Task 4 - Project management and implementation
AlBeSa

*Authors: Beloslava Malakova (TU/e: 1923404), Alicja Gwiazda(TU/e:2017830),
Stanimir DImitrov(TU/e: 1932217), Aleksandra Nowińska (TU/e: 2008580)*

Sprint 1: The scrum master was Aleksandra (Ola). In the first sprint, we focused on setting up the Agile strategy and formulating our User/Functional/AI requirements into Trello Boars tasks. We also set up a GitHub repository. We started off by gathering an initial dataset for our task (we had to pivot from the previous one), and writing code automated data collection, so that our work is reproducible. We have also created a simple MVP of the UI,which includes a basic image upload interface and display functionality for artwork retrieval. The UI MVP is designed so that users can upload images and it displays search results in a clean, intuitive playout using streamlit - establishing the foundation for our multimodal art algorithm. Two new branches were created in the repository, one called "UI" for the previously mentioned point, the other "MVP" for this current iteration of tasks. For both of the branches, team members have been committing regularly, with adequate git messages. The meetings were well organized thanks to the scrum master.

Sprint 2: The scrum master was Stanimir (Stan). The second sprint was research heavy, as we had to decide which exact vector database to use, which embeddings and retrievers would be most suitable. After a detailed comparison of Weaviate and Qdrant we set up a cluster in Qdrant Cloud (free tier). We then focused on research related to the embeddings since the free tier requires us to generate embeddings outside of the database and then populate it. During this sprint, Aleksandra focused on implementing data preprocessing and evaluation pipeline for our retrieval system. We developed two files for generating synthetic photo variants that simulate real-world user queries (including transformations like blur, glare ect) and second for baseline evaluation using CLIP embeddings. This work created a standardised test dataset of 250 variants from 50 artworks, achieving baseline metrics. This sprint we also included proper data structure organization compatible with the kaggle dataset that we are using and created clear documentation for easier group collaboration.The scrum master kept us organized.

GitHub repository: https://github.com/alicegwiazda/822196_SE