CS Capstone '26
JaPolCha

UC CompSci 2026

Members

Jack Nguyen

- *nguye2lo@mail.uc.edu*
- CS Senior '26
- Area of interest: AI, OCR, Game Dev

Tiep Tran

- *tran2tp@mail.uc.edu*
- CS Senior '26
- Area of interest: ML devops, Agentic AI, Context Engineering

Chau Nguyen

- *nguye2cu@mail.uc.edu*
- CS Senior '26
- Area of interest: AI, Backend, Infra

Ryan Persaud - *ryan.persaud@midea.com* - Project Leader at Midea

Project Abstract

- **Matcha** is an automated system designed to streamline the process of matching bank statements with receipts using **Natural Language Processing (NLP)** and **Image Processing techniques