ASCII CHALLENGE —LLM は画家になれるか—

吉田遥音 *1　羽根田賢和 *1　斉藤いつみ 1,2　坂口慶祐 1,2
1 東北大学 2 理化学研究所
{yoshida.haruto.p1, haneda.kento.t6}@dc.tohoku.ac.jp

概要

アスキーアートはイラストや画像を文字で表現するテキストアートである．文字だけを用いて様々な表現を可能にするアスキーアートは現代社会で広く用いられている一方で，その作成は容易ではない．また既存の生成ツールは画像を機械的に変換するなど柔軟性が低い方法に限定されている．本研究では，自然言語からアスキーアートを生成する手段としての LLM・LVLM の利用可能性の検証を行った．結果として現行のモデルでは生成が困難だが，アスキーアートに特化したデータセットを用いて学習することで生成可能になる兆しが見えた．

1　はじめに

アスキーアートとは文字を用いて動物や人の表情などをグラフィカルに表現するテキストアートの一種である [1, 2, 3]．アスキーアート (以後 AA とする) はテキストのみで画像やイラストを代替し豊かな感情表現などを可能にするため，世界的に広く用いられている [4]．一方で読み手が AA の意図を汲み取れるように適切な文字の選択・配置を行うことは容易ではない．そのため AA を自動で生成するツールも多く提案されている [4, 5, 6, 7] が，これらのツールの多くは元となる画像を機械的に変換するものであり，描く対象の画像を用意しなくてはならないという点において柔軟性が低い．例えばまったく新しい架空の生物を描く場合や，AA の構図を変更したいといった場合，これらを高品質に素早く描くことは容易ではない．そのため自然言語に基づく AA の自動生成には一定の需要が存在すると考えられる．そこで本研究では，自然言語を入力とした AA の高品質な生成手段として，大規模言語モデル（LLM）や大規模視覚言語モデル（LVLM）が活用可能であるかを調査した．

* Equal Contribution

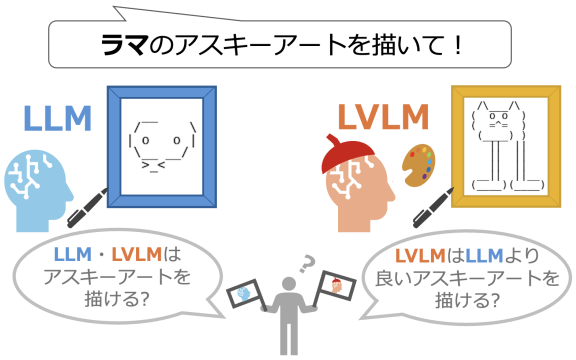

図 1　本研究の概要．自然言語を入力として LLM や LVLM でのアスキーアート（AA）の生成を試みる

図 1 に本研究の概要を示す．AA は文字のみによって描かれるため，テキストを出力する能力を持つ LLM は原理上，AA も生成可能であると考えられる．また，テキストと画像を入力として，画像についての高度な推論を実現している LVLM [8, 9, 10] は，画像を用いて学習されている．したがって純粋な LLM と比較して二次元的な情報処理能力があると考えられ，AA の生成能力が高い可能性がある．しかしながら，自然言語文ではない AA に対する LLM や LVLM の生成能力は未知数であり，この点において十分な検証が必要である．

本研究では，自然言語から高品質な AA を生成する手段として，LLM や LVLM が活用可能であるかを調査し，LLM と LVLM の間に生成における性能差が存在するかを検証した．実験は複数のモデルに同じテーマの AA を生成させ，その結果を人手評価することによって実施した．結果として，いずれのモデルも AA の生成手段としての実用性は低く，画像での学習の有無が生成能力に与える影響は小さいという示唆を得た．

2　関連研究

画像入力に基づく AA の自動生成　Chung ら [6] や O'Grady ら [7] は白黒の二値画像を元に AA を生

成する手法を提案している．また，Xu ら [4, 5] は二値に限らない自然画像から AA を生成する手法や，文字の濃淡ではなく輪郭や構造に着目して画像を AA に変換する手法を発表している．いずれも元となる画像を用意する必要があるという点において，本研究の対象とは異なる．

