

画像生成 AI を用いた読者を誘引する書籍表紙画像の生成

池田 諒真¹ Shan Junjie¹ 安尾 萌² 西原 陽子¹
Ryoma Ikeda¹ Junjie Shan¹ Megumi Yasuo² Yoko Nishihara¹

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² 立命館グローバル・イノベーション研究機構

² Ritsumeikan Global Innovation Research Organization

Abstract: 書籍の表紙は、読者に与える第一印象を形成する。既存研究では、書籍表紙のデザインは読者の興味や選択に影響を与えることが明らかになった。しかし、表紙画像に書籍のタイトルやカテゴリがどのように反映されれば、読者を誘引する表紙画像となるかはまだ不明である。本研究では、テキストから画像を生成する AI モデルを活用し、書籍の説明文から表紙画像を生成するシステムを構築した。このシステムを用い、カテゴリ別の書籍に対して、異なるプロンプト構成によって生成された表紙画像が読者を誘引するかについて主観実験を行い評価した。実験の結果、書籍の説明文にある名詞を全て画像生成 AI のプロンプトとした表紙画像が最も読者を誘引した。

1 はじめに

商品パッケージのデザインは、消費者に与える第一印象を形成し、その購買意欲に大きな影響を及ぼす [Bloch 95, Silayoi 07]。書籍の表紙もまた「書籍商品のパッケージ」として、読者に第一印象を与え、手に取るかどうかの判断を左右する重要な要素である [Hagtvedt 08, Jian 19]。既存の研究では、表紙に描かれたポジティブな感情を喚起する写真やイラストが、結果として購買意欲を高める可能性が示唆されている [Liu 17]。しかし、表紙画像の具体的なデザイン属性と、読者の興味との関連性については、十分な分析が行われていないのが現状である [Orth 08]。さらに、これまでの書籍表紙デザインは専門デザイナーの感覚や経験に大きく依存してきたため、統一的な基準による定量的・定性的な分析は困難であった [Bruce 97]。

本研究では、画像生成 AI を用いて書籍の説明文から表紙画像を自動生成し、どのような属性をプロンプトの構成とすると書籍表紙の誘引力を高める表紙画像を生成できるかを明らかにすることを目的とする。近年注目されている入力テキストから画像を生成する技術 [Ramesh 21, Rombach 22] を用いて、書籍の説明文から表紙画像を生成するシステムを構築した。本研究で達成した主な点は以下の通りである。

- 画像生成 AI を利用し、書籍の説明文から表紙画像を生成するシステムを構築した。
- 書籍の説明文にある単語の品詞と重要度に基づき、画像生成 AI に入力するプロンプトの内容を

構成する 4 つの手法を提案し、それぞれ表紙画像を生成した。

- 主観実験を通じて、4 つの手法を用いて生成された、18 種類の書籍カテゴリ別の表紙画像の誘引力を評価し、カテゴリ別の各提案手法で生成した表紙画像が、どれだけ読者を誘引するかを分析した。

2 関連研究

2.1 認知に関する既存研究

Coherence Principle [Mayer 05] は、視覚情報とテキスト情報が意味的に一致しているほど、学習者の認知的な負荷が下がり、内容の理解が促進され、結果として好意的な評価につながるとしている。この原則は、情報の整合性が人間の認知プロセスにおいて重要な役割を果たすことを示唆している。

2.2 心理学に関する既存研究

Processing Fluency Theory [Reber 04] は、人間は知覚的に処理しやすい、つまりスムーズに理解できる情報を好む傾向があることを示している。これは、「分かりやすいものは心地よい」という直感的な感覚を理論的に説明するものである。

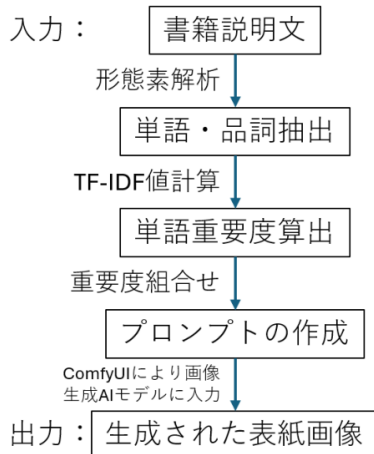


図 1: 提案システムのフロー図

2.3 書籍表紙・商品パッケージに関する既存研究

商品パッケージのデザインが消費者の購買意欲に与える影響については、Bloch ら [Bloch 95] や Silayoi ら [Silayoi 07] をはじめ、多くの研究が行われてきた。書籍の表紙も同様に「パッケージ」として、読者の選択に影響を与える重要な要素である [Hagtvedt 08]。Gudinavičius ら [Gudinavičius 17] は、読者の選択における表紙の役割を分析し、その重要性を指摘している。特にオンライン書店の普及により、読者は中身を読む前に表紙画像だけで購入を判断する機会が増えており [Jiang 13], Park ら [Park 23] もオンライン環境での表紙の役割について分析している。しかし、表紙画像の具体的なデザイン属性と、読者の興味との関連性については、十分な分析が行われていないのが現状である。

3 書籍表紙画像の生成手法

本研究では、書籍の説明文から読者を誘引する表紙画像の生成を目的とするため、画像生成 AI に入力するプロンプトの作成により表紙画像の生成を調整している。図 1 に示す提案システムのフロー図に基づき、入力された「書籍説明文」に対し、形態素解析と TF-IDF 値の計算を行うことで、プロンプトを作成するための「単語重要度」を算出する。次に、抽出された単語や重要度の組合せにより、プロンプトの作成手法を提案する。作成したプロンプトを画像生成 AI モデルに入力し、表紙画像を生成する。本章では、これらの手順について詳細に述べる。

