から得られた解析結果を受け取り、ユーザへの提示を行う。出力結果は、画像を解析することにより得られた文章として提示されるユーザがより直感的に使いやすく操作できるように、図 3 に例示されるようなグラフィカルユーザインターフェースを採用した。

画像解析部では、ユーザインターフェースモジュールからの入力を受け取り、画像の色彩情報の解析及び画像内の物体認識を行う。詳細は後述するが、画像特徴から、「単語－色彩データベース」及び「物体認識の学習データ」を参照し、単語リストを生成し、「文章生成手法の学習データ」に基づき文章を生成する。

図 3 のシステムでは、画像の色彩情報の解析で、画像の色彩情報を 45 次元のベクトル空間へ変換する。入力画像が持つ色彩ベクトルと類似した色彩ベクトルを持つ単語を単語－色彩データベースから抽出し、その単語を元に歌詞を生成するためである。なお、単語と色彩印象が紐づいたデータベースの構築手順は 2.1 節で説明した楽曲検索システムと同様である。単語－色彩データベースは、単語データベース、単語の色彩ベクトルデータベースの 2 つで構成され、色彩想起に強く影響をあたえる単語（2.1 節で紹介した「プリミティブワード」）用とプリミティブワード以外の単語用の 2 種類ずつあるため、計 4 つから構成される。

(2) カギとなるプリミティブワードについて

2.1 節で紹介した楽曲検索システム実装から 5 年経過し、随時更新しているが、歌詞でよく使われる色彩想起との結びつきが強い単語としてプリミティブワードを選定している。プリミティブワードは、歌詞でよく使われ、かつ汎用性のある単語がよい。そこで、一般的に知られた楽曲を収集対象とした。1968 年から現在までのオリコン年間 CD シングルランキングの 1 位～100 位に含まれる歌詞を収集した。歌詞の収集は「Lyrics Master 2.4.8.1」(<http://www.kenichimaehashi.com/lyricsmaster/>)を使用した。収集した歌詞のうち、歌詞を持たない楽曲と英語のみで構成される楽曲は除いた 4458 曲をプリミティブワード選定のための歌詞コーパスとして使用した。収集

した楽曲を単語に分割するために形態素解析を行い、28918 語の単語を取得した。形態素解析されたコーパスから、品詞が「名詞」「動詞」「形容詞」「形容動詞」かつ出現回数が 50 回以上の単語を抽出対象とし、プリミティブワードとして 710 語を選定した。次に、2.1 節で解説した手法により、各プリミティブワードの色彩ベクトルを算出し、プリミティブワード以外の色彩と結びつきのある単語についても色彩ベクトルを推定するようにした。最終的に、10 万 9259 曲の日本語歌詞コーパスに対して形態素解析し、全 1501 万 4087 語、異なり語数 13 万 1565 語の単語が得られた。さらに 13 万 1565 語のうち 2 つ以上の歌詞に含まれていた単語のみを抽出し、最終的に 6 万 7434 語と、10 万 9259 曲をそれぞれ行、列とした単語文脈行列について潜在意味解析を行い、歌詞コーパスの意味空間を構築した。次に、構築した単語文脈行列に対して特異値分解を行い、パフォーマンスが良いと報告されている 300 次元に圧縮した。意味ベクトル間のコサイン類似度をとることで、単語-単語間、単語-歌詞間、歌詞-歌詞間の意味的類似度を測定することができるため、この類似度を用いて未知語データベースを構築した。未知語の色彩ベクトルの推定は、未知語とプリミティブワードの類似度をもとに行い、新たに未知語 2 万 9226 語に色彩ベクトルが付与された。

入力画像に基づいた単語リストは、入力画像が持つ色彩ベクトルと、単語-色彩データベース内にある各単語が持つ色彩ベクトルのコサイン類似度を算出することで求められる。

さらに、このようにして色彩により抽出された単語群に加えて、物体認識により単語名を追加した。物体認識については、当時の Tensorflow のプログラムを利用した。画像を入力すると、その画像内に含まれる物体名が確信度に基づいて出力されるため、確信度が最も高い物体名を、文章生成に用いる入力単語の一つとして抽出した。また、出力されるラベルデータは英語であるが、単語-色彩データベースや文章生成のための機械学習では日本語の歌詞データを用いるため、ラベルデータを日本語翻訳した。