テキスト入力に基づく AA の生成　Zhang ら [11] は，LLM を用いてチャート図などを AA で表現することを試みている．Few-shot 学習により一定の成功を収めたものの，複雑な論理構造の表現には課題があり，LLM におけるテキストでの図形生成が容易でないことが確認されている．また上記の研究における調査対象は論理的な要素を持つ構造化された表現である．イラストなど，構造化されていない AA の生成を目指しているという点において，本研究は既存研究とは異なる特徴を持つ．

3　実験

各種 LLM，LVLM に AA を生成させ，結果を 5 段階で人手によって評価した．

3.1　モデル

生成用のモデルは計 8 種使用した．そのうちオープンアクセスなモデルが 6 種，クローズドアクセスなモデルが 2 種である．

オープンアクセスなモデルは，LLM と LVLM を 3 種ずつ用いた．LLM と LVLM の間の生成能力の差を検証するために，LLM は各モデルを構成要素に持つ学習済み LVLM が利用可能であることを基準に選定した．具体的には，Llama-3.1-8B-Instruct [8]，Llama-3.2-11B-Vision-Instruct [8]，Phi-3.5-mini-instruct [9]，Phi-3.5-vision-instruct [9]，Qwen2-7B-Instruct [10]，Qwen2-VL-7B-Instruct [10] を用いた．

クローズドアクセスなモデルとしては，GPT-4o [12] と，Claude 3.5 Sonnet[13] を利用した．

3.2　プロンプト

各モデルには「Draw *subject* as ASCII art.」というプロンプトを与えた．*subject* には手作業で選定した AA のテーマとなる 50 種類の単語が入る．

テーマ単語は，ASCII Art Arcive[1] というサイトをもとに選定した．ASCII Art Arcive では AA がカテゴリごとにまとめられており，これらのうち animals，nature，mythology，sports，food and drinks の 5 つを選び，それぞれに 10 のテーマ単語を用意した．多様な生成対象の検証のため，カテゴリは AA の登録数が多いものからバランスを考慮して選定した．

各カテゴリの単語群は，ASCII Art Archive 上にそれをモチーフとした AA が登録されていることが確認された 5 種の単語と，そうでない 5 種とで構成されている．例えば，animals カテゴリのテーマ単語は，a cat，an owl，a dog，a shark，a monkey の 5 単語と，a capybara，a llama，an ostrich，a tuna，a seal の 5 単語である．前半の 5 つは当該サイトに AA が登録されている単語であり，そのうち登録数の多いものを基本として選定した．後半の単語に関しては前半の単語と比較して知名度や性質に大きな差が出ないように留意して著者らが選定した．使用したテーマ単語の一覧は Appendix の表 3 を参照されたい．

3.3　人手評価

生成結果に対して 3 人の評価者による人手評価を行った．評価軸は**受容可能度**と**一致度**の 2 つである．受容可能度は「AA として解釈可能か」を意味し，一致度は「AA がお題のテーマ単語を表しているか」を意味する．これらの 2 軸に対して，それぞれ 5 段階の評価を実施した．

受容可能度では，生成結果が何らかの対象を描いた AA であるとして解釈可能なものであるかを評価した．評価者には題材となったテーマ単語は明らかにせず，生成結果のみを提示した．一般的な文章のみを出力している場合を最も低く評価し，全体を図形としてみた際にある特定の事物を描いたものだと断言可能であるような場合を最高評価とした．

一致度では，生成結果がテーマ単語を十分に表現したものであるかを評価した．評価者には生成結果とともに題材となったテーマ単語を提示した．一般的な文章のみの生成を最低評価とし，一目見てテーマ単語を表すものであると自信を持って解釈可能である結果を最も高く評価した．

各軸における 5 段階評価の詳細な基準についてはそれぞれ Appendix の A，B のとおりである．

4　結果・考察

4.1　現行のモデルでの AA 生成は困難

表 1 に各モデルに対する人手評価の結果を示す．なお 3 人の評価者間の Krippendorff の α 値 [14] は，受容可能度，一致度でそれぞれ 0.570 と 0.649 であっ

1　https://www.asciiart.eu/

表 1　各モデルの平均スコア

	受容可能度	一致度
Llama 3.1	3.07	2.13
Llama 3.2 vision	3.19	2.12
Phi 3.5 mini	2.51	2.15
Phi 3.5 vision	2.75	2.01
Qwen 2	2.83	2.05
Qwen 2 VL	2.66	2.05
GPT-4o	3.44	2.77
Claude 3.5	3.19	2.77
全体	2.96	2.26

