Data Wrangling & Visualization: Second Checkpoint

# Interactive Visualization of `UMAP` Algorithm on User Data

Nikita Zagainov, Dmitry Tetkin, Nikita Tsukanov

April 2025

## 1 Overview of Our Progress

By the time of the second checkpoint, we have almost completed the plan of our project, specifically:

- Algorithmic part:
  - Original `UMAP` algorithm GitHub repository was used as a backbone
  - We modified the original implementation to enable streaming of intermediate embeddings for visualization of concrete steps
  - We implemented data preprocessing pipeline to filter user input date for better quality of visualization
- Backend:
  - We implemented a `FastAPI` backend server that accepts user input data, performs data preprocessing, and applies the `UMAP` algorithm to it, returning the history of intermediate embeddings along with the final embedding
- Frontend:
  - We created a `Express.js` frontend application
  - We added a single page with submission form where user can upload their data
  - We added frames for 2D (using `D3.js`) and 3D (using `Plotly.js`) visualizations
  - We added a slider to allow user to select the step of the algorithm to visualize
- Deployment:
  - Every of the services is wrapped in a `Docker` container
  - We provide `docker-compose` file to run the whole application
  - **Note:** We do not have a running deployment of the application since `UMAP` algorithm requires a lot of resources to run. We do not plan to deploy the application, it is intended to run locally only

## 2 Next Steps

Our project is almost done, but it still requires some polishing and end-to-end testing. We plan to try to run the whole application on different real-world datasets and fine-tune our application where necessary