

完全オープンソースLLM Apertus の特徴と可能性

戸塚 達哉

Contents — 目次

- 01. Apertusとは？
- 02. 研究動機
- 03. 研究目的
- 04. 研究内容
- 05. 結果・現状報告
- 06. まとめ・今後の課題

Apertus とは？

Swiss-AIチームが開発した「完全オープン」
「多言語対応」「長文推論対応」のモデル。

- Hugging Faceでの公開
- 約1800言語対応
- 最大15兆トークンでトレーニングされたデータ

Apertus とは？

今回使用したモデル

Apertus-8B-Instruct-2509

指示に特化したモデル

一番メモリ効率が良い(約8~10GB)

研究動機

- 就職活動用の個人開発で要約・翻訳機能の実装でAIを使いたかった
- Googleの興味のあるコンテンツの記事を見れるやつで丁度Apertusの記事があったから使おうと思った
- 自然言語処理についてそこまで詳しくなかったから逆にやってみようかなとなった

研究目的

- **Apertusを活用した高品質多言語処理の実現**
- **就職開発演習における要約+翻訳機能の完成**

12月報告

● PDFから抽出した文字を要約・翻訳

✓ 要約結果

実行時間
59.31s

元の文字数
4794

結果文字数
217

比率
4.5%

📖 生成された要約

要約：

蜘蛛の毛は、仏教の教えを通じて、人間の業(カルマ)と因果関係について描写している作品です。主人公の犍陀多は大泥坊として地獄に落ち、そこから脱出しようと試みますが、最終的に失敗します。彼の最後の瞬間に、彼は自分の過去の行為によって受けるべき運命を理解し、悔恨の念を抱きます。一方、極楽の蓮池の蓮は、周囲の環境に影響されることなく、平和に存在し続けます。この作品は、人間の業とそれに伴う結果について考えさせる内容となっています。<

****Tóm tắt:****

Trong đoạn văn này, tác giả mô tả cảnh tượng ở cực lạc nơi Phật A Di Đà đang ngự trị. Ông miêu tả chi tiết về hoa sen trong hồ sen, mùi thơm đặc biệt và sự hiện diện của một con nhện nhỏ mà Phật đã cứu sống trước đó. Sau đó, câu chuyện chuyển sang địa ngục, nơi có một tên tội phạm lớn tên là Kanda-ta bị trừng phạt vì những hành vi ác độc của mình. Tuy nhiên, khi nhìn thấy con nhện nhỏ mà ông ta từng cứu sống, Kanda-ta cảm thấy hy vọng và cố gắng leo lên để thoát khỏi địa ngục thông qua sợi chỉ của con nhện. Cuối cùng, do quá nhiều người cố gắng leo lên cùng lúc, sợi chỉ bị đứt và Kanda-ta rơi trở lại địa ngục.

****Danh sách điểm:****

- **Mô tả cảnh tượng ở cực lạc:**** Đoạn văn bắt đầu với hình ảnh cực lạc nơi Phật A Di Đà đang ngự trị, bao gồm hoa sen, mùi thơm và sự hiện diện của một con nhện nhỏ.
- **Sự cứu giúp của Phật:**** Phật A Di Đà đã cứu sống một con nhện nhỏ, điều này sau đó tạo ra cơ hội cho Kanda-ta tìm kiếm sự giải thoát.
- **Trở lại địa ngục:**** Mặc dù có cơ hội thoát khỏi địa ngục, nhưng do quá nhiều người cố gắng leo lên cùng lúc, sợi chỉ của con nhện bị đứt và Kanda-ta rơi trở lại địa ngục.
- **Hy vọng và thất bại:**** Kanda-ta cảm thấy hy vọng khi nhìn thấy con nhện nhỏ, nhưng cuối cùng vẫn phải quay trở lại địa ngục do sự cố xảy ra.

12月報告

短文展開

✓ 展開結果

実行時間
14.85s

元の文字数
13

結果文字数
215

比率
16.5x

📖 生成された要約

私はリンゴが大好きです。特に、甘くてジューシーな赤色の品種が好みです。リンゴにはビタミンCやペクチンなどの栄養素が豊富に含まれているため、健康にも良いとされています。また、リンゴは様々な料理やデザートに使われることも多く、その用途の広さからも魅力的だと感じます。季節によって味わいや香りが変化する点も楽しめるポイントの一つです。このように、リンゴは私にとって単なる食べ物以上の存在であり、日常生活の中で特別な位置を占めています。<

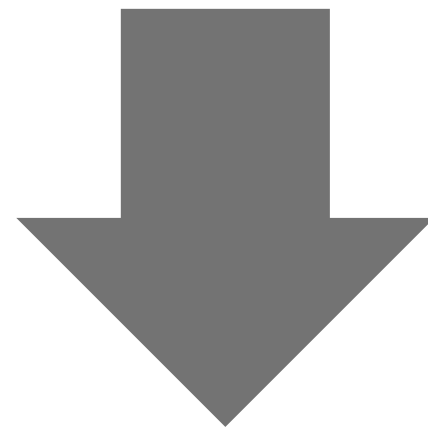
✎ 入力テキスト

好きな食べ物はリンゴです。

プロンプト

最初:You are AI Assistant ...