3.1 説明文の単語抽出

本研究では、画像生成 AI への指示となるプロンプトを作成するため、形態素解析により書籍の説明文から単語と品詞の抽出を行う。形態素解析エンジンには MeCab¹ を使用し、新しい単語に対応するため追加辞書 NEologd² を用いる。画像生成 AI のプロンプトには名詞で構成されており、また書籍の説明文は内容を客観的に伝えるために主観的な記述（形容詞、副詞など）を避けて名詞中心の構成となる傾向があることを考慮する。本研究では形態素解析により抽出された単語の中から名詞をプロンプトの要素として採用した。抽出された名詞のうち、プロンプトとして意味内容が希薄となる代名詞、非自立、数、接尾、副詞可能、助動詞語幹、接続詞的といった単語を除外し、データのクリーニングを行う。

3.2 単語重要度の算出

プロンプトに使用する単語の重要度を決めるため、形態素解析後の各単語に対して TF-IDF 値を計算する。これにより、各書籍の内容を特徴づける重要な単語が数値化される。IDF 値の作成には、最新版の日本語 Wikipedia の全記事データ³ を基に構築されたテキストコーパスを使用した。IDF 値について、ユニーク単語数が 3,725,459 (総単語数が 720,205,244)、総文書数 (記事数) が 2,374,911 となっている。表 1 に、書籍説明文の入力と抽出された名詞のサンプル例を示す。

3.3 プロンプトの生成

画像生成 AI において、プロンプトは大きく分けて「Positive」プロンプトと「Negative」プロンプトの 2 種類があり、それぞれが生成される画像の内容を制御する上で重要な役割を果たす。Positive プロンプトは「生成したい要素」や「画像に含めたい属性」を AI に指示するものであり、一方 Negative プロンプトは「生成したくない要素」や「除外したい属性」（例えば、奇形になりやすい文字、不完全・歪みな画像など）を指示している。本研究では、説明文から抽出された各単語と重要度の組み合わせにより、以下の 4 種類のプロンプト作成方法を提案する。本研究では、3.1 で抽出した名詞を英訳した上で、書籍の内容を反映させるための 4 つの手法を比較した。

- 手法 1 (全名詞 Positive) : 書籍説明文から抽出された名詞を全て平等 (重み 1.0) に Positive の

¹<https://taku910.github.io/mecab/>

²<https://github.com/neologd/mecab-ipadic-neologd>

³<https://dumps.wikimedia.org/jawiki/latest/jawiki-latest-pages-articles.xml.bz2>

表 1: 形態素解析による単語抽出の具体例

説明文入力	
DNA や生体情報の収集、顔や声を常時スキャンする街頭センサー、移動・購買履歴を自動で吸い上げる端末。すべてのデータは中央 AI に集約され、アルゴリズムが個人ごとに「危険度スコア」を算出する仕組みになっていた。	
抽出単語 単語 (上位 20 件)	TF-IDF 値
吸い上げる	9.7826
危険度	8.4344
スキャン	7.5788
購買	7.3223
街頭	7.3178
履歴	6.6423
アルゴリズム	6.6275
生体	6.6052
センサー	6.6024
AI	6.5507
算出	6.4946
DNA	6.3542
常時	6.3176
端末	6.1619
スコア	6.0428
集約	6.0317
仕組み	5.5566
自動	5.2906
収集	5.2541
顔	4.6522

内容としてプロンプトを作成する。これは、書籍説明文にある単語は、書籍の内容を表現するには全て重要であると考えられるためである。

- **手法 2 (Positive/Negative 分割)** : 書籍説明文から抽出された名詞を重要度 (TF-IDF 値) により並び替え、重要度の上位半分を Positive プロンプト、重要度の下位半分を Negative プロンプトとしてプロンプトを作成する。各部分に入力する単語の重みは全て 1.0 にする。これは、書籍説明文にある単語は「独自性のある単語」と「一般的な単語」が含まれ、独自性のある単語を強調させて、一般的な単語を避けるために考えられた。
- **手法 3 (上位 10 件)** : 書籍説明文から抽出された名詞を重要度 (TF-IDF 値) により並び替え、重要度上位 10 件の名詞のみを Positive プロンプトとしてプロンプトを作成する。なお、入力する各単語の重みは全て同じ (重み 1.0) である。これは、手法 2 の考え方と同じく、書籍の内容に繋がる単語を強調する。
- **手法 4 (重み付け)** : 書籍説明文から抽出された名詞を重要度 (TF-IDF 値) により並び替え、重要度上位 2 件の名詞に 1.15 倍、3 位から 10 位の

名詞に 1.1 倍、残る単語に 1.0 倍の重みを付けて、Positive プロンプトとして入力してプロンプトを作成する。この処理は、書籍説明文にある単語が書籍の内容を表現する上で重要なものであると考えられるために行うものである。重み付けの基準については、画像生成 AI の重み付けに関する検証記事 4 を参考に重みを決定し、Positive プロンプトとして入力する。

