

## 報告書

### 1 今週の進捗

- アニメミームデータセットの作成

### 2 アニメミームデータセット

大規模言語モデル (LLM) にアニメ文化の深い文脈やユーモアを学習させることを目的とし、アニメ由来のインターネットミームを集約したデータセットを構築している。しかし、データセットの作成過程でいくつかの課題に直面している。

まず、「アニメミームとは何か」という定義の問題がある。当初はアニメキャラクターの有名なセリフを収集していたが、探していくうちに、セリフそのものではなく視聴者の反応や、作品から派生したネット文化など、その境界が曖昧になってきた。

また、以下のようなプロセスでデータセットを作成している。最初に、OpenAI 社の GPT-5 や Google 社の Gemini 2.5 Pro といった LLM を用いて、アニメミームの候補リストを網羅的に生成させる。次に、そのリストを人間が 1 つずつ検証する。検証作業は、インターネット上の百科事典サイト (ニコニコ大百科、ピクシブ百科事典など) での調査や、アニメに詳しい人に聞き、そのミームが実際に通用するかどうかを判断している。

#### 2.1 データセットの構造

現在構築中のデータセットは、1 つのミームにつき以下の 4 つの要素で構成されている。表 1 にデータセットの構造例を示す。phrase はミームのテキスト、source は元ネタの作品名、character は発言者、category はミームの性質を示す分類である。

表 1: データセットの構造例

phrase	source	character	category
お前も ○○ にならないか？	鬼滅の刃	猗窩座	anime
○○ よ 死になさい シンジ君	新世紀エヴァンゲリオン	桂木みさと	anime
おそろしく速い手刀 オレでなきゃ見逃しちゃうね	HUNTER × HUNTER	モブ	anime
だが断る	ジョジョの奇妙な冒険	岸辺露伴	anime
ある日突然 - 紅葉兄ちゃんのコトが大好きなんです…	シスタープリンセス	?	anime
尊い	アニメファン文化	視聴者	concept

#### 2.2 データセットの特性と課題

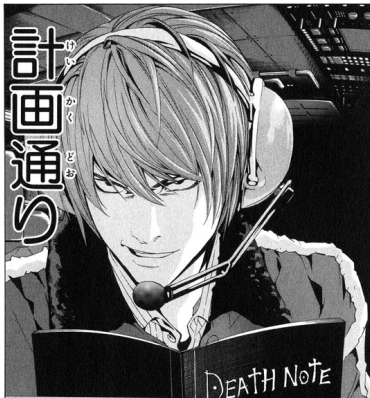
収集を進める中で、アニメミームが持ついくつかの特性が見えてきた。データセットとして特に重視しているのは、「お前も ○○ にならないか？」のように、構文として穴埋め形式で応用できるかどうかという点である。

一方で、収集対象となるフレーズには、非常に長いものも存在する。例えば、『機動武闘伝 G ガンダム』の「俺のこの手が真っ赤に燃える！お前を倒せと輝き叫ぶ！」といった長台詞は、どこまでを1つのフレーズとして切り取るべきか判断が難しい。

また、全てのミームがキャラクターのセリフに由来するわけではないことも興味深い点である。表1の「尊い」のように、キャラクターのやり取りや関係性を見た視聴者の感情表現が、ひとつのミームとして定着しているケースも少なくない。

さらに、ミームにはテキスト情報だけでは完結しない、マルチモーダルなものが多数存在するという課題もある。本データセットは現在テキストのみを対象としているが、将来的には以下のような画像ベースのミームも考慮に入れる必要がある。

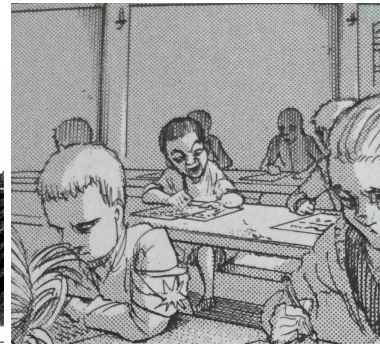
- 画像とそのままのテキスト: 特定のシーンのスクリーンショットとセリフ
- 画像と適当なテキスト: いわゆる「コラ画像」
- 画像のみ: 特定のポーズや表情だけで意味が通じるもの



(a) Death Note1



(b) Death Note2



(c) 進撃の巨人 1

### 3 今後の課題

- アニメミームデータセットの作成
- SRAG の実装

### 参考文献