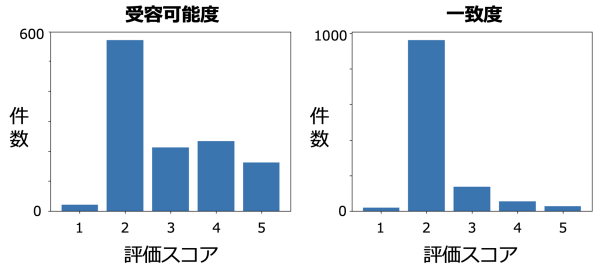

図 2　各評価軸での点数分布

た．受容可能度，一致度に対しての全体の平均スコアはそれぞれ 2.96 と 2.26 であった．クローズドアクセスモデルである GPT-4o，Claude 3.5 は，どちらも全体平均よりも高い平均スコアを記録し中でも GPT-4o はどちらの評価軸においても最も高く評価されていた．しかしその GPT-4o においてもスコアは決しては高いとは言えず，現行のモデルを AA の生成手段とすることは困難であると考えられる．特に一致度でのスコアは全体的に低く，特定の事物を対象とした AA の生成はモデルを問わず困難なタスクであるといえる．

図 2 に示した各評価軸での点数分布を示す．ここから評価者はほとんどの生成に対し 2 点を与えていることが確認できる．これはどちらの評価軸においても，生成が通常の文章ではなかった場合に与えられるスコアのうち最低の点数である．Krippendorff の α 値が低く，評価の難しいタスクであったことを差し引いても，モデルが実用に足るだけの AA を生成できていたとは言い難い．

一方で同じく図 2 から 1 点とされたケースはごく一部であり，通常のテキストのみでの生成がほとんどないことが読み取れる．図 3 に示した実際の生成例や，受容可能度のスコアが比較的高いことからも，**AA のようなもの**の生成は十分に行えていたといえる．このことから LLM はテキストという一次

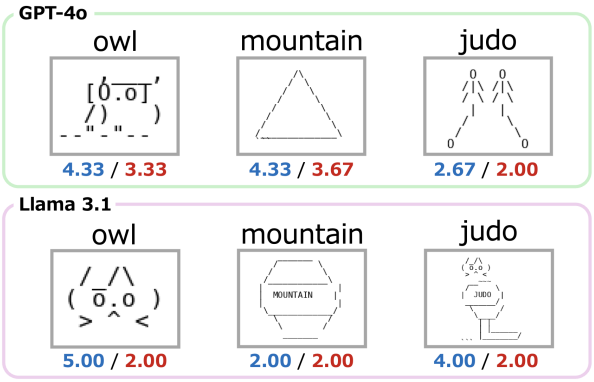

図 3　実際の生成例．3 人の評価者の平均スコアを受容可能度は青字で一致度は赤字で示した

表 2　単語カテゴリごとの平均スコア

	受容可能度	一致度
animals	3.78	2.43
nature	2.67	2.45
mythology	3.09	2.13
sports	2.36	2.03
food and drinks	2.88	2.24
全体	2.96	2.26

元的な学習データを元に，二次元的な表現能力を部分的には獲得していると考えられる．

4.2　LLM と LVLM の性能差

表 1 の結果から，LLM と LVLM の AA 生成能力に明確な差は見らなかった．したがって，画像による学習が AA の生成に何らかの効果をもたらしているとは考えにくい．画像での学習を上手く AA の生成に活用するためには，現行のモデルアーキテクチャや学習方法を改善する必要があると考えられる．

4.3　生成対象による難度差

テーマ単語のカテゴリによるスコアは表 2 にのようにまとめられた．受容可能度についてはカテゴリごとに明確なスコア差が見られ，一致度についてもスコア差がやや見られた．どちらの評価軸においても animals は高い評価を獲得しており，一方で sports をテーマにした生成は最低の評価を受けていた．特に受容可能度においては，両カテゴリの平均スコアに 1 以上の開きがあり，生成対象によって生成難易度に明確な差があることが示唆されている．これには主に二つの要因が影響していると考えられる．

