報告書

1 今週の進捗

- llama.cpp の試行および評価
- Gemini と Yomitoku の OCR 精度比較

2 llama.cpp の試行

llama.cpp を用いた各種モデルの動作検証をした.

2.1 マルチモーダルモデルによる OCR

マルチモーダルに対応したモデルを用いて OCR (光学文字認識) を試行した. 使用した画像説明に特化したモデルでは, OCR 性能は限定的であった. 別途, OCR 性能が高いとされる ”gemma-3” モデルについて検証した. その結果, チャット形式での対話実行では良好な文字認識結果を示したが, プログラムコードを介した実行では再現性が得られなかった. このため, 実装コードの再検討が必要であると考えられる. 以下にいくつかのモデルにおける OCR の試行結果を示す.

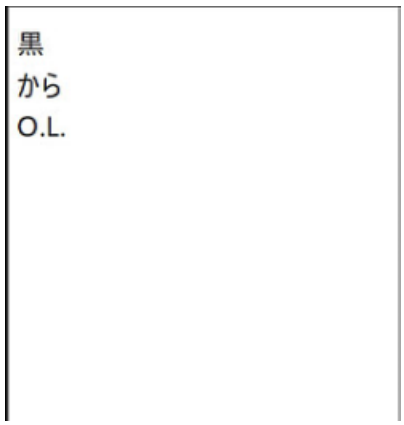


図 1: 入力画像

2.1.1 ”llava-v1.5-7b-Q2_K” による試行例

出力テキスト:

```
--- Assistant’s Response 1 ---
オルコン

--- Assistant’s Response 2 ---
10000000000000000000000000000000
```

2.1.2 ”llava-v1.6-34b” による試行例

出力テキスト:

```
--- Assistant's Response ---
.....
```

2.2 テキスト翻訳モデルの調査

テキスト入力による翻訳モデルの性能調査も実施した. 一般的な文章においては一定の翻訳精度が確認された. しかし, 固有名詞や特定の専門用語, あるいは文脈依存性の高い表現に関しては, 翻訳結果の正確性に課題が見られる場合があった.

翻訳例 1: 一般的な文章

```
Input Japanese Text:
こんにちは、世界！今日は良い天気ですね。

Translated Text (English):
Hello, world! Today is a nice day, isn't it?
```

翻訳例 2: 固有名詞を含む文章

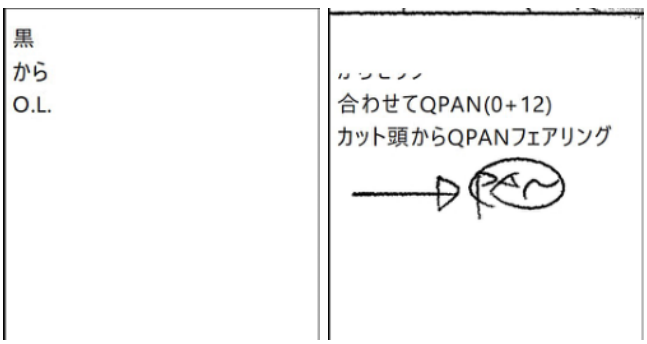
```
Input Japanese Text:
木剣のヨリから

Translated Text (English):
From the Yorikuni of the Wooden Sword
```

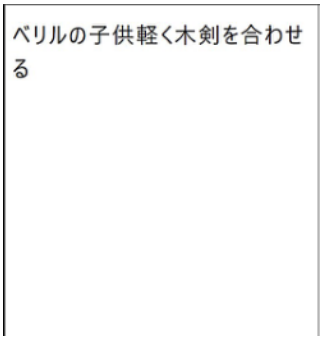
3 OCR 精度比較: Gemini vs. Yomitoku

Google の Gemini モデルと Yomitoku の OCR 精度について比較検証をした. 本比較では, 目視による事前確認において Gemini の方がより正確に文字認識ができていたため, Gemini の OCR 結果を正解データとして採用した. Gemini の結果においても, 空白の不適切な挿入などにより意図しない不正解と判定されるケースが一部存在した. Gemini と Yomitoku の OCR 比較の正解率は 68.7% であった.

3.1 比較に使用した入力画像例



(a) page002_row1 (b) page007_row2



(c) page014_row1

(d) page030_row5

3.2 比較結果

Gemini の出力を正解とした場合の Yomitoku の認識結果について, サンプルを以下の表に示す.

比較項目	Gemini による認識 (正解基準)	Yomitoku による認識
page002_row1	黒@から@O.L.	から@O.L.
page007_row2	ソソソソ@合わせて QPAN(0+12)@ カット頭から QPAN フェアリング@→ PAN	合わせて QPAN(0+12)@ カット頭から QPAN フェアリング
page014_row1	ベリルの子供軽く木剣を合わせ@る	ベリルの子供軽く木剣を合わせ
page030_row5	- +F (F)@follow back@ ベリル窓外をぼんやり眺めながら	follow back@ ベリル窓外をぼんやり眺めながら

4 今後の課題

- ”gemma-3” モデルのプログラムコード経由での実行における動作の安定化および精度改善
- Mixture of Experts (MoE) アーキテクチャを採用した OCR モデルに関する調査

参考文献