(3) 文章生成手法

文章生成部では、画像解析部によって抽出された単語リストをもとに、文章を生成する。当時の文章生成手法は、「文字単位の LSTM」「単語単位の LSTM」「単語 2-gram のマルコフ連鎖」の 3 パターンのうち、どれか 1 つを選択して用いることとした。詳細は割愛するが、「文字単位の LSTM」「単語単位の LSTM」学習モデルに入力する文章として、単語リスト内に含まれる単語をランダムに連結したものをを用いている。両モデルではデータベース内にある学習済みのデータを用いて、入力文章の次に来る文字または単語を確率的に選択して入力文章に追加する。その後、この操作を連鎖的に繰り返して文章内の総文字数または総単語数が任意の数に達するまで生成を行う。「単語 2-gram のマルコフ連鎖」の入力として、単語リスト内の単語からランダムに選択されたもののうち、その単語が先頭に含まれる 2-gram 表現を用いた。入力となる 2-gram 表現の次に来る単語を確率的に選択し、入力の 2-gram 表現の 2 個目の単語と選択された単語を合成し、新たな 2-gram 表現を生成する。新たに生成された 2-gram 表現を入力として、再度単語の予測を繰り返し行うことで文章の生成を行う。なお、マルコフ連鎖部では画像印象に近い単語を生成するため、単語予測の際に、画像から抽出された単語リスト内の単語が候補としてある場合にはそれを優先し、次につながる単語の候補として選択する。その後、3 パターンの文章生成手法のうち、いずれかの手法で生成された文章をユーザーインターフェースに渡す。

以上の手法は、2017 年特許出願当時の手法であるが、その後の文章生成 AI 手法の進化はすさまじいことは 1 章で述べたとおりである。大規模データの規模や生成 AI 実装コストの規模については現在主流のものと比較にならないが、単語の連鎖の予測の考え方についての本質は同様であると考えている。

2. 3 アイドルグループの描く絵から歌詞を生成

2017 年 4 月 6 日、東京秋葉原の仮面女子カフェで、人工知能で作詞した「電☆アドベンチャー」の記者発表イベントが開催された。新聞各紙の記者が集まり、終了後は囲み取材もあった。翌日以降、新聞各紙に掲載された。産経新聞「アイドルソング AI が作詞 「仮面女子」と電通大コラボ」(産経新聞)、「ええっ! ? AI がアイドル楽曲作詞♪」(スポーツニッポン)、「AI 作詞曲」(日刊スポーツ)、「作詞 アイドル×AI」(朝日新聞) など多種多

様な新聞に掲載された。その後、NHK や民放の各番組の取材も相次いだ。

このように反響が大きかったのは、硬い国立大学で開発された人工知能とアイドルとのコラボレーションで、実際にアイドルに歌ってもらったことによると思われる。そもそもこのコラボレーションはどのように実現したのか、経緯について説明する。詳細は坂本（2019）も参照されたい⁽⁸⁾。

2016年8月に東京ビッグサイトで開催されたイノベーションジャパン 2016 に、研究室のオノマトペ（擬音語・擬態語の総称）関連のシステムを出展していた。筆者の研究室では、任意のオノマトペを五感や感性に関する多数の尺度で数値化できる技術を開発しており、関連する特許が多数ある。そこに、仕事で会場を訪れていた仮面女子のメンバーの月野もあさんが、研究室のブースに立ち寄ってくれた。「つきのもあ」と入力すると名前前の音の印象を多次元尺度で数値化するとともに、イメージからも出力できる iPad に実装したシステムなどを体験していただいた。当時、現在も筆者が所属している芸能事務所との契約の話が進んでいたこともあり、何かコラボしましょう、という話になった。何をしたら面白いかを考えた結果、仮面女子さんは歌う仕事をしており、筆者は人工知能要素技術を使って言語の感性情報処理をしていることから、人工知能で作詞して歌ってもらおうと面白いのではないかと、という結論に至った。2.2 節で解説した特許技術を活用し、仮面女子のメンバーに色付きの絵を描いていただき、それを元に作詞する、ということに挑戦することにした。実際に仮面女子のメンバーのみなさんからいただいた色付きのイラストから作詞した。実際の楽曲と歌詞の全文は、こちらを参照されたい。<https://youtu.be/UpfzVJgSD8U>