一つ目は，事前学習データ中に存在する AA の影響である．Ascii Art Archive には，Cats というカテ

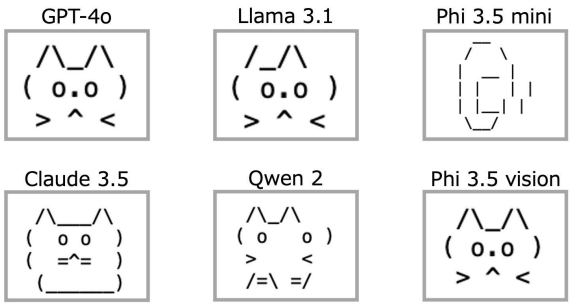

図 4　テーマ単語を与えない場合の生成結果（一部抜粋）．temperature などのパラメータはデフォルト値を使用

ゴリに 86 件もの AA が作品として登録されており，これは同サイトにおいて，一単語に対する登録件数として二番目に多いものである．このことから AA の描画対象として猫は比較的普遍性が高く，事前学習データに登場する頻度も高いと推測される．

図 4 に，今回用いたモデルに対してテーマを与えずに AA を生成させた際の結果を示した．多くのモデルが猫のような AA を生成する傾向を持つことが確認できる．これらは生成例の一部であるが，各モデルに対してそれぞれ 10 回，計 80 回行った生成のうち，実に 85%以上が猫のような AA であった．

この結果からモデルは AA という概念のうち登場頻度の高い猫の AA に関して，既に学習していると考えることができる．そのため猫の AA の生成が上手く，同じ animals というカテゴリに属し見た目が近い他の動物に対しても比較的良い生成を行うことができたと考えられる．

二つ目の要因として考えられるのが，テーマ単語の具体性・限定性による影響である．平均スコアが比較的高かった animals に含まれる単語は，a cat や an owl といった言葉であり，それらが指す対象は具体的かつ実体を持つものに限定される．そのため対象の輪郭を描くことさえできれば，受容可能度においては高く評価されやすい．一方で，sports カテゴリに含まれる単語は，basketball や soccer などであり，これらの単語が指す対象は曖昧性を孕む．例えば basketball という単語は，競技そのものを指す場合とボールを指す場合とが考えられ具体性が低い．さらに競技を指していた場合，プレイする選手のみを描くのか，コートの様子まで描くのかなど自由度が高く，何をどこまで描くかが不明瞭である．実際に，Ascii Art Archive には，選手とボールが描かれたものから，リングのみが描かれたものまで様々な AA が登録されている．こうした多様性により，学習に必要な AA の種類や量が多く，追加学習をしていない現行のモデルでの生成が難しかったと考えられる．

4.4　学習データの暗記の可能性

Ascii Art Archive に存在する単語群の平均スコアは受容可能度で 3.09，一致度で 2.36 と全体平均よりも若干スコアが高かった．一方，そうでない単語群の平均スコアは受容可能度，一致度で 2.82，2.15 となり，全体平均よりも低かった．このことから，学習データを丸暗記した生成が行われている可能性が考えられる．実際，図 4 で示した例の中に，Ascii Art Archive に登録された AA とほぼ同一の生成が確認されており，元の AA の作者への権利侵害が懸念される．一方で，大量のデータを学習に利用すれば，一般的な AA の生成が可能になることを示唆しているとも考えられ，今後の発展性は十分に期待できる．

5　おわりに

本研究ではアスキーアートについて，その生成手段として LLM を用いることができるかを調査し，LVLM との比較検証により画像での学習が AA の生成能力に与える影響を調査した．結果としていずれのモデルでも実用に足る品質での AA の生成は困難であることがわかり，LVLM の優位性は確認されなかった．また生成対象により難易度に差があることが考えられ，学習データを丸暗記した生成を行っている可能性を示唆する結果が得られた．

今後の課題としては，AA に対する明快で実用的な評価指標の確立と，生成能力の強化の二点が挙げられる．本研究では 5 段階での人手評価を行ったが現状の評価軸では評価者間の一致を得ることが難しい．そのため説得力の高い評価を行うことが困難であり，新たな評価指標を確立する必要性がある．生成能力の強化も今後の目標の一つである．本研究を通して LLM による AA の生成は発展途上であることが確認された．一方で，Web 上に多くのデータが存在するテーマでは高品質な AA が生成されている例もあり，AA に特化したデータセットを用いて学習することで，LLM は AA を生成しうると考えられる．LLM が言葉の絵を描く日を目指して，挑戦を続けていきたい．

