Anec.AI- Turning Bytes Into Comedy

Introduction

Anec.AI is a system that can generate short, funny stories or statements based on given keyboards.

Week I progress

Agenda:

- Road map
- Team formation
- Data collection

Problem statement

Humor is inherently subjective, and what one person finds funny, another might not. But with machine learning and deep learning, we can try to model some patterns of humor. To successfully implement the project, we created road-map:

☒ *Preliminary Planning & Research:*

- Team formation.
- Identify the project's objectives and scope.
- Research existing humor datasets and potential sources.

☐ *Data Collection & Processing:*
- Gather a large dataset of anecdotes or short stories based on what users find humorous (upvotes/downvotes).
- Preprocess this data: tokenization, stemming, removing special characters, etc.

☐ *Model Selection & Prototyping:*
- RNNs (like LSTM or GRU) for text generation.
- Transformers (like GPT and BERT) for NLP tasks.

☐ *Keyword Integration:*
- Embed the keywords into a vector space (like Word2Vec or FastText).
- Provide these embeddings to guide the generation process.

☐ *Training & Evaluation:*
- Using a pre-trained model (like GPT), fine-tune it on anecdotes dataset.
- Use NLP metrics (like BLEU or ROUGE) and user feedback on the generated humor f0r evaluation.
- Implement filters to avoid generating inappropriate content.

☐ *Deployment & Scalability:*
- Develop a user-friendly interface for the anecdote generator bot.
- Deploy the model on a scalable platform (Telegram).
- Monitor usage and ensure uptime and performance metrics are met.

At the moment, *Preliminary Planning & Research* is fully completed, and work on the second point of road-map.

Team Formation

The team was created on the basis of each member's stack. Some additional roles had to be shared among the participants, due to the small size of the team. You can see the table below with all members and their contacts:

Team Member	Elina Akimchenkova	Ruslan Abdullin	Anatoliy Pushkarev
Telegram ID	@akmchnkv	@Fliegende_Rehe	@anatoliy_pus
Email	e.akimchenkova@innopolis.university	ru.abdullin@innopolis.university	a.pushkarev@innopolis.university
Role	Data engineer	ML engineer	Software Developer Quality Assurance Engineer

Invited expert, head of the stand-up club of Innopolis - Albert Nasybullin. We will consult with him on the theory of humor.

Data collection

We started building data-sets using several sources:

- Public datasets: Kaggle humor datasets.
- Web scraping: Python Reddit API Wrapper (subreddits like r/Jokes, r/Anecdotes, r/funny).

Since we don't have much time to collect data. Now we try to use special techniques to augment the data:

- Back-translation (translate anecdotes to another language and then translate it back to original).
- Paraphrasing tools (use tools to rephrase the anecdotes).

Useful links

- GitHub link