

# LLMによる画像内のバリア描写の基礎的分析

松林 瞭<sup>1</sup> 池田 悠星<sup>1</sup> 藤井 大輝<sup>1</sup> 土岐田 力輝<sup>2</sup> 呉 健朗<sup>3</sup> 宮田 章裕<sup>1,a)</sup>

**概要:** 公共空間や都市施設のバリアフリー化が進む一方で、障害者や高齢者が移動時に直面する物理的バリアは依然として多く、詳細なアクセシビリティ情報が求められている。ボランティアによるクラウドソーシングを用いたバリア情報収集が行われることもあるが、彼らに詳細なバリアに関するテキスト記述を求めることは困難であり、収集される情報には記述の質や粒度にばらつきが生まれる。一方、大規模言語モデル (Large Language Model, LLM) と視覚情報処理技術の統合により、画像から自然言語による説明文の生成が可能となった。このアプローチを用いると低コストで大量のバリア画像の描写が取得できる可能性があるが、LLMによる生成文は視点の偏りを含むという問題点がある。そこで本研究では、LLMに特定の“役割”や“視点”を与えるロールプレイ手法を用い、異なる障害種別や利用者視点に応じた多様な説明文生成の可能性を検討する。

## 1. はじめに

公共空間や都市施設のバリアフリー化が進展する一方で、障害者や高齢者が移動時に直面する物理的バリアは依然として多様であり、現地の詳細なアクセシビリティ情報の提供が求められている。特に、移動支援や経路推薦システムなどにおいて、現実空間のバリア情報を正確に収集し、地図上で可視化することは不可欠である [1], [2]。こうした情報の整備において、ボランティアによる地理情報提供やクラウドソーシングが重要な役割を果たしてきた [3], [4], [5]。しかし、屋外で作業するクラウドワーカーに詳細なテキスト記述を求めることは困難であり、作業負担や作業環境の制約から、記述の質や粒度に大きなばらつきが生じやすい。

近年、大規模言語モデル (Large Language Model, LLM) と視覚情報処理技術の統合が進み [6], [7]、画像から自然言語による説明文を生成することが可能となった。この技術を用いれば、クラウドワーカーが撮影した現地画像に対して、自動的にバリアに関するテキスト説明を付与することが期待される。しかし、LLMによる生成文はしばしば視点の偏りを含むことが指摘されており [8], [9]、障害種別によるバリア認識の差異を十分に反映できていない [10]。

この課題に対し、LLMに特定の“役割”や“視点”を与えることで、多様な人間の立場を模倣させる手法が提案されている [11], [12]。ただし、これらの手法は一般的な役割

のみを与えている研究であり、障害者等の役割についても適切な生成が行えるか不明である。よって、障害者向けの視点でも十分な記述ができているか確認する必要がある。

本研究では、LLMによる画像説明生成手法をバリアに関する説明文生成に応用することで [13]、異なる障害種別や利用者視点に応じた多様な説明文生成の可能性を検討する。

## 2. 関連研究

### 2.1 LLMによる画像説明生成

Liuら [6] や Alayracら [7] は、画像入力に対して自然言語の説明を生成するマルチモーダルLLMの基礎研究を行い、視覚的文脈理解の精度向上を示した。一方、Kazaiら [14] は、クラウドソーシング環境での人間のテキスト記述の精度やバイアスを分析し、人間ベースのアノテーションに限界があることを明らかにしている。

### 2.2 LLMのバイアスと多視点生成

Benderら [8] と Navigliら [9] は、LLMが持つ社会的バイアスや単一的な記述傾向の危険性を指摘した。Liら [10] は、障害種別によってバリアの深刻度認識が異なることを実証的に示し、パーソナライズされた地図設計の必要性を論じている。さらに、Parkら [11] および Shanahanら [12] は、LLMを“ロールプレイエージェント”として活用し、多様な視点を模倣させる試みを報告している。これらの研究は、本研究が目指す多視点記述生成の理論的基盤を与えるものである。