主流なオープンソースの画像生成 AI モデルは、英語の入力しか受け入れられないので、各手法により抽出された単語を英訳した上で、画像生成モデルに入力する。その上、全提案手法において、表紙画像の生成を指示する「(cover image:1.2), key visual」の Positive のプロンプトと、奇形・劣化の画像を抑制する「text, title, watermark, cropped, out of frame」の Negative のプロンプトを入力している。本モデルにおいて最も高品質かつ安定した画像生成が期待できる標準解像度であることから、生成される画像の解像度は 1024 × 1024 とした。

表 2 は各手法により作成されたプロンプトの構成例を示している。本研究では、ComfyUI⁴と Stable Diffusion 3.5⁵モデルを利用して表紙画像の生成を行う。図 2 に各手法の入力プロンプトから生成された表紙画像の例を示す。

4 評価実験

4.1 実験方法と評価指標

生成した表紙画像が読者の興味をどの程度誘引するかを評価するため、主観実験により評価した。書籍の購入決定には、価格、著者、レビュー、販売環境など様々な外的要因が複雑に絡み合う。そのため本研究では、これらの外的要因を捨象し、表紙デザインそのものが持つ、読者の「興味を惹き、手に取らせる力 (誘引力)」に焦点を当てることとした。この指標を、本稿では便宜上「魅力度」と呼ぶこととする。ただし、評価実験の実施にあたっては、この「魅力度」という概念を被験者により具体的に理解してもらうため、指示文や評価基準の説明において「書籍の表紙としてどの程度購買意欲が湧くか」という言葉を用いた。評価実験で使用した UI を図 3 と図 4 に示す。図 3 は 4 つの表紙画像の内 1 つを選択した際の、「書籍タイトルと生成された表紙画像」に対する 5 段階評価の画面である。図 4 は書籍タイトルと各手法により生成された表紙画像を表示しており (左から順に手法 1, 手法 2, 手

⁴<https://github.com/comfyanonymous/ComfyUI>

⁵<https://stability.ai/stable-diffusion-3>



図 2: 表 2 に示した各手法で作成したプロンプトにより生成された表紙画像例（左から手法 1, 2, 3, 4）

表 2: 各手法におけるプロンプト構成例

手法	Positive プロンプト	Negative プロンプト
手法 1	(cover image:1.2), key visual, Monitoring, data, City, Risk level, Vivid, cage, scan, Near Future, Tech, Buy, street, back side, history, algorithm, living organisms, sensor, AI, calculation, technology, transparent, DNA, Terminal, Score, Aggregation, structure, daily, future, automatic, collection, Major, escape, net, digital, girl, shaft, huge, the same, Device, face, city, residents, story, Domination, society, voice, central, move, system, personal, people, figure, enterprise, information, development, record	text, title, watermark, cropped, out of frame
手法 2	(cover image:1.2), key visual, Monitoring, data, City, Risk level, Vivid, cage, scan, Near Future, Tech, Buy, street, back side, history, algorithm, living organisms, sensor, AI, calculation, technology, transparent, DNA, Terminal, Score, Aggregation, structure, daily, future, automatic	text, title, watermark, cropped, out of frame, collection, Major, escape, net, digital, girl, shaft, huge, the same, Device, face, city, residents, story, Domination, society, voice, central, move, system, personal, people, figure, enterprise, information, development, record
手法 3	(cover image:1.2), key visual, Monitoring, data, City, Risk level, Vivid, cage, scan, Near Future, Tech, Buy	text, title, watermark, cropped, out of frame
手法 4	(cover image:1.2), key visual, (Monitoring:1.15), (City:1.1), (Risk level:1.1), (Vivid:1.1), (cage:1.1), (scan:1.1), (Near Future:1.1), (Tech:1.1), (Buy:1.1), street, back side, history, algorithm, living organisms, sensor, AI, calculation, technology, transparent, DNA, Terminal, Score, Aggregation, structure, daily, future, automatic, collection, Major, escape, net, digital, girl, shaft, huge, the same, Device, face, city, residents, story, Domination, society, voice, central, move, system, personal, people, figure, enterprise, information, development, record	text, title, watermark, cropped, out of frame



図 3: 「書籍タイトルと生成された表紙画像のペア」に対する主観的購買意欲の評価画面の 1 例

法 3, 手法 4), 5 段階評価後の最も「購買意欲が湧く」画像を選択する画面である。

実験は図 3 および図 4 の UI を用いて行われ、被験者には以下の手順で評価を依頼した。

1. 基本情報（性別、年齢など）を入力させる。
2. ランダムに 1 つの書籍タイトルが選択され、4 つの手法により生成された表紙画像を表示する。
3. 被験者に、図 3 に示すように、表示された「書籍タイトルと生成された表紙画像のペア」に対して、他の画像と比較せず、「書籍の表紙としてどの程度購買意欲が湧くか」という基準で 5 段階評価する（絶対評価）。
4. 被験者に、図 4 に示すように、4 つの画像の中から、「最も購買意欲が湧く」と感じた画像を 1 つ選択させる（相対評価）。

被験者には、評価に正解はなく、自身の感覚に基づいて正直に回答するよう教示した。

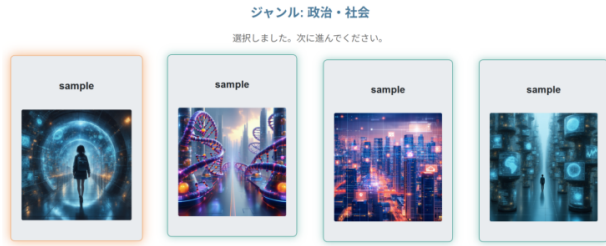


図 4: 書籍タイトルと各手法により生成された画像の表示画面 (左から手法 1, 手法 2, 手法 3, 手法 4 の順で書籍タイトルと生成された表紙画像のペア)

