





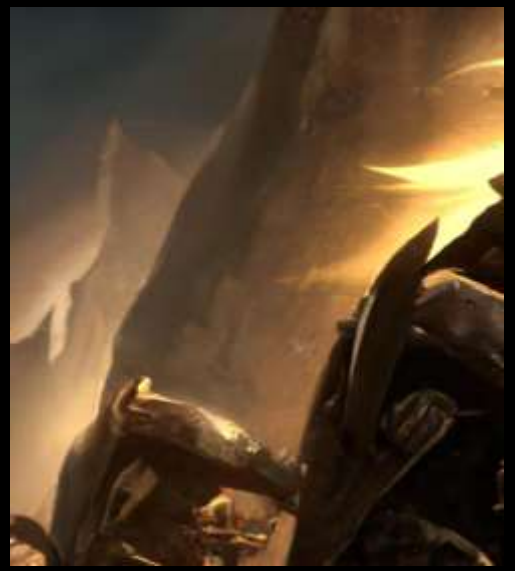

# NLP Anime Quotes Classification & Sentiment Analysis

Jorge Barrios, 20759401

Anime is a popular form of entertainment around the world, with millions of fans watching it every day (including myself). One of the most popular aspects of anime is the quotes from characters, which often become popular catchphrases or memes, so it's a good approach to develop a system that can classify anime quotes based on their sentiment and emotional content. Let's get otaku … !!

## Objectives

Build a dataset of anime quotes, along with their associated sentiment labels, using web scraping techniques, preprocess the dataset using natural language processing techniques, such as tokenization, stemming, and stopword removal, classify anime quotes into different sentiment categories, such as positive, negative, and neutral and provide insights into the language used in anime using machine learning algorithms.

## Techniques

- Data Collection: Use web scraping methods to collect anime quotes from popular anime websites, such as Zoro.to, and categorize them into different sentiment categories using crowdsourcing.
- Data Preprocessing: Preprocess the dataset using natural language processing library tools in jupyter for tokenization, stemming, and stopword removal, to prepare it for machine learning.
- Machine Learning: Train a machine learning model, such as a neural network, naïve bayes or support vector machine, to classify anime quotes into different sentiment categories based on the preprocessed dataset.
- Sentiment Analysis: Perform sentiment analysis on the dataset using techniques such as lexicon-based analysis and deep learning to finally provide insights into the language used in anime.

## Expected Outcomes

Analyze dataset of anime quotes, categorized into different sentiment categories using the machine learning model that can classify anime quotes finally to have insights into the language used in anime, and the sentiment and emotional content of popular anime quotes. The results of this project can be used to improve our understanding of the language used in anime and develop new tools and applications for analyzing and categorizing text data in other web domains.