Dungeons and Data

By: Patrick Brown

Case

The client is Latitude. They are an emerging gaming company. They specialize in AI generated content. Their initial release AI Dungeon is a text based roleplay adventure game using language models to generate content.

Design

The goal of this project is to improve player experience with our AI DM. We hope to find the kind of interactions that people have during a DnD session.

1. We will scrape our text data.
2. We will use clustering to find kinds of interactions.
3. We will fine tune a pretrained language model with our transcripts leveraging our discovered clusters.

Data

We scraped 871 DnD session transcripts. Session lengths ranged from 1 to 5 hours in length. We seperated our transcripts into individual lines spoken (documents). After cleaning these documents, we had 1,300,000 documents leftover.

Tools

- Numpy
- Sklearn
- Spacy
- Hugging Face
- PyTorch
- Google Colab

Conclusions

Found clusters were relatively interpretable. There was still a large amount of variance within clusters. The average DnD typically consists of a large amount of social interactions vs gameplay.

Communication

End results were presented via a slide presentation. The language models developed are hosted on HuggingFace, here.