

Projet SMART 2024

# ANIME ATLAS

Machine Learning / Large Language Model  
Multimedia Recommendation System

*(Demonstrated on Anime Dataset)*

*Team*

*NGUYEN Mathis*

*NGUYEN Le Tuan Khai*

*Mohamed-Ali Lajnef*

*FARHAT Widad*

*Yikang Su*

*Zijing Weng*

*Remy Etienne*

*Elie Andrianarisolo*

## Project Description

Our project focuses on developing a sophisticated multimedia recommendation system tailored to anime enthusiasts.

Unlike traditional recommendation systems that rely on user historical data, our system will leverage collaborative and content-based filtering techniques along with natural language processing (NLP) to generate personalized recommendations based solely on the user's favorite anime list. By analyzing the attributes of the user's favorite anime, such as genres, themes, and characters, our content-based filtering approach will identify similar anime to recommend.

Additionally, we will utilize NLP to extract meaningful insights from the synopses of animes, enriching the recommendation process.

Moreover, we aim to integrate large language models (LLMs) to further enhance the system's ability to understand user preferences and generate contextually relevant recommendations.

This system will help users to discover new anime aligned with their unique tastes, fostering a more engaging and immersive anime viewing experience.

This system can be generalized to recommend books and films when given appropriate datasets such as IMDb reviews for TV shows or Goodreads dataset for books.

## Development Progression

1. Clean and boost the dataset
2. Develop a website as front-end
3. Develop a recommendation system based in collaborative filtering.
4. Use NLP to develop content-based filtering to enhance existing solution.
5. Integrate LLM / GPT to handle more complex queries

Step 1 has the highest priority while step 2, 3, 4 can be developed in parallel. Step 5 is a possible but optional evolution to the project if we have enough time after finishing the first 4 steps.

## Positioning in relation to existing products

According to our research, services specific to this domain has already existed, although not too numerous because of how niche the genre is. Of the services we tested, they all employ Machine Learning technique in one way or another in order to categorize and recommend suitable titles

However, we think that with extension such as NLP and/or LLM, we can achieve a more sophisticated service that can not only better model anime titles based on extracted text data but also able to handle more complex request from users.

## Requirements (software, hardware, data)

The requirement includes

1. Anime Dataset
  - Available on Kaggle (2Gb):  
<https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset>
  - In need of a more lightweight and faster loading dataset than the dataset suggested.
2. GPU
  - For potential uses in Deep Learning or LLM Finetuning
  - Preferably on a Laptop or through Cloud services (Kaggle, Colab, ...)
3. Software
  - Python, Pandas, Numpy – native programming language
  - Streamlit, Flask (optional) – Front-end
  - Scikit-learn, Pytorch – ML/DL
  - OpenAI API, HuggingFace API – LLM/GPT support
  - Langchain – LLM Integration