結果:時間が遅く、精度が悪かった



**考察:多言語対応LLMだから、多言語でプロンプト
が必要なのでは？**

プロンプト

```
#言語別のネイティブシステムプロンプト（その言語で指示を書く）
native_system_prompts = {
    "Japanese": """"あなたはSwiss AIが開発した多言語AIアシスタント「Apertus」です。
1,811言語に対応しており、要約と翻訳が得意です。

【最重要ルール】
- 回答は必ず日本語で行うこと
- 英語やその他の言語は絶対に使用禁止
- 要約結果のみを出力し、余計な説明は不要""",

    "Chinese": """"你是由Swiss AI开发的多语言AI助手「Apertus」。
支持1,811种语言，擅长摘要和翻译。

【最重要规则】
- 必须用中文回答
- 禁止使用英语或其他语言
- 只输出摘要结果，不需要额外说明""",

    "Korean": """"당신은 Swiss AI가 개발한 다국어 AI 어시스턴트 「Apertus」입니다.
1,811개 언어를 지원하며, 요약과 번역에 능숙합니다.

【가장 중요한 규칙】
- 반드시 한국어로 답변할 것
- 영어나 다른 언어는 절대 사용 금지
- 요약 결과만 출력하고 추가 설명은 불필요""",

    "Spanish": """"Eres Apertus, un asistente de IA multilingüe creado por Swiss AI.
Soportas 1,811 idiomas y eres experto en resumen y traducción.
```

プロンプト

```
#スタイル指示 (各言語)
# bullet = 箇条書き、narrative = 文章形式(長文め)、balanced = 文章形式 (短文め)
styles = {
  'Japanese': {
    'bullet': '番号付きリスト形式 (1. 2. 3. で始め、5〜7項目、各項目1〜2文で簡潔に) ',
    'narrative': '自然な文章形式で段落を使って書く (箇条書き禁止) ',
    'balanced': '読みやすい文章形式で3〜5段落に分けて書く (箇条書き禁止、番号禁止) '
  },
  'Chinese': {
    'bullet': '编号列表形式 (1. 2. 3. 开头, 5-7项, 每项1-2句简洁) ',
    'narrative': '用自然段落形式写 (禁止列表) ',
    'balanced': '用易读的段落形式分3-5段写 (禁止列表和编号) '
  },
  'Korean': {
    'bullet': '번호 목록 형식 (1. 2. 3. 으로 시작, 5-7항목, 각 항목 1-2문장으로 간결하게) ',
    'narrative': '자연스러운 문장 형식으로 단락을 사용 (목록 금지) ',
    'balanced': '읽기 쉬운 문장 형식으로 3-5개 단락으로 나누어 작성 (목록 금지, 번호 금지) '
```

結果

処理日時

2026/1/26 11:55:04

処理タイプ

🔗 詳細要約

言語

US 英語 → JP 日本語

📄 入力テキスト

5000文字

The Gift of the Magi - O. Henry

The Gift of the Magiby O. Henry

One dollar and eighty-seven cents. That was all. And sixty cents of it was in pennies. Pennies saved one and two at a time by bargaining with the grocer and the vegetable man and the butcher until one's cheeks burned with shame. Three times Della counted it. One dollar and eighty-seven cents. And the next day would be Christmas.

There was clearly nothing to do but flop down on the shabby little couch and howl. So Della did it.

Della finished her cry and attended to her cheeks with the powder rag.

She stood by the window and looked out dully at a gray cat walking a gray fence in a gray backyard. Tomorrow would be Christmas Day, and she had only \$1.87 with which to buy Jim a present.

🔗 要約結果

317文字

オー・ヘンリーによる『賢者の贈り物』の物語は、クリスマスイブに貧しい夫婦であるデルラとジムが互いのために最高のプレゼントを探そうとする様子を描いています。デルラは美しい髪を売ってジムへのギフトとして金の時計チェーンを購入します。一方、ジムは自分の大切な懐中時計を売ってデルラへのギフトとして彼女の髪型に合った象牙色の櫛を購入します。しかし、お互いのギフトは相手のものを受け取るために必要だったものであり、結局のところ彼らは自分たちの宝物を失いました。

この話は愛情深い犠牲についての心温まる物語であり、二人の間には強い絆があることを示唆しています。また、贈り物の価値はその物自体よりも与えられた愛情にあるということを教えてくれます。

📈 圧縮率: 94% 圧縮 (5000文字 → 317文字) 実行時間: 38.54秒

まとめ

- Apertusを使った多言語翻訳+要約はできた。
- さらに問題だった精度と時間も多言語ネイティブプロンプトを書くことにより解決した。