

[招待講演]

## 生成 AI を用いた電子書籍の図版説明の実証的研究

仁科 哲<sup>†</sup> ファイサル・マウラナ<sup>‡</sup>

<sup>† ‡</sup> 株式会社和文 〒151-0051 東京都渋谷区千駄ヶ谷 3-53-17-307

E-mail: <sup>†</sup> [t-nishina@wabun.co.jp](mailto:t-nishina@wabun.co.jp), <sup>‡</sup> [m-faisal@wabun.co.jp](mailto:m-faisal@wabun.co.jp)

**あらまし** 顕著に能力が向上した現在の生成 AI の画像認識能力は、電子書籍のアクセシビリティに必要な図版内容の説明に活用できる可能性が高い。書店で販売されている実際の電子書籍の本文と図版を主要3社5種類の汎用生成 AI に読み込ませ、図版の内容を説明させた場合の実用性を検証した。「実用性」判定の基準は、説明によって図の概要が把握可能なこととした。本文と画像の両方を与えた場合、生成 AI による図版内容の説明は、93% の高い確率で実用性があると判定された。商用電子書籍の製作において図版の内容を説明するテキストの作成は人的作業では難しい点があるが、現在の AI の文章／画像認識の能力を用いれば、技術的には低コストで実現できる可能性をこの結果は示している。

**キーワード** 生成 AI, 画像認識, 文脈理解, 電子書籍, 商業出版

[Invited Lecture]

## An Empirical Study on AI-Generated Figure Descriptions in Commercial E-books

Tetsu NISHINA<sup>†</sup> and Maulana FAISAL<sup>‡</sup>

<sup>† ‡</sup> Wabun Inc., 3-53-17-307 Sendagaya, Shibuya-ku, Tokyo 151-0051, Japan

E-mail: <sup>†</sup> [t-nishina@wabun.co.jp](mailto:t-nishina@wabun.co.jp), <sup>‡</sup> [m-faisal@wabun.co.jp](mailto:m-faisal@wabun.co.jp)

**Abstract** The significantly improved image recognition capabilities of current generative AI suggest a high potential for use in describing figures essential for the accessibility of e-books. In this study, we examined the practicality of using five general-purpose generative AI models from three major companies to describe figures in commercially available e-books by providing both the text and the figures. The criterion for determining "practicality" was whether the generated description allowed for a general understanding of the figure. The results showed that when both textual content and images were provided, the AI-generated figure descriptions were deemed practical in 93% of cases. Providing spoken descriptions of figures in books is essential for accessibility but has been challenging to implement. This study demonstrates that, with the current capabilities of AI in image recognition, such descriptions could be achieved at a low cost from a technical perspective.

**Keyword** Generative AI, Image Recognition, Contextual Understanding, E-books, Commercial Publishing

### 1. 生成 AI の画像認識は電子書籍の画像を正しく説明できるか

視覚障がい者の利用度が高いスマホでは、近年、デバイスの機能により、画像内の文字を認識して読み上げることが可能になってきた。一方、汎用的に使われる生成 AI でも画像ファイルの内容を認識する能力が顕著に向上しており、それらの機能を用いたアプリも実用化されている。すなわち技術的には、電子書籍の

図版内にある文字の認識、および図版の内容の読み上げ説明をデバイスで行うことは十分可能となっているといえる。

本研究は、電子書店で販売されるリフロー型電子書籍の実際の本文と画像を用いて、現在の汎用的な生成 AI がどの程度まで画像を正しく認識し、説明する能力があるのかを検証したものである。

生成 AI が書籍の文章を瞬時に把握し、画像の内容

を正確に説明できれば、また、その技術を電子書籍ビューワでも利用できる状況になれば、電子書籍のアクセシビリティは格段に向上すると思われる。さらに、現在多くの出版社が取り組む画像の代替テキストの作成・追加の工程にも寄与するであろう。今回の検証ではこの画像の説明自動化の技術的可能性を確認するため、生成AIの出力を文字単位で詳細に評価した。

## 2. 全文校正による検証と判定

### 2.1. タイプ別 11 画像を主要 3 社 5AI で認識

テストでは汎用 AI を提供する主要 3 社の 5 種類の生成 AI を API 環境で使用した。画像はシンプルなグラフから複雑なダイアグラムまでの 11 種の図版(図 1)とし、画像とそれに関係する本文を読み込ませ、統一した短いプロンプトを与えた。生成された回答は商業出版の校正者が全文を本文と文字照合し、画像内文字および画像の説明の誤植/間違いを確認した。

①

図 1 米国の大学における障害学授業の増加 (Cushing & Smith (2009) Chart 4 を参照)

②

表 1 2008 年における障害学授業の概要 (Cushing & Smith (2009) Figure 1 を改変)

国	障害学課程の授業数		科目数の合計		障害学学位課程数 (学位)	
	他課程の授業数	選択科目数	BA (学士)	MA (修士)	博士	修士課程等
アメリカ	224	66	290	1	13	3
カナダ	57	29	86	3	1	1
イギリス	79	10	89	7	0	2
オセアニア	60	3	63	4	0	1
合計	420	108	528	15	14	7

図 2 障害学課程の設置学科分布 (Cushing & Smith (2009) Chart 6 を参照)

③

図 2 障害学課程の設置学科分布 (Cushing & Smith (2009) Chart 6 を参照)

④

人の背丈を超える草束 (ロコンゴ)

⑤

ロンドンのデザインミュージアムにあるオレンジ色の自動車。キューブリック監督による映画『時計じかけのオレンジ』で使用されたもの。

⑥

図 1 SDGs の 17 目標

⑦

図 1 宮城県民の年齢構成

⑧

ピクトグラムに見る男性と女性 (2010年代)