<sup>1</sup> 日本大学文理学部

<sup>2</sup> 日本大学大学院総合基礎科学研究科

<sup>3</sup> ソフトバンク株式会社

a) miyata.akihiro@acm.org

### 3. 実験

#### 3.1 実験目的

本実験の目的は、各当事者の役割を与えた LLM 間で対話させた場合に、バリアに関する説明文がどのように生成されるかを明らかにすることである。具体的には、異なる移動困難者の視点に基づく発話と、専門家視点に基づく発話を組み合わせた対話を通じて説明文生成を行い、生成される記述の内容の違いや傾向を観察・整理する。これにより、複数の視点を対話的に組み合わせることが、バリアに関する多様な理解や表現を引き出す手段としてどのような可能性を持つかを検討する。

#### 3.2 実験用パイプライン

研究課題を達成するため、本論文では図 1 に示す実験用パイプラインを構築した。このパイプラインの主要モジュールは次のとおりである。

**Describer**：車椅子利用者、杖利用者、視覚障害者の介助者などの当事者の立場から、与えられた画像に写る歩道の問題点を描写するモジュールである。問題点の描写は、LLM (gpt-5-mini) に、各当事者の役割を指示するプロンプトと、各当事者の知識を与えることで実現する。

**Questioner**：都市空間のアクセシビリティの専門家の立場から、与えられた画像と Describer の描写に基づいて、Questioner が Describer に対してより厳密な描写を求める質問を作成するモジュールである。質問生成は、LLM (gpt-5-mini) に、専門家の役割を指示するプロンプトと、専門家の知識を与えることで実現した。

**Meta-Describer**：与えられた描写の集合を集約するモジュールである。Describer と Questioner が 1 組 (N=1) の場合は、Describer の描写をそのまま集約結果とする。Describer と Questioner が 2 組以上 (本稿では N=5) の場合は、複数の Describer による描写に共通する内容を集約結果とする。LLM (gpt-5.1) に上記作業を指示するプロンプトを与えることで実現する。加えて、画像についてバリアがどの程度問題であるのかを 3 段階でスコアリングも行う。

本実験においては、Describer の最初の描写 (1st Description) を、Monologue とし、その後の Questioner から質問を受け、Describer が出力しなおした描写 (2nd Description) を Dialogue とする。さらに、これらの結果を Meta-Describer に集約させた生成文をそれぞれ Mono-C と Dia-C とする。

Describer に与える各当事者知識 (図 1 中の First-Person Knowledge) と、Questioner に与える専門家知識 (図 1 中の Allly Expert Knowledge) は、対話型 AI (ChatGPT 5, Gemini 2.5 Pro) を用いて各当事者・専門家による事実

に基づく記述を 2025 年 10 月～11 月に Web から収集し、これを対話型 AI (Gemini 2.5 Pro) で要約したものをを用いた。

#### 3.3 実験手順

**Step 1**: 日本における、典型的な市街地の写真を何枚か収集した。

**Step 2**: 収集した画像を使って、図 1 に示すパイプラインを実装したプログラムを実行し、そのテキスト形式の生成文を確認、集計した。

**Step 3**: 生成文を一文ごとに分割し、Sentence-BERT で意味ベクトルに変換し、それらを比較するプログラムを実行した。

#### 3.4 出力結果の集計

本節では、各役割を与えた LLM による問題点を描写した生成文を集約し、その傾向と差異を分析する手法について述べる。生成文の焦点の違いを定量的に評価するために、意味的距離の分析を行った。

具体的には、各描写文を Sentence-BERT により埋め込みベクトルへ変換し、コサイン距離を算出した。同一の役割間 (例: 複数の“杖利用者”ロール間) と、異なる役割間 (例: “杖利用者”と“車椅子利用者”) の距離分布を比較することで、役割設定が出力文の意味的差異にどの程度影響を与えているかを評価した。