表 3: 実験に用いた書籍カテゴリとタイトル数

カテゴリ	タイトル数
アート・建築・デザイン	10
コミックス	10
サイエンス・テクノロジー	10
スポーツ・アウトドア	10
ノンフィクション	10
ビジネス・経済	10
音楽	10
絵本・児童書	10
教育・自己啓発	10
芸能・エンターテインメント	10
事典・年鑑・本・ことば	10
趣味・実用	10
人文・思想・宗教	10
政治・社会	10
文芸・評論	10
暮らし・健康・料理	10
旅行・紀行	10
歴史・地理	10
合計	180

4.2 実験データ

本実験では、表 3 に示す 18 カテゴリから各 10 タイトル、計 180 タイトルの書籍データを使用した。実験参加者は 34 名 (男性 17 名, 女性 17 名, 年齢は 15~24 歳) であった。

4.3 実験結果

4.3.1 全体の評価とカテゴリ別分析

実験全体の結果をまとめる。表 4 は、4 つの手法の「魅力度」に関する 5 段階評価の平均点と標準偏差 (SD) をカテゴリ別に示し、最終行に全体の平均点と SD を加えたものである。表 4 の全体平均では全ての名詞を Positive プロンプトに用いる手法 1 が 3.00 点と最も高

表 4: 手順 3「絶対評価」により得られたカテゴリ別および全体の 5 段階評価平均点と標準偏差 (最大値:太字・下線, SD: 標準偏差)

ジャンル	手法 1(SD)	手法 2(SD)	手法 3(SD)	手法 4(SD)
アート・建築・デザイン	2.84 (1.21)	2.49 (1.12)	2.62 (1.09)	2.88 (1.21)
コミックス	3.03 (1.07)	2.46 (1.21)	2.52 (1.13)	2.98 (1.15)
サイエンス・テクノロジー	3.12 (0.99)	2.41 (1.09)	2.61 (1.11)	2.91 (1.11)
スポーツ・アウトドア	2.98 (1.12)	2.69 (1.24)	2.40 (1.16)	2.94 (1.24)
ノンフィクション	3.01 (1.25)	2.48 (1.10)	2.55 (1.15)	2.75 (1.21)
ビジネス・経済	3.11 (1.16)	2.49 (1.14)	2.49 (1.14)	3.02 (1.13)
音楽	3.27 (1.09)	2.51 (1.03)	2.78 (1.12)	3.16 (1.18)
絵本・児童書	3.26 (1.09)	2.31 (1.01)	2.71 (1.17)	3.34 (1.11)
教育・自己啓発	2.98 (1.18)	2.65 (1.13)	2.42 (1.06)	2.50 (1.10)
芸能・エンターテインメント	2.74 (1.15)	2.61 (1.04)	2.41 (1.16)	2.72 (1.19)
事典・年鑑・本・ことば	3.01 (1.10)	2.69 (1.10)	2.86 (1.13)	3.13 (1.33)
趣味・実用	2.85 (1.06)	2.39 (1.05)	2.67 (1.18)	2.76 (1.22)
人文・思想・宗教	3.08 (1.05)	2.83 (1.10)	2.74 (1.15)	3.03 (1.20)
政治・社会	2.96 (0.99)	2.75 (1.13)	2.70 (1.15)	2.71 (1.18)
文芸・評論	2.86 (1.11)	2.62 (1.15)	2.35 (1.09)	2.84 (1.08)
暮らし・健康・料理	2.94 (1.06)	2.75 (1.18)	2.89 (1.19)	2.78 (1.10)
旅行・紀行	2.84 (1.09)	2.67 (1.12)	2.64 (1.14)	2.99 (1.11)
歴史・地理	3.12 (1.11)	2.57 (1.13)	2.71 (1.13)	2.99 (1.17)
全体平均	3.00 (1.12)	2.58 (1.13)	2.61 (1.15)	2.91 (1.18)
最大値の合計	14	0	0	4

表 5: 手順 4「相対評価」により得られたカテゴリ別の選択手法の分布 (最大数:太字・下線, %: 割合)

ジャンル	手法 1(%)	手法 2(%)	手法 3(%)	手法 4(%)
アート・建築・デザイン	57 (33.5%)	39 (22.9%)	28 (16.5%)	46 (27.1%)
コミックス	53 (31.2%)	45 (26.5%)	22 (12.9%)	50 (29.4%)
サイエンス・テクノロジー	69 (40.6%)	29 (17.1%)	35 (20.6%)	37 (21.8%)
スポーツ・アウトドア	44 (25.0%)	54 (31.8%)	25 (14.7%)	47 (27.6%)
ノンフィクション	70 (41.2%)	27 (15.9%)	36 (21.2%)	37 (21.8%)
ビジネス・経済	69 (40.6%)	33 (19.4%)	26 (15.3%)	42 (24.7%)
音楽	67 (39.4%)	19 (11.2%)	31 (18.2%)	53 (31.2%)
絵本・児童書	66 (38.8%)	21 (12.4%)	24 (14.1%)	59 (34.7%)
教育・自己啓発	63 (37.1%)	53 (31.2%)	24 (14.1%)	30 (17.6%)
芸能・エンターテインメント	52 (30.6%)	48 (28.2%)	37 (21.8%)	33 (19.4%)
事典・年鑑・本・ことば	48 (28.2%)	26 (15.3%)	39 (22.9%)	57 (33.5%)
趣味・実用	51 (30.0%)	33 (19.4%)	42 (24.7%)	44 (25.9%)
人文・思想・宗教	47 (27.6%)	47 (27.6%)	35 (20.6%)	41 (24.1%)
政治・社会	52 (30.6%)	43 (25.3%)	39 (22.9%)	36 (21.2%)
文芸・評論	57 (33.5%)	41 (24.1%)	30 (17.6%)	42 (24.7%)
暮らし・健康・料理	56 (32.9%)	42 (24.7%)	35 (20.6%)	37 (21.8%)
旅行・紀行	56 (32.9%)	32 (18.8%)	33 (19.4%)	49 (28.8%)
歴史・地理	63 (37.1%)	37 (21.8%)	35 (20.6%)	35 (20.6%)
全体合計	1,040 (34.0%)	669 (21.9%)	576 (18.8%)	775 (25.3%)
最大数の合計	16	2	0	1