謝辞

本研究は JSPS 科研費 JP21K21343 の助成を受けたものです．人手評価に際しご協力いただきました株式会社バオバブ様（https://baobab-trees.com/）に深く感謝を申し上げます．また研究を進めるにあたりご協力いただいた東北大学 Tohoku NLP グループの皆様へ感謝いたします．

参考文献

[1] Anders Carlsson. Beyond encoding: A critical look at the terminology of text graphics. 2017.

[2] Karin Wagner. **From ASCII Art to Comic Sans: Typography and Popular Culture in the Digital Age**. The MIT Press, 09 2023.

[3] Anders Carlsson and A Bill Miller. Future potentials for ascii art cac. 3, paris, france. **In Postdigital art-Proceedings of the 3rd computer art congress**, 2012.

[4] Xuemiao Xu, Linyuan Zhong, Minshan Xie, Xueting Liu, Jing Qin, and Tien-Tsin Wong. Ascii art synthesis from natural photographs. **IEEE Transactions on Visualization and Computer Graphics**, Vol. 23, No. 8, pp. 1910–1923, 2017.

[5] Xuemiao Xu, Linling Zhang, and Tien-Tsin Wong. Structure-based ascii art. **ACM Trans. Graph.**, Vol. 29, No. 4, July 2010.

[6] Moonjun Chung and Taesoo Kwon. Fast text placement scheme for ascii art synthesis. **IEEE Access**, Vol. 10, pp. 40677–40686, 2022.

[7] Paul D. O'Grady and Scott T. Rickard. Automatic ascii art conversion of binary images using non-negative constraints. In **IET Irish Signals and Systems Conference (ISSC 2008)**, pp. 186–191, 2008.

[8] AI @ Meta Llama Team. The llama 3 herd of models, 2024.

[9] Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, Alon Benhaim, Misha Bilenko, Johan Bjorck, Sébastien Bubeck, Martin Cai, Qin Cai, Vishrav Chaudhary, Dong Chen, Dongdong Chen, Weizhu Chen, Yen-Chun Chen, Yi-Ling Chen, Hao Cheng, Parul Chopra, Xiyang Dai, Matthew Dixon, Ronen Eldan, Victor Fragoso, Jianfeng Gao, Mei Gao, Min Gao, Amit Garg, Allie Del Giorno, Abhishek Goswami, Suriya Gunasekar, Emman Haider, Junheng Hao, Russell J. Hewett, Wenxiang Hu, Jamie Huynh, Dan Iter, Sam Ade Jacobs, Mojan Javaheripi, Xin Jin, Nikos Karampatziakis, Piero Kauffmann, Mahoud Khademi, Dongwoo Kim, Young Jin Kim, Lev Kurilenko, James R. Lee, Yin Tat Lee, Yuanzhi Li, Yunsheng Li, Chen Liang, Lars Liden, Xihui Lin, Zeqi Lin, Ce Liu, Liyuan Liu, Mengchen Liu, Weishung Liu, Xiaodong Liu, Chong Luo, Piyush Madan, Ali Mahmoudzadeh, David Majercak, Matt Mazzola, Caio César Teodoro Mendes, Arindam Mitra, Hardik Modi, Anh Nguyen, Brandon Norick, Barun Patra, Daniel Perez-Becker, Thomas Portet, Reid Pryzant, Heyang Qin, Marko Radmilac, Liliang Ren, Gustavo de Rosa, Corby Rosset, Sambudha Roy, Olatunji Ruwase, Olli Saarikivi, Amin Saied, Adil Salim, Michael Santacroce, Shital Shah, Ning Shang, Hiteshi Sharma, Yelong Shen, Swadheen Shukla, Xia Song, Masahiro Tanaka, Andrea Tupini, Praneetha Vaddamanu, Chunyu Wang, Guanhua Wang, Lijuan Wang, Shuohang Wang, Xin Wang, Yu Wang, Rachel Ward, Wen Wen, Philipp Witte, Haiping Wu, Xiaoxia Wu, Michael Wyatt, Bin Xiao, Can Xu, Jiahang Xu, Weijian Xu, Jilong Xue, Sonali Yadav, Fan Yang, Jianwei Yang, Yifan Yang, Ziyi Yang, Donghan Yu, Lu Yuan, Chenruidong Zhang, Cyril Zhang, Jianwen Zhang, Li Lyna Zhang, Yi Zhang, Yue Zhang, Yunan Zhang, and Xiren Zhou. Phi-3 technical report: A highly capable language model locally on your phone, 2024.