当時、文章生成については、様々な研究が行われていた。人工知能による小説創作に取り組むプロジェクトとして「きまぐれ人工知能プロジェクト 作家ですよ」と「人狼知能プロジェクト」がメディアで注目されていた。しかし、小説のようなストーリー性と理解可能性が重要な文章を生成するのは難しいようであった。それに対し、歌詞の場合、詩的な表現を用いることが出来るため、違和感のない文章を作成しやすいという利点があった。そのような状況であったことを考えると、すでに小説も生成可能になっている現在の ChatGPT を代表とする文章生成 AI の進化は驚くべきものである。

3. VTuber を用いた作詞 AI 活動による感性への挑戦

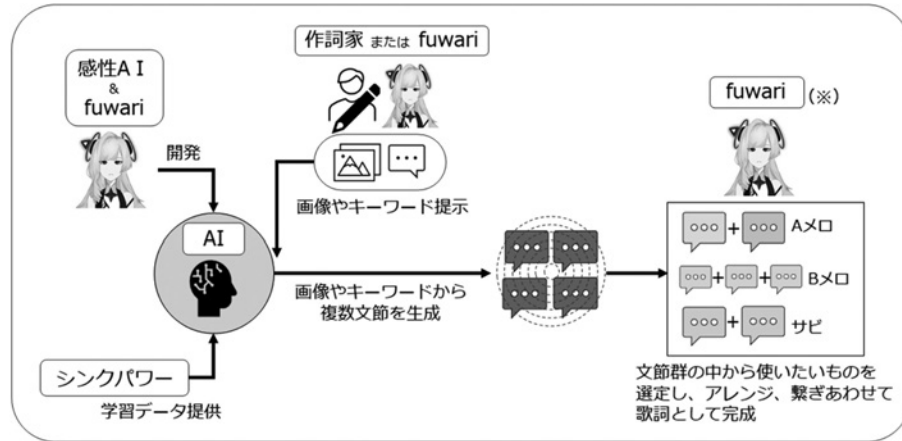
アイドルグループとのプロジェクトは一度だけだったが、一過性のプロジェクトでは意義が薄いと考え、継続的に作詞 AI による歌詞で歌ってくれるアーティストを探していた。コンピュータが何かを生成するのは難しいことではないが、その精度が重要である。特に、作詞の場合は、ただ意味が通じる、というだけでなく、人々の感性に刺さることが重要であり、やはりヒットする曲を出せるか、が挑戦なのではないかと考えた。

AI 作詞曲を歌ってくれるプロのアーティストが見つからないまま5年が経過した。その間に、時代はメタバースに注目が集まり、VTuber が YouTube などでの活動の幅を広げるようになっていた。

そこで、2022年3月に AI 作詞家 VTuber の fuwari を創出し、fuwari が文章生成 AI で作詞を行い、歌手として活動するようプロデュースした。この活動を開始するまでも大変な道のりがあり、特に、AI 作詞というものが世の中に受け入れられていないことから、AI による作詞が著作権などの観点で問題が生じないかなど、さまざまヒアリングを行いながら進めた。内閣府や文化庁の現段階での検討状況を見た限りでは、AI だけで自律的に生成されたものは著作物性はなく、人が思想感情を創作的に表現するための「道具」として AI を使用した場合は著作物性がある、と言えるようである（https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoseido/r05_01/pdf/93918801_03.pdf）。しかし、AI の進化がすさまじいこともあり、継続的な検討が必要とのことである。