く、次いで重要度の順に単語に重み付けを行う手法 4 が 2.91 点となった。標準偏差は全体平均で 1.12 から 1.18 の範囲にあり、各手法ともに評価には個人差があることが示された。特に手法 1 は平均点が最も高い一方で、標準偏差は 1.12 と最も小さく、比較的评价が安定していた。表 4 のカテゴリ別に見ると、「音楽」(3.27 点)や「ビジネス・経済」(3.11 点)などで手法 1 の評価点が特に高い。一方で、「絵本・児童書」(3.34 点)、「事典・年鑑・本・ことば」(3.13 点)など 4 つのジャンルでは手法 4 が最高評価となった。

表 5 は、各手法が最終的に「最も購買意欲が湧く」として選択された回数を示す。表 5 の全体合計では手法 1 が 1,040 件 (34.0%) と最も多く選択された。表 5 のカテゴリ別に見ても、多くのジャンルで手法 1 の選択数が最も多い (例:「ノンフィクション」で 70 件、「ビジネス・経済」で 69 件)。一方で、「事典・年鑑・本・ことば」では手法 4 が 57 件と最も多く選択され、「絵本・児童書」でも手法 4 が 59 件と手法 1 (66 件) に次いで多く選択された。

表 6: 手順3「絶対評価」により得られた性別・カテゴリ別の5段階評価平均点と標準偏差（最大値:太字・下線, SD: 標準偏差）

性別	ジャンル	手法 1(SD)	手法 2(SD)	手法 3(SD)	手法 4(SD)
男性	アート・建築・デザイン	2.89 (1.13)	2.65 (1.13)	2.68 (0.95)	2.89 (1.12)
	コミックス	3.11 (1.04)	2.54 (1.17)	2.68 (1.17)	3.05 (1.13)
	サイエンス・テクノロジー	3.06 (0.94)	2.60 (1.12)	2.68 (1.01)	2.95 (1.00)
	スポーツ・アウトドア	3.08 (1.08)	2.66 (1.22)	2.46 (1.22)	3.00 (1.21)
	ノンフィクション	2.99 (1.21)	2.61 (1.13)	2.67 (1.16)	2.73 (1.12)
	ビジネス・経済	3.16 (1.14)	2.84 (1.12)	2.55 (1.19)	3.14 (1.12)
	音楽	3.39 (1.06)	2.69 (1.04)	2.80 (1.09)	3.19 (1.17)
	絵本・児童書	3.33 (1.08)	2.49 (1.00)	2.72 (1.13)	3.36 (1.06)
	教育・自己啓発	3.12 (1.18)	2.80 (1.13)	2.46 (1.01)	2.55 (1.15)
	芸能・エンターテインメント	2.72 (1.10)	2.56 (1.01)	2.60 (1.12)	2.81 (1.22)
	事典・年鑑・本・ことば	3.15 (1.03)	2.61 (1.09)	2.88 (1.08)	3.15 (1.29)
	趣味・実用	2.99 (1.02)	2.40 (1.08)	2.67 (1.18)	2.75 (1.21)
	人文・思想・宗教	3.18 (1.05)	2.84 (1.10)	2.92 (1.12)	3.07 (1.18)
	政治・社会	2.98 (0.93)	2.85 (1.10)	2.81 (1.22)	2.75 (1.17)
	文芸・評論	2.91 (1.08)	2.64 (1.10)	2.54 (1.12)	2.89 (0.95)
	暮らし・健康・料理	3.04 (1.09)	2.71 (1.10)	3.02 (1.24)	2.80 (1.03)
	旅行・紀行	2.91 (1.08)	2.79 (1.14)	2.75 (1.08)	2.99 (1.01)
	歴史・地理	3.24 (1.02)	2.62 (1.12)	2.66 (1.16)	3.08 (1.17)
	男性平均	2.98 (1.08)	2.66 (1.11)	2.64 (1.14)	2.91 (1.15)
女性	アート・建築・デザイン	2.78 (1.28)	2.34 (1.08)	2.55 (1.20)	2.86 (1.29)
	コミックス	2.95 (1.09)	2.38 (1.25)	2.36 (1.07)	2.91 (1.16)
	サイエンス・テクノロジー	3.18 (1.04)	2.21 (1.03)	2.54 (1.19)	2.87 (1.21)
	スポーツ・アウトドア	2.88 (1.15)	2.72 (1.25)	2.34 (1.09)	2.88 (1.26)
	ノンフィクション	3.02 (1.28)	2.34 (1.06)	2.44 (1.13)	2.78 (1.30)
	ビジネス・経済	3.06 (1.17)	2.14 (1.05)	2.42 (1.08)	2.91 (1.12)
	音楽	3.15 (1.10)	2.32 (0.99)	2.75 (1.15)	3.14 (1.19)
	絵本・児童書	3.20 (1.10)	2.12 (0.99)	2.69 (1.21)	3.31 (1.15)
	教育・自己啓発	2.85 (1.16)	2.51 (1.12)	2.38 (1.10)	2.45 (1.05)
	芸能・エンターテインメント	2.76 (1.20)	2.65 (1.06)	2.22 (1.16)	2.62 (1.15)
	事典・年鑑・本・ことば	2.87 (1.15)	2.76 (1.11)	2.85 (1.17)	3.11 (1.37)
	趣味・実用	2.72 (1.07)	2.39 (1.03)	2.67 (1.17)	2.78 (1.24)
	人文・思想・宗教	2.99 (1.05)	2.82 (1.10)	2.55 (1.15)	2.99 (1.22)
	政治・社会	2.95 (1.04)	2.66 (1.14)	2.59 (1.06)	2.66 (1.15)
	文芸・評論	2.82 (1.13)	2.60 (1.20)	2.15 (1.01)	2.78 (1.20)
	暮らし・健康・料理	2.84 (1.03)	2.79 (1.25)	2.76 (1.12)	2.75 (1.18)
	旅行・紀行	2.78 (1.10)	2.55 (1.08)	2.52 (1.17)	2.99 (1.20)
	歴史・地理	3.01 (1.18)	2.52 (1.14)	2.76 (1.09)	2.91 (1.15)
	女性平均	3.03 (1.14)	2.49 (1.13)	2.59 (1.15)	2.92 (1.22)
最大値の合計		13	0	0	7

