

## System Architecture

This section provides a high-level system architecture of the prototype. This architecture will include different components involved in the process, from data retrieval to model training and ethical classification. Figure 1 provides an overview of the components and their interactions within the transformation.

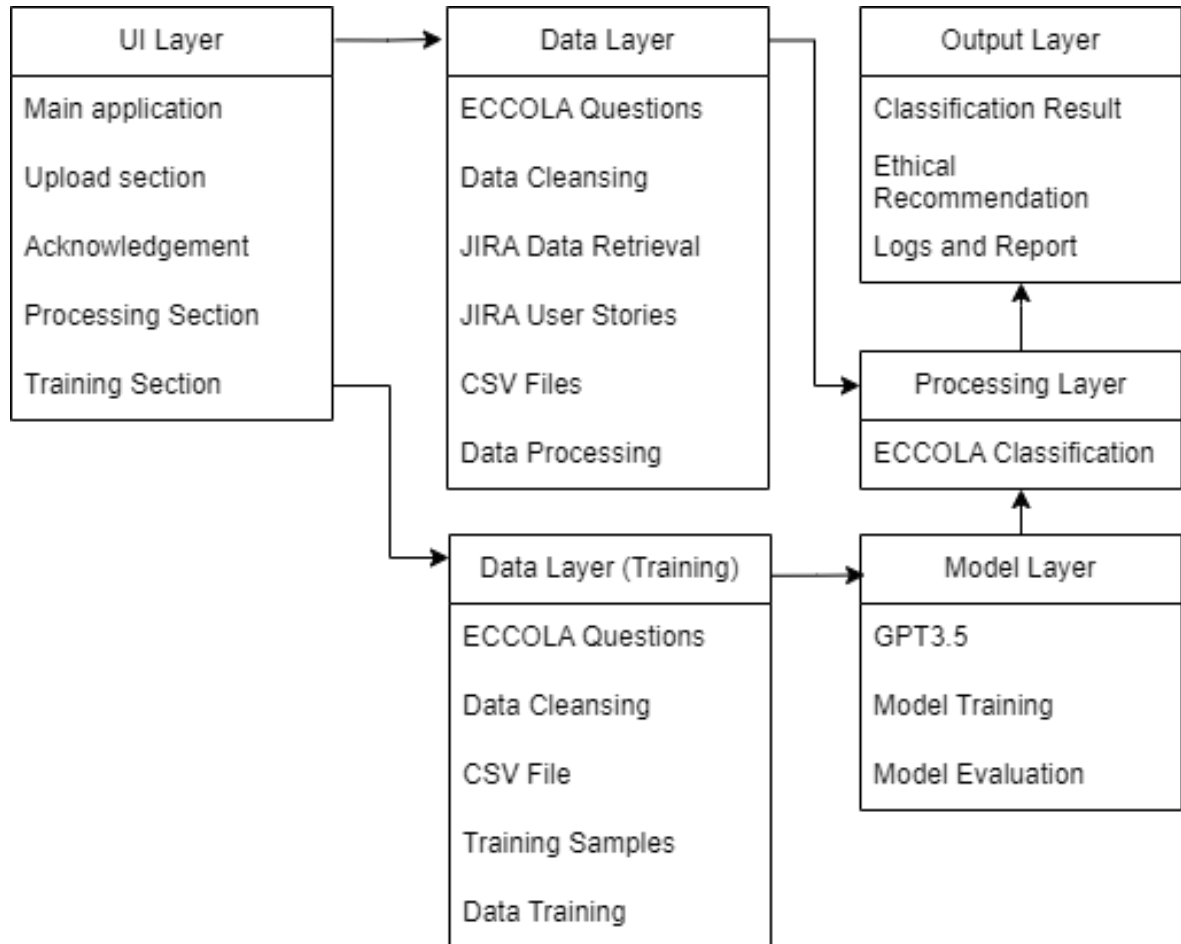


Figure 1: System Architecture

The User Interface (UI) Layer is designed to offer a user-friendly platform for interacting with the system. Its primary goals include providing an interface that allows users to upload data, initiate processes, and view acknowledgment points and logs. The functions of this layer encompass providing a graphical user interface (GUI) where users can upload ECCOLA questions, JIRA user stories, and training samples. It also enables users to start processing and training tasks, displays acknowledgment points, and outputs logs and final classification results. The key components of this layer are the main application, whose front-end was built in Python using Tkinter. The front-end has sections for uploading various datasets (ECCOLA questions, JIRA user stories, and training samples), as well as sections dedicated to processing, training, and receiving suggestions. In any section where user stories will be passed to the LLM, an acknowledgment notice will be displayed.

The Data Layer focuses on managing data storage and retrieval, ensuring data is cleaned and structured appropriately for further processing. Its goals include handling data storage using local CSV files and preparing data through preprocessing for the processing layer. The

functions involve managing data storage and performing preprocessing tasks to clean and prepare the data. Components of this layer include local CSV files (for ECCOLA questions, JIRA user stories, and training samples) and data preprocessing mechanisms.

In the Processing Layer, the system retrieves data from JIRA using an API, cleans and prepares this data for model training and classification, and applies the ECCOLA model to classify data and suggest ethical considerations. The goals are centered on data preparation and cleaning, along with ethically classify user stories based on the ECCOLA model. Functions of this layer include handling JIRA data retrieval or manual data retrieval, data cleaning, and processing, as well as using the ECCOLA model for classification and ethical recommendations. The key components are the JIRA Data Retrieval Module, Data Cleaning and Preparation Module, and ECCOLA Classification Module.

The Model Layer integrates the GPT-3.5 API for model training, evaluation, and validating. Its goals include integrating with the LLM model (GPT-3.5 API), fine-tuning LLM model, and evaluating it. Functions in this layer are aimed at fine-tuning the model using ECCOLA questions and training data, evaluating the model's performance, and using the trained model for classifying new user stories and generating ethical recommendations. Components consist of the GPT-3.5 API Integration, LLM-Model Training Module, and LLM-Model Evaluation mechanisms.

Finally, the Output Layer is responsible for outputting classification results and ethical recommendations, logging actions, and providing detailed reports. The primary goal is to offer easily-followed ethical suggestions for IT practitioners. The ethical suggestion is outputted from ethical recommendations derived from ECCOLA. For future enhancements, we suggest adding customized ethical recommendations generated by the LLM-model. This layer also maintains logs and provides detailed reports for users. Components of this layer include sections for Classification Results, Ethical Recommendations, and Logs and Reports.