



## I My Hero Academia - Magic: The Gathering Set

### 1 Task Overview

Generative AI has shown remarkable capabilities in creative domains. This project applies these techniques to the domain of game design, specifically for the collectible card game *Magic: The Gathering* (MTG). The goal is to leverage a pre-trained Transformer model to generate a complete, novel MTG card set based on the thematic universe of the anime "My Hero Academia". This involves not only generating grammatically correct card text but also ensuring thematic consistency, adherence to game mechanics, and appropriate statistical distributions of card types, colors, and rarities.

### 2 Objective

The main objective is to fine-tune a decoder-only Transformer model on a dataset of existing MTG cards to learn their structure and language. The fine-tuned model will then be used to generate a new set of cards that captures the essence of "My Hero Academia" characters and abilities ("Quirks").



Figure 1: Example of the expected output (The illustration generation is not required).

### Sub-Objectives

- **Acquire and Preprocess Data:** Download a comprehensive dataset of MTG cards and format it into a structured text file suitable for training.
- **Thematic Mapping:** Develop a clear mapping between "My Hero Academia" concepts (characters, factions, powers) and the five colors and mechanics of MTG.



- **Fine-Tune a Transformer Model:** Implement a training pipeline to fine-tune a pre-trained Decoder-only model on the expected MTG card dataset.
- **Conditional Generation:** Develop a strategy to prompt the model to generate specific cards based on the thematic mapping (e.g., generate a legendary creature card for "All Might").
- **Curate the Final Set:** Generate a large number of cards and then curate a final, balanced set that respects the typical statistical distributions of rarity, color, and card type found in a standard MTG expansion.

### 3 Magic the Gathering Game: an Overview

*Magic: The Gathering* (MTG) is a popular collectible card game created by mathematician Richard Garfield and published by Wizards of the Coast. The game was first released in 1993 and has since gained a massive following with a rich history of expansions and sets. In MTG, players take on the roles of powerful wizards, known as "planeswalkers," and use their decks of magical cards to cast spells, summon creatures, and engage in battles with their opponents.



(a) The most expensive MTG card of all times



(b) The most valuable card of *Wilds of Eldraine* set

Figure 2: Two *Magic the Gathering* cards

Here's a detailed description of the game, the structure of the cards, and the Scryfall notation:

#### Game Overview

- **Objective:** The primary goal of MTG is to reduce your opponent's life total to zero. Players do this by casting spells and summoning creatures to attack their opponents.
- **Deck Construction:** Each player builds a deck of at least 60 cards. The deck is composed of spells, creatures, lands, and other card types. A player may also have a sideboard with additional cards to swap in between games in a match.



- **Mana:** Players use "mana" to cast spells and summon creatures. Mana is generated by playing land cards, which are a fundamental part of a player's deck. There are five colors of mana (white, blue, black, red, and green), and some cards require specific colors of mana to be cast.

## Card Structure

MTG cards have a standard layout with various components:



Figure 3: Caption

- **Card Name:** The name of the card, which describes its identity and theme.
- **Mana Cost:** The mana cost is located in the upper-right corner of the card. It indicates the amount and type of mana required to cast the spell or summon the creature.
- **Card Type:** The card type (e.g., Creature, Instant, Sorcery, Enchantment, Artifact, Planeswalker) appears below the card name and describes the card's function.
- **Rules Text:** The rules text provides the card's abilities and how it functions (**Oracle text**). It often contains **flavor text** for thematic context.
- **Power and Toughness:** For creatures, the power and toughness are shown as a pair of numbers (e.g., "2/2"). The power represents the creature's attacking strength, and the toughness represents its ability to withstand damage.
- **Expansion Symbol:** This symbol, located on the center-right of the card, indicates the set from which the card originates.



- **Illustration:** MTG cards feature unique artwork to enhance the game's aesthetics.
- **Collector Number:** This number identifies the card's position within its set.

## 4 Dataset

You can download datasets containing of card descriptions and images from scryfall:  
Scryfall data download: <https://scryfall.com/docs/api/bulk-data>

### Scryfall Notation

Scryfall is a popular website and database for searching and exploring Magic: The Gathering cards. It uses a specific notation to describe cards, allowing for precise searches and referencing. Here is a basic breakdown of Scryfall notation (**pay attention:** the dataset notation is slightly different from the API here provided):

- **Card Names:** Card names are typically written within double curly braces, e.g., Lightning Bolt.
- **Mana Symbols:** Scryfall uses X for generic mana, W for white mana, U for blue mana, B for black mana, R for red mana, and G for green mana.
- **Card Types:** Card types are represented in lowercase, e.g., "instant," "creature," "enchantment."
- **Set Codes:** Sets are referred to by their official abbreviations, e.g., "DOM" for Dominaria, "M20" for Core Set 2020.
- **Rarity:** Rarity is indicated as "common," "uncommon," "rare," and "mythic."
- **Color Identity:** Color identity is represented as "C" for colorless, "W" for white, "U" for blue, "B" for black, "R" for red, and "G" for green.

## 5 Data Preprocessing

Load the JSON dataset and parse it to create a single, large text file. Each card should be converted into a structured string format, delimited by special tokens. This format will be the basis for the model's learning.

```
Example Format:  
<|startofcard|>  
name: Shivan Dragon  
cost: {4}{R}{R}  
type: Creature | Dragon  
rarity: rare  
text: Flying\n{R}: Shivan Dragon gets +1/+0 until end of turn.  
pt: 5/5  
<|endofcard|>
```

This structured text must then be tokenized using the pre-trained model's tokenizer.



## 6 Model Architecture

Students must implement a fine-tuning pipeline for a **decoder-only Transformer model**.

- **Core Model:** A pre-trained decoder-only model is required.
- **Task:** The model should be trained on a causal language modeling objective (i.e., predicting the next token in the sequence).

## 7 Hints

- Start by fine-tuning on a smaller subset of the data to ensure your pipeline works correctly.
- The quality of your structured text format in preprocessing is critical to the model's success.
- Consider using online resources for access to free GPU resources for fine-tuning.

## 8 Training and Evaluation

Train the model until the validation loss plateaus. Evaluation for this task is primarily qualitative.

- **Quantitative Metric:** Track the model's **Perplexity** on a held-out test set.
- **Qualitative Evaluation:** The final report must analyze the generated cards based on:
  - **Coherence:** Is the generated rules text grammatically correct and sensible?
  - **Thematic Fit:** Does the card's name, mechanics, and flavor align with the "My Hero Academia" character or concept it represents?
  - **Playability:** Does the card seem balanced within the rules of MTG? (A brief analysis is sufficient).

## 9 Model Analysis

Perform a critical analysis of the model's performance.

- Discuss common failure modes (e.g., generating nonsensical mana costs, repetitive text, mechanically broken cards).
- Analyze the effect of different prompting strategies on the generated output.
- Provide recommendations for improving the generation quality in future work.

## 10 Deliverables

- A well-commented Python code implementation of the data preprocessing and model fine-tuning pipeline.
- A .json or .txt file containing the final curated card set (at least 100 cards).



- A report documenting the detailed approach, including data preprocessing, the thematic map, model training, and evaluation results.
- Analysis and discussion of the model's performance, challenges faced, and recommendations for improvement.
- A presentation that will be discussed at the time of the Oral Exam.

## Notes

You are encouraged to utilize existing deep learning libraries such as **TensorFlow or PyTorch** for implementing the model. Make sure to properly document your code and provide clear explanations in the report to demonstrate your understanding of the concepts and techniques used.