施設	一般向け情報		WC (男性)		WC (女性)	
	男性	女性	男性	女性	男性	女性
駅	男性	女性	男性	女性	男性	女性
バス	男性	女性	男性	女性	男性	女性
劇場	男性	女性	男性	女性	男性	女性
大学	男性	女性	男性	女性	男性	女性
劇場	男性	女性	男性	女性	男性	女性

左の欄が一般向けの情報である。駐車料金の支払い、ゴミはゴミ箱に、緊急時には集合などの呼びかけ。右の欄は男性用トイレと女性用トイレの構成。男性性の普遍性で抽象的な表現が強く示されている。一般的な多行者の表示も、男性用トイレと同じデザインだ。男性が優先して人間を代表しているとするれば、女性は常に自分のジェンダーに紐づけられている。ドレス、ヘアスタイル、閉じた胸、そして子どもの世話。(写真は2018年にフランスで撮影された。ただし、一部はドイツ・と中国 \*\*)

⑨

図表 2.2 スウェーデンにおける子どもアドボカシーの全体像

⑩

アドボカシー理解の枠組

子どもが経験する問題の根本原因は制度にある。制度改革は個別課題のより良い解決を支援する。

システムアドボカシー (政策提言・制度改革) ↔ 個別アドボカシー (意見表明支援・代弁)

図表 1.1 個別アドボカシーとシステムアドボカシー

出所: Office of the Child, Youth and Family Advocate 1997, p.3

⑪

図表 2.2 スウェーデンにおける子どもアドボカシーの全体像

### テスト対象の生成 AI

- Google 社 Gemini
- (1) 「gemini-1.5-pro-exp-0827」
- (2) 「gemini-1.5-flash-exp-0827」
- OpenAI 社 ChatGPT
- (3) 「gpt-4o(omuni)」
- Anthropic 社 Claude
- (4) 「claude-3-opus-20240229」
- (5) 「claude-3.5-sonnet-20240620」

図 1 テストに用いた 11 の図版と出典

- ① 棒グラフ 『障害学の展開』明石書店 (2024) P62-63 図 1
- ② 複雑な表 同上 表 1
- ③ 円グラフ 同上 P63-64 図 2
- ④ モノクロ写真 『コンゴ民主共和国を知るための 50 章』明石書店 (2024) P35-36 写真
- ⑤ カラー写真 『オックスフォード哲学者奇行』明石書店 (2022) P27-29 写真
- ⑥ 広く知られた図版 『SDGs と地域社会』明石書店 (2022) P6 図 1
- ⑦ 1 色の濃淡地図 同上 P64 図 1
- ⑧ 図形と Cap が複雑な図版 『マチズモの人類史』明石書店 (2024) P97 図
- ⑨ Cap が長い写真 同上 P144 写真
- ⑩ シンプルなダイアグラム 『子どもアドボカシー Q & A』明石書店 (2024) P23 図表 1.1
- ⑪ 複雑なダイアグラム 同上 P68 図表 2.2

## 2.2. 「実用的か」が評価の基準

生成 AI による出力はハルシネーションによる間違いが含まれるケースがある。判定ではハルシネーションの存在を聴者が把握していることを前提とし、微細な間違いが含まれる場合でも「図版の概要を大きな不明がなく把握できる」場合は「実用性あり」とした。図の内容が間違っただけの場合、および理解ができない／理解が難しい説明は「実用性なし」と判定した。

## 3. 画像内文字の認識結果

画像の文字認識は概ね正確であり、間違いがある場合も内容の理解を損なわない程度が多かった。

要素が多い、構成が複雑な図版は文字の認識に間違いが見られるケースが多かった。ただし多くの場合、数カ所にとどまった。

広く社会的に知られた図版では画像が複雑であっても文字認識は極めて正確だった。

ハルシネーションによる文字の認識違いは程度の差があるが一定数が認められた。多くは許容できる範囲だが、図の誤読に直結するケースも見られた。

## 4. 画像の説明の評価結果

画像を関係する本文とともに読み込ませた場合、図の説明で「実用性あり」と評価されたケースは45例中42例(93.3%)であった(図2)。

プロンプトで本文を参照して説明させた場合、ほとんどのテストで本文の内容を踏まえた説明がされた。図版にも本文にもない要素を付加する例も見られた。

## 5. 結論

生成 AI が画像と本文を読み込んで出力する「画像内の文字認識」および「画像の内容の説明」は、画像の概要の理解を目的とした場合は現状でも実用的と考えられる。

## 6. 謝辞

テストに用いた電子書籍の本文、図版のデータは、株式会社明石書店刊行の販売用リフロー型電子書籍から使用している。アクセシビリティの拡大に役立てば、と快く許諾していただいた明石書店の大江道雅社長と製作部に感謝いたします。

図2 画像別の実用性評価

テスト番号	画像の種類	画像の説明の実用性あり	文字起こしの実用性あり
1	棒グラフ	4 AI中 4 AI	4 AI中 3 AI
2	表	4 AI中 4 AI	4 AI中 3 AI
3	円グラフ	4 AI中 4 AI	4 AI中 3 AI
4	モノクロ写真	4 AI中 4 AI	——
5	カラー写真	5 AI中 4 AI	——
6	アイコン図版	4 AI中 4 AI	(「画像の説明」に含める)
7	1色の濃淡分け地図	4 AI中 4 AI	(「画像の説明」に含める)
8	図形とCapが複雑な図版	4 AI中 2 AI	(「画像の説明」に含める)
9	Capが長い写真	4 AI中 4 AI	——
10	シンプルダイアグラム	4 AI中 4 AI	(「画像の説明」に含める)
11	複雑ダイアグラム	4 AI中 4 AI	(「画像の説明」に含める)