

CS319 PROJECT PROPOSAL

Project Title: CodeArena

Team Members:

Ali Şen / 22203368

Barış Peksak / 22201659

Berkay Şimşek / 22303338

Gökay Nuray / 22302913

Muhammet Furkan Demir / 22302644

Description

- **Motivation**

Both humans and AI now produce software artifacts like code, diagrams, and test cases. Yet deciding which is clearer, more accurate, or more useful remains subjective. Our project introduces a platform that enables comparisons between these artifacts, with features like blinded evaluations, comments, and customizable rating criteria. This not only helps researchers measure quality more objectively for AI research but also reveals where human creativity still provides unique value compared to AI generated outputs.

- **Goals and Important Problems to be Solved**

The goal of this project is to build a modular web platform where researchers can easily set up and run studies comparing human-made and AI-generated software artifacts. Today, evaluations are often unstructured, biased, and time-consuming. Our system addresses these issues by offering a user-friendly and controlled environment: researchers can upload artifacts, assign participants, test their skills, and define comparison tasks. By simplifying this process, the platform makes evaluation tasks faster, more reliable and easier to manage, while producing feedback that can be easily analyzed and shared. Additionally, the system will be designed with extensibility in mind, allowing new artifact types, evaluation criteria, and analysis tools to be integrated with minimal rework for future studies.

- **Features**

The application will have several core features.

- **User Management:** A system with distinct roles such as Researcher, Participant, and Admin, featuring registration and role-based control.
- **Artifact Management:** The module will allow researchers to upload, tag, and organize various artifact types like source code, test cases, and UML diagrams.

- **Participant Competency Assessment:** This feature will consist of questionnaires and technical quizzes to evaluate the participant's expertise level before a study begins.
- **Artifact Comparison Interface:** Presents two or more artifacts in a side-by-side view with synchronized scrolling, enabling participants to add comments and rate them based on criteria. This interface will also support a blinded evaluation mode to prevent any possible bias.
- **Dashboards:** For both researchers to monitor study progress and for participants to view their assigned tasks.
- **Selling Points and Interesting Parts:**

This project's key selling point is its focus on enabling formal, blinded human-subject studies. Architecture is designed to allow new artifact types and analysis tools to be integrated with minimal rework. The ability to handle artifacts from various diverse sources, including dynamic generation via LLM APIs and bulk uploads, makes it highly versatile. Furthermore, the dual deployment mode ensures accessibility for different use cases.

Extra Features

- **Dashboard Visualizations with Charts:** In order to provide the user with easier to understand data and evaluations, we visualize compatible data with charts.
- **Evaluation report exported as PDF:** Evaluation data must be easy to share between your colleagues, team members and other parties involved, therefore we introduce a system that takes the final data and exports it as a PDF so that the user may share them with the people they need to in an easier and accessible manner.
- **Task pending/finished notification:** Some evaluation tasks may run in the background as the user is busy with something else, therefore we will add a notification feature for the task's status so the user may know when it is finished or not.