この分析により、Meta-Describer を通じた集約処理が、多様な視点を統合しつつも、役割間の特徴的差異を保持できているかを定量的に検証した。

### 4. 結果と考察

本章では、Meta-Describer による出力結果および各役割の描写内容を分析し、役割設定や対話構造が出力に与える影響について考察する。

#### 4.1 実際の出力の一例

図 2 に示す歩道画像に対して、LLM が出力した問題点の描写の例を次に示す。

- narrow\_sidewalk 3/3: 歩道の有効幅が非常に狭く (0.6~0.8 m 程度)、車いすの幅に合う連続した通行ルートがなく、車道に出ないと自力で通行することはほぼ不可能です。
- uncut\_curb 2/3: 車道との境に縁石が連続しており、近くに車いす用のスロープがないため、歩道から降りるには危険な段差を越えて車道に出るか、スロープを探して大きく迂回する必要があります。
- obstacle 3/3: 電柱とそのコンクリート基礎が歩行ラインの真上にあり、前方への進行をふさいでいるため、車道側へ迂回しなければならず、多くの利用者にとってこの歩道部分は通行不能です。

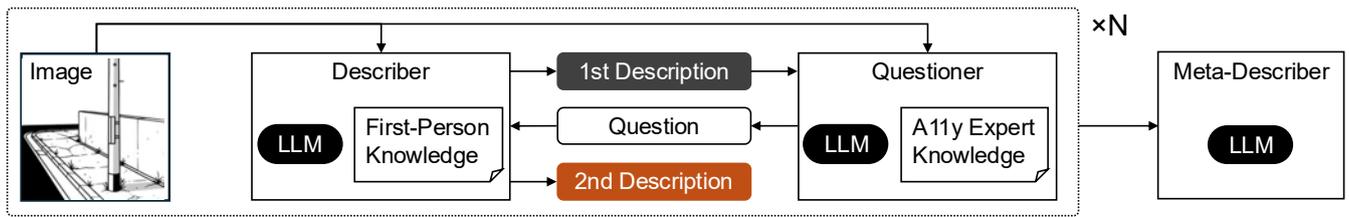


図 1 実験用パイプライン



図 2 実験で用いた画像の一例

- bad\_surface 2/3: 路面にはひび割れや穴、緩んだ補修材があり、走行抵抗や前輪キャストが挟まる危険が増えています、努力や介助があればなんとか通行は可能です。
- tactile\_paving 2/3: 建物側に細い点字ブロック帯があり、有効幅がさらに狭くなるうえ、連続した振動と不快感が生じて、方向転換や長距離の移動が難しくなります。
- step 1/3: 目に見える歩行ルート上には、はっきりした単独の段差はありません。
- stairs 1/3: 目に見える歩行ルート上には、階段はありません。
- steep\_slope 1/3: 歩行面はほぼ水平で、通行に影響するような急な傾斜はありません。
- other 3/3: 極端な狭さ、電柱などの障害物、悪い路面状態、縁石に切り下げがないことが重なり、この区間は多くの車いす利用者にとって実質的に使えず、歩行者全体にとっても非常に危険です。

生成文は、指摘する問題のタグ (bad\_surface や steep\_slope など)、その問題点がどのくらい深刻かを 3 段階評価、そしてその問題についての詳細な記述で構成されている。

車椅子利用者が直面しそうな問題についてはほぼ正確な指摘がされていることがわかる。また、図 2 内では存在しない、ただし、縁石の問題については uncut\_curb で事実誤認を含む描写があることがわかった。

#### 4.2 出力内容の傾向

Meta-Describer の集約結果を確認したところ、著者らが

生成文を確認したところ、多くの画像で“道が狭い”“電柱が邪魔”などの表現が頻出した。これらの語句は、全生成のうち約 3 割の文中に共通して現れており、屋外空間における代表的な問題点として LLM が一貫して注目していることが分かった。異なる役割を与えた場合でも、空間構造そのもの (例: 通路幅, 段差, 坂) に関する指摘が支配的であり、感覚的・体験的なバリア表現 (例: “出入口に入りにくい” など) は相対的に少なかった。この結果は、モデルが一般的な視覚的特徴に強く依存していることを示唆している。

また、生成文のスコアリングについても、Mono-C や Dia-C など Meta-Describer が集約した結果については、評価が厳しくなる例も確認された。