4.3.2 性別による傾向の分析

表 6 は、5 段階評価の平均点と SD を性別とカテゴリでクロス集計し、最終行に性別ごとの全体平均と標準偏差を加えたものである。表 6 によれば、女性は多くのジャンルで手法 1 を最も高く評価し、全体平均でも**3.03**点 (SD 1.14) と最も高かった。男性も全体平均では手法 1 が**2.98**点 (SD 1.08) と最も高かった。標準偏差を見ると、男性は手法 1 の SD が 1.08 と最も小さく、評価のばらつきが小さかったのに対し、女性は手法 2 の SD が 1.13 と最も小さかった。

表 7 は、最終選択数を性別とカテゴリでクロス集計し、最終行に性別ごとの合計と割合を加えたものである。全体合計で見ると、女性は手法 1 を選択した割合が 36.1%に達したのに対し、男性は 31.8%であった。男性は「音楽」(39 件)や「ノンフィクション」(30 件)などで手法 1 を最も多く選択した。女性は「サイエンス・テクノロジー」(40 件)や「ノンフィクション」(40 件)で手法 1 を最も多く選択したが、「音楽」(31 件)や「事典・年鑑・本・ことば」(31 件)では手法 4 を最も多く選択した。

5 考察

本章では、主観評価実験の結果に基づき、提案した 4 つのプロンプト構成手法の有効性、評価のばらつき、

表 7: 手順4「相対評価」により得られた性別・カテゴリ別の選択手法（最大数:太字・下線, %: 割合）

性別	ジャンル	手法 1(%)	手法 2(%)	手法 3(%)	手法 4(%)
男性	アート・建築・デザイン	29 (34.1%)	25 (29.4%)	11 (12.9%)	20 (23.5%)
	コミックス	25 (29.4%)	25 (29.4%)	10 (11.8%)	25 (29.4%)
	サイエンス・テクノロジー	29 (34.1%)	20 (23.5%)	17 (20.0%)	19 (22.4%)
	スポーツ・アウトドア	23 (27.1%)	26 (30.6%)	14 (16.5%)	22 (25.9%)
	ノンフィクション	30 (35.3%)	14 (16.5%)	24 (28.2%)	17 (20.0%)
	ビジネス・経済	31 (36.5%)	20 (23.5%)	11 (12.9%)	23 (27.1%)
	音楽	39 (45.9%)	13 (15.3%)	11 (12.9%)	22 (25.9%)
	絵本・児童書	30 (35.3%)	14 (16.5%)	10 (11.8%)	31 (36.5%)
	教育・自己啓発	32 (37.6%)	27 (31.8%)	10 (11.8%)	16 (18.8%)
	芸能・エンターテインメント	26 (30.6%)	17 (20.0%)	23 (27.1%)	19 (22.4%)
	事典・年鑑・本・ことば	22 (25.9%)	16 (18.8%)	21 (24.7%)	26 (30.6%)
	趣味・実用	26 (30.6%)	15 (17.6%)	21 (24.7%)	23 (27.1%)
	人文・思想・宗教	24 (28.2%)	25 (29.4%)	18 (21.2%)	18 (21.2%)
	政治・社会	23 (27.1%)	23 (27.1%)	25 (29.4%)	14 (16.5%)
	文芸・評論	21 (24.7%)	23 (27.1%)	20 (23.5%)	21 (24.7%)
	暮らし・健康・料理	25 (29.4%)	19 (22.4%)	23 (27.1%)	18 (21.2%)
	旅行・紀行	25 (29.4%)	18 (21.2%)	21 (24.7%)	18 (21.2%)
	歴史・地理	27 (31.8%)	20 (23.5%)	20 (23.5%)	18 (21.2%)
	男性合計	487 (31.8%)	360 (23.5%)	310 (20.3%)	373 (24.4%)
最大数の合計		12	3	2	3
女性	アート・建築・デザイン	28 (32.9%)	14 (16.5%)	17 (20.0%)	26 (30.6%)
	コミックス	28 (32.9%)	20 (23.5%)	12 (14.1%)	25 (29.4%)
	サイエンス・テクノロジー	40 (47.1%)	9 (10.6%)	18 (21.2%)	18 (21.2%)
	スポーツ・アウトドア	21 (24.7%)	28 (32.9%)	11 (12.9%)	25 (29.4%)
	ノンフィクション	40 (47.1%)	13 (15.3%)	12 (14.1%)	20 (23.5%)
	ビジネス・経済	38 (44.7%)	13 (15.3%)	15 (17.6%)	19 (22.4%)