fuwari プロジェクトでは、AI により生成された文章をもとに、楽曲に合うように VTuber の fuwari が創作活動を行う、というものであることに加え、学習に用いる歌詞も、著作権処理がされたものを用いる、という慎重な体制をとっている。さらに、少なくとも現段階では、歌詞の著作権も放棄する形で音楽活動を行っている。しかし、VTuber というものについての一般の理解が広まっていないせいか、fuwari そのものも AI である、という誤解があるようである。VTuber とは、「Virtual YouTuber」であり、「2D または 3D のアバターを使って活動して

いる YouTuber」であるため、人間の活動といえる。歌っている声も機械によるものという誤解があるが、少なくとも fuwari については人間が歌っている。つまり、fuwari プロジェクトは、図 4 に示されるように、人間の作詞家に相当する fuwari が、ツールとして AI を用いて作詞し、歌っている。電気通信大学の文章生成 AI 技術のライセンスを受け、株式会社シンクパワーから提供を受けた歌詞データを学習データとし、電気通信大学発ベンチャーの感性 AI 株式会社が文章生成 AI を構築し、その AI をツールとして fuwari が作詞をして歌っている、というプロジェクトである。



(※) …歌詞のまとめ上げは作詞家側で実施するケースもあり

図 4 AI 作詞家 VTuber の取り組み概要 (2022 年 3 月 30 日 株式会社シンクパワープレスリリースより <https://syncpower.jp/ja/news/2022/0330/1>)

2022 年 4 月 27 日、ポニーキャニオンが運営する PR 型デジタルディストリビューションサービス EarlyReflection (<https://earlyreflection.com/>) で fuwari が作詞を手掛け楽曲の配信が開始した。第 1 弾シングルは「夜舞う fuwari」という曲で、2023 年 8 月 10 日現在、YouTube では 3200 回再生されている。以下は fuwari の第 1 弾シングル歌詞である。

夜舞う fuwari

作曲：みゅーと

作詞：fuwari

編曲：みゅーと・結・今給黎 奏人

fuwari ふるえる世界を 綴るメロディー

遅い夜溺れる fuwari 舞う 二人 僕が僕を 演じても君には どんな日も 過去を描きながら
暗闇 で手をかざし届くまで 僕が消えてしまいそうな 雨君の深い影 静かに

あなたと泳ぐ 同化 踊ろう かけがえのない世界は 咲き誇る 夢の先へ

fuwari 鮮やかに 変わって 色づく 世界の向こうへ
轟け (ひびけ) 交わり合う傷は 止められない 強く
暗いトンネルを 吹き消して
出会った 場所が変わっても
変わらないものを なぞり 確かめたい

遅い夜溺れる 後悔 咲き 花は消える 君はずっと 強く抱きしめてくれて 奪いたい 現実と

夢が終わらせているよ 寂しさたちと 見えない oh lonely lonely night 空は星が 降りそそぐ

ねえ 奇跡を呼ぶ 街の光が ああ 二人は今叫ぶよ oh your love 夜にのって

fuwari 高く飛べる世界の果て 乗り越えて見上げ

あと 少し星を繋ぎ 描きかけた夜

遅い夜 溺れる闇に ふるえるように 見上げる

果てないけど どこへ向かう 消えない この愛

fuwari もっと 君だけの愛を 知りたい風がそよげば

fuwari ふるえる世界を 綴るメロディー

AIが使われていることを感じさせられるところはあるだろうか。

Turing Test 的には、人の作詞か、AIによる作詞かわからないところまで行っていれば、ある程度成功とみなせるのかもしれないが、やはり、この歌詞が、人の心を打つレベルまで行くことが目標である。

その後も、第2弾シングル「熱くキラリ光る夜空に駆け出す」、第3弾シングル「FlickFlickKiss」、第4弾シングル「FUWALUTION」とリリースを続け、本稿が出版されるまでには、第5弾シングルもリリースされている予定である。