#### 4.3 意味的距離の比較

生成文の焦点の違いを定量的に評価するために意味的距離の分析を行った。具体的には、各描写文を Sentence-BERT により埋め込みベクトルへ変換し、コサイン距離を算出した。表 1 のように、同一の役割間 (例: 複数の“杖利用者” ロール間) と、異なる役割間 (例: “杖利用者” と “車椅子利用者”) の距離分布を比較することで、役割設定が出力文の意味的差異にどの程度影響を与えているかを評価した。この結果、同一役割間の出力文同士でも、異なる役割間でも文の意味的な距離が大きく変わらないことが判明した。この理由として、先に述べたような指摘の偏り (“道が狭い” という指摘が多いなど) が影響していると考えられる。

#### 5. おわりに

本稿では、画像内のバリアの問題を LLM に説明させるタスクにおいて、Monologue と Dialogue、またそれぞれの生成文を集約した結果を比較した。その結果、(1) 各役割でほぼ正確な記述が得られたこと、(2) Meta-Describer による集約により、画像内のバリアの評価が厳しくなるなど変化があることがわかった。今後の展望として、障害当事者と協働し、生成モデルの出力を検証・改善していきたい。

#### 参考文献

- [1] Biagi, L., Brovelli, M. A., Stucchi, L. et al.: Mapping the accessibility in openstreetmap: A comparison of different

表 1 役割間の意味ベクトルの距離

	blv_caregiver	cane_user	wheelchair_user
blv_caregiver	0.475	0.473	0.476
cane_user	0.473	0.476	0.476
wheelchair_user	0.476	0.476	0.462

techniques, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 43, No. B4, pp. 229–236 (2020).

- [2] Goodchild, M. F.: Citizens as sensors: the world of volunteered geography, *GeoJournal*, Vol. 69, No. 4, pp. 211–221 (2007).
- [3] Haklay, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets, *Environment and planning B: Planning and design*, Vol. 37, No. 4, pp. 682–703 (2010).
- [4] Cardonha, C., Gallo, D., Avegliono, P., Herrmann, R., Koch, F. and Borger, S.: A crowdsourcing platform for the construction of accessibility maps, *Proceedings of the 10th international cross-disciplinary conference on web accessibility*, pp. 1–4 (2013).
- [5] Mobasheri, A., Deister, J. and Dieterich, H.: Wheelmap: the wheelchair accessibility crowdsourcing platform, *Open Geospatial Data, Software and Standards*, Vol. 2, No. 1, p. 27 (2017).
- [6] Liu, H., Li, C., Wu, Q. and Lee, Y. J.: Visual instruction tuning, *Advances in neural information processing systems*, Vol. 36, pp. 34892–34916 (2023).
- [7] Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M. et al.: Flamingo: a visual language model for few-shot learning, *Advances in neural information processing systems*, Vol. 35, pp. 23716–23736 (2022).
- [8] Bender, E. M., Gebru, T., McMillan-Major, A. and Shmitchell, S.: On the dangers of stochastic parrots: Can language models be too big?, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623 (2021).
- [9] Navigli, R., Conia, S. and Ross, B.: Biases in large language models: origins, inventory, and discussion, *ACM Journal of Data and Information Quality*, Vol. 15, No. 2, pp. 1–21 (2023).
- [10] Li, C., Pang, R. Y., Labbé, D., Eisenberg, Y., Hosseini, M. and Froehlich, J. E.: Accessibility for Whom? Perceptions of Mobility Barriers Across Disability Groups and Implications for Designing Personalized Maps, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–19 (2025).
- [11] Park, J. S., O’Brien, J., Cai, C. J., Morris, M. R., Liang, P. and Bernstein, M. S.: Generative agents: Interactive simulacra of human behavior, *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22 (2023).
- [12] Shanahan, M., McDonell, K. and Reynolds, L.: Role play with large language models, *Nature*, Vol. 623, No. 7987, pp. 493–498 (2023).
- [13] Miyata, A.: Can Role-Based LLM Dialogue Improve Understanding of Accessibility Issues?, *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–5 (2025).
- [14] Kazai, G., Kamps, J. and Milic-Frayling, N.: An analysis of human factors and label accuracy in crowdsour-

ing relevance judgments, *Information retrieval*, Vol. 16, No. 2, pp. 138–178 (2013).