	音楽	28 (32.9%)	6 (7.1%)	20 (23.5%)	31 (36.5%)
	絵本・児童書	36 (42.4%)	7 (8.2%)	14 (16.5%)	28 (32.9%)
	教育・自己啓発	31 (36.5%)	26 (30.6%)	14 (16.5%)	14 (16.5%)
	芸能・エンターテインメント	26 (30.6%)	31 (36.5%)	14 (16.5%)	14 (16.5%)
	事典・年鑑・本・ことば	26 (30.6%)	10 (11.8%)	18 (21.2%)	31 (36.5%)
	趣味・実用	25 (29.4%)	18 (21.2%)	21 (24.7%)	21 (24.7%)
	人文・思想・宗教	23 (27.1%)	22 (25.9%)	17 (20.0%)	23 (27.1%)
	政治・社会	29 (34.1%)	20 (23.5%)	14 (16.5%)	22 (25.9%)
	文芸・評論	36 (42.4%)	18 (21.2%)	10 (11.8%)	21 (24.7%)
	暮らし・健康・料理	31 (36.5%)	23 (27.1%)	12 (14.1%)	19 (22.4%)
	旅行・紀行	31 (36.5%)	14 (16.5%)	12 (14.1%)	28 (32.9%)
	歴史・地理	36 (42.4%)	17 (20.0%)	15 (17.6%)	17 (20.0%)
	女性合計	553 (36.1%)	309 (20.2%)	266 (17.4%)	402 (26.3%)
最大数の合計		14	2	0	3

およびカテゴリ・性別による傾向の違いについて考察する。

実験全体の結果（表 4 および表 5 の最終行）では、5 段階評価の平均点と最終選択率の順位が一貫しており、手法 1 が最も高い評価（平均点 3.00 点（表 4, 全体平均）、選択率 34.0%（表 5, 全体合計））を獲得した。これは、書籍説明文中の名詞をすべてプロンプトとして利用すること（手法 1, 全名詞 Positive）で、画像生成 AI に網羅的な情報を与え、結果として書籍の内容を意味的に一貫して反映する画像が生成されたためと推察される。全体平均において、手法 1 は標準偏差 1.12（表 4, 全体平均）と最も小さく、評価のばらつきが最も小さかった。これは、情報量の多い手法 1（全名詞 Positive）では生成の方向性が安定し、被験者間で評価のコンセンサスが最も得られやすかったことを示唆している。一方で、手法 4（重み付け）は標準偏差が 1.18（表 4, 全体平均）と最もばらつきが大きかった。これは、特定の単語を強調する処理が、画像生成 AI に与えた内容の間の一貫性を保ちにくくなり、評価の振れ幅を大きくした可能性を示唆している。

手法 2（Positive/Negative 分割）と手法 3（上位 10 件）は、全体的に低い評価となった（表 4 の全体平均で、手法 2 は 2.58 点、手法 3 は 2.61 点）。手法 1 や手法 4 は Positive プロンプトに平均して約 30 語の名詞が含まれているのに対し、手法 3 は、書籍の複雑な内容をわずか 10 語の名詞のみで表現しようとしたため、情報量が不足し、生成画像が書籍の内容との意味的な一貫性を損ねた可能性がある。手法 2 は、重要度の低い

単語を Negative プロンプトに指定することで、かえって内容のバランスを欠いた画像を生成させ、評価を下げる要因となったと推察される。これら手法2と手法3の結果は、視覚情報とテキスト情報が意味的に一貫しているほど好意的な評価につながるという Coherence Principle[Mayer 05] の観点から、書籍の内容をスムーズに理解させる上で有効ではなかったことを示唆している。

「ノンフィクション」(評価点 3.01, 選択数 70 件), 「ビジネス・経済」(評価点 3.11, 選択数 69 件), 「サイエンス・テクノロジー」(評価点 3.12, 選択数 69 件)といったジャンルにおいて、手法1(全名詞 Positive)が最も高い評価と選択数を獲得する傾向が見られた。これらのジャンルの説明文には、実際に「システム」「構造」「社会」「経済」といった、具体的形状を持たない抽象的な単語や、要素間の関係性を示す単語が多く含まれている。説明文中的名詞をすべて均一な重みで網羅的に用いる手法1は、特定の単語を突出させることなく、これらの抽象的な単語と背景や周辺要素を表現する単語を平等に扱う。そのため、書籍が持つ「複合的な文脈」や「世界感」を画像全体で表現するのに適していたと推察される。すなわち、テキスト情報の複雑さと画像の視覚的な情報量が一致することで、Coherence Principle[Mayer 05] に基づき、好意的な評価につながったと考えられる。