図5は fuwari channel という YouTube チャンネルの画面である。登録者が非常に少なく、運営継続自体厳しいチャンネルではあるが、人の感性に刺さるまでがんばりたいと思っているため、よろしければチャンネル登録をして応援していただければ幸いである。



図5 fuwari channel の YouTube 画面 (https://www.youtube.com/channel/UCq5f7id9SbIL_LgL57HbdQ 2023 年 8 月 10 日アクセス)

4. まとめ

本稿では、筆者の研究室での文章生成 AI 関連の取り組みについて、10 年前からさかのぼって紹介した。今や、一研究室では太刀打ちできない大規模言語モデルに基づく文章生成 AI の開発競争が世界的に展開されている。しかし、感性を切り口に、人を対象とした丁寧な実験により取得されるデータの価値を今後も追及していきたいと思う。一般的に理解できるレベルでの一見よさそうな文章が出てくる、というだけでなく、人の感性に響く文章を作れる AI を目指すことは、歌詞の生成において重要である。AI による作詞がヒット作を生み出せるようになると、人間の作詞家の仕事が奪われるのではないか、といった懸念の声がきかれる。人間対 AI という対立構造で語られることが多いが、人間の作詞家も様々いて、たくさんの個性のあつまりであるように、その中に、AI が参加してもよいのではないかと思う。ただし、AI だけで自律的に、時に大量に生成された歌詞は、創作意図を持たないものとして、創作物として認められない、という考え方ももっともである。作曲するためのソフトウェアはツールとして普及していると同様に、作詞するためのソフトウェアである作詞 AI も一つのツールとして気軽に使ってもらえるようになるとよいと思う。人間が、作詞する AI をツールとして用いただけで、その作詞は著作物として認められないとなると、誰も使えない技術になるだろう。作詞してみたいがなかなかよい表現が思いつかない、という人も AI の支援で作詞を楽しみながら、自分の感性に刺さる表現を発見したりできるようになれば良いと思う。人間社会の発展に資するべく、人間の作詞家の創作活動の支援になる AI 技術の開発を行っていきたい。

(参考文献)

- (1) 坂本真樹・仲村哲明・内海彰、情報処理装置、情報処理方法、及びプログラム（プリミティブな色彩連想語を用いた色彩に基づく楽曲推薦システム）、特許第 5618150（出願日：2011 年 1 月 19 日、登録日：2014 年 9 月 26 日）
- (2) 坂本真樹・仲村哲明、情報処理装置、情報処理方法、及びプログラム（色彩と歌詞の相関に基づく音楽検索システム）、特許第 5344756 号（出願日：2009 年 7 月 17 日、登録日：2013 年 8 月 23 日）
- (3) 仲村哲明、川西鉦平、坂本真樹：歌詞と色彩に基づいた楽曲推薦の可能性、電子情報通信学会論文誌、J94-A (2)、85-94 (2011)
- (4) 仲村哲明、内海彰、坂本真樹：色彩想起と歌詞の関係に基づく楽曲検索、人工知能学会論文誌、27 (3)、163-175 (2012)
- (5) 坂本真樹・鎌田清一、音楽と色彩の連想メカニズムに関する研究、日本認知科学会第 24 回大会発表論文集、pp.184-187 (2007)
- (6) 坂本真樹、情報処理システム、情報処理方法及びプログラム、特許第 7170299（出願日：2017 年 12 月 3 日、登録日：2022 年 11 月 4 日）
- (7) Masato Konno, Kohei Suzuki, Maki Sakamoto, Sentence Generation System Using Affective Image, Proceedings of the 2018 Joint 10th International Conference SCIS and 19th ISIS, 678-682.
- (8) 坂本真樹、作詞 AI による人の創造力と想像力増幅の試み—電☆アドベンチャー作詞の経緯と今後の可能性—、電子情報通信学会誌、102 (3)、234-239 (2019.3.1)

(原稿受領 2023.8.12)