[10] Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution, 2024.

[11] Yimin Zhang and Mario de Sousa. Exploring llm support for generating iec 61131-3 graphic language programs, 2024.

[12] OpenAI. Hello gpt-4o, 2024.

[13] Anthropic. The claude 3 model family: Opus, sonnet, haiku, 2024.

[14] Klaus Krippendorff. Computing krippendorff's alpha-reliability. 2011.

表 3　カテゴリごとのテーマ単語一覧

カテゴリ名	AAA にある単語	AAA にない単語
animals	a cat, an owl, a dog, a shark, a monkey	a capybara, a llama, an ostrich, a tuna a seal
nature	the sun, an island, snow, rain, a mountain	a river, a volcano, a forest, a cave, fog
mythology	a dragon, a mermaid, a devil, an unicorn, a fairy	a cerberus, a nessie, a zombie, a medusa, a minotaur
sports	billiards, ice hockey, basketball, bowling, soccer	table tennis, rugby, badminton, pole vault, judo
food and drinks	beer, chocolate, ice cream, pizza, an apple	coke, a donut, an eclair, a hamburger, a strawberry

A　受容可能度の評価基準

受容可能度における評価スコアと評価基準の対応は以下のとおりである．また，評価スコアが 1 と 5 の例を図 5 に示す．

1. 文章のみ出力している
2. 全体を図形として見た場合に特定のある対象としての解釈が難しく，部分的に解釈することも困難
3. 全体を図形としてみた場合には何らかの事物としての解釈が難しいが，部分的には何らかの解釈が可能
4. 全体を図形としてみた場合，何らかの事物を描いたものであると解釈可能
5. 全体を図形としてみた時に，ある特定の事物を描いたものであると断言可能

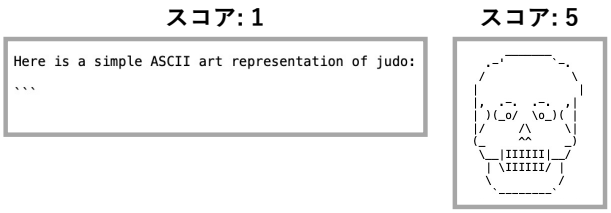

図 5　受容可能度においてスコアが 1 と 5 の例

B　一致度の評価基準

一致度における評価スコアと評価基準の対応は以下のとおりである．また，評価スコアが 1 と 5 の例を図 6 に示す．

1. 文章のみ出力している
2. 入力として与えた単語を知っていても，それとして解釈することが困難
3. 一見するとそれとしての解釈は困難であるが，入力として与えた単語を知っている状態であれば，それとしての解釈の可能性を見出しうる．
4. 一目見て入力として与えた単語としても解釈できるが，それ以外の単語としての解釈の余地が多分にある
5. 一目見て入力した単語を表すものであると自信をもって解釈可能である

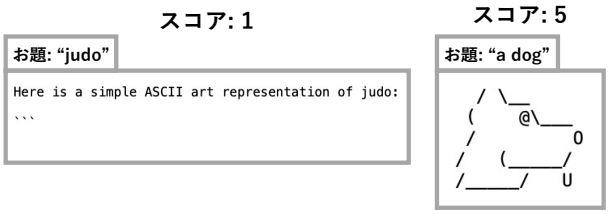

図 6　一致度においてスコアが 1 と 5 の例

C　使用したモデル

使用したモデルの一覧を表 4 に示す．

表 4　使用したモデル一覧

タイプ	モデル
LLM	Llama-3.1-8B-Instruct [8]
	Llama-3.2-11B-Vision-Instruct [8]
	Phi-3.5-mini-instruct [9]
LVLM	Phi-3.5-vision-instruct [9]
	Qwen2-7B-Instruct [10]
	Qwen2-VL-7B-Instruct [10]