一方で、「事典・年鑑・本・ことば」(選択数 57 件で最多, 評価点 3.13 で2位)や「絵本・児童書」(評価点 3.34 で最多, 選択数 59 件で2位)といったジャンルでは、手法4(重み付け)も高い評価を得た。これらのジャンルでは、複雑な背景描写よりも、特定の「主題(キャラクターやトピック)」が明確であることが表紙デザインとして求められる傾向にある。手法4は重要度の高い単語を強調することで、画像生成 AI に明確な焦点を与え、主題が際立った画像を生成する効果がある。これにより、読者が内容を把握しやすくなる「知覚的流暢性(Processing Fluency)」[Reber 04] が高まり、好評価につながったと考えられる。

以上のことから、全ての情報を網羅する手法1は「文脈や世界観」を重視するジャンルに、重要語を強調する手法4は「主題の明確さ」を重視するジャンルに適しているという、ジャンルごとの最適なプロンプト構成の違いが示唆された。「スポーツ・アウトドア」カテゴリでは、他のジャンルと異なり手法2が最も多く選択された(選択率 31.8%, 表5)。これは、重要度の低い情報を Negative プロンプトで排除する操作が、かえって主題(アクションや選手など)を際立たせる効果を生んだ可能性がある。

性別による傾向(表6, 表7)にも、平均点と選択率のねじれが表れている。男性は、平均点(2.98点)、選択率(31.8%)において手法1を最も支持した。一方、

女性は、平均点(3.03点)と選択率(36.1%)では手法1を最も支持したが、手法4に対しても一部のジャンル(事典・年鑑・本・ことば、音楽など)で強い支持を示した。これは、女性被験者において、手法4(重み付け)が生成する「焦点が明確な画像」が好まれるケースがあったことを示唆している。

6 結論

本研究では、画像生成 AI を用いて書籍の説明文から表紙画像を生成し、どのプロンプト構成が書籍表紙の誘引力を高めるかについて主観評価実験を行った。その結果、説明文中的名詞をすべて用いる手法(手法1)が全体的に最も好まれる傾向がある一方で、重要語を重み付けする手法(手法4)も特定のジャンルや性別においては有効であることが明らかになった。この結果は、情報量が多く意味的に一貫した画像や、焦点が明確な画像が好まれやすいという認知心理学の理論によって支持される[Reber 04, Mayer 05]。

今後は、被験者層の拡大や、生成された画像自体の定量的分析に取り組むことで、AI による書籍デザイン支援の可能性をさらに追求できると期待する。

参考文献

- [Orth 08] Orth, U. R., & Malkewitz, K.: Holistic package design and consumer brand impressions, *Journal of Marketing*, Vol. 72, No. 3, pp. 64-81 (2008).
- [Liu 17] Liu, Y., Li, K., Hu, H.: The impact of positive imagery in visual marketing on consumer approach behavior, *Journal of Business Research*, Vol. 79, pp. 106-114 (2017).
- [Bloch 95] Bloch, P. H.: Seeking the ideal form: Product design and consumer response, *Journal of marketing*, Vol. 59, No. 3, pp. 16-29 (1995).
- [Bruce 97] Bruce, M., & Cooper, R.: Marketing and design in the new product development process, *Marketing intelligence & planning*, Vol. 15, No. 3, pp. 100-106 (1997).
- [Gudinavičius 17] Gudinavičius, A., & Šuminas, A.: Choosing a book by its cover: analysis of a reader's choice, *Journal of Documentation*, Vol. 74, No. 2 (2017).

- [Hagtvedt 08] Hagtvedt, H., & Patrick, V. M.: Art and the brand: The role of visual art in enhancing brand extendibility, *Journal of Consumer Psychology*, Vol. 18, No. 3, pp. 212-222 (2008).
- [Jian 19] Jian, W., Lyu, S., & Li, Y.: The influence of book cover design on reader's perception of content and quality, *Publishing Research Quarterly*, Vol. 35, pp. 466-476 (2019).
- [Jiang 13] Jiang, Z., Chan, J., Tan, B. C., & Chua, W. Y.: Effects of interactivity on website involvement and purchase intention, *Journal of the Association for Information Systems*, Vol. 14, No. 1, p. 3 (2013).
- [Mayer 05] Mayer, R. E.: The Cambridge handbook of multimedia learning, Cambridge university press (2005).
- [Park 23] Park, J. Y., Kim, C., Park, S., Dio, K.: Do you judge a book by its cover? Online book purchases between Japan and France, *Asia Pacific Journal of Marketing and Logistics*, Vol. 35, No. 2 (2023).
- [Ramesh 21] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... Sutskever, I.: Zero-shot text-to-image generation, *International Conference on Machine Learning*, PMLR, pp. 8821-8831 (2021).
- [Reber 04] Reber, R., Schwarz, N., Winkielman, P.: Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?, *Personality and social psychology review*, Vol. 8, No. 4, pp. 364-382 (2004).
- [Rombach 22] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684-10695 (2022).
- [Silayoi 07] Silayoi, P., & Speece, M.: The importance of packaging attributes: a conjoint analysis approach, *European Journal of Marketing*, Vol. 41, No. 11/12, pp. 1495-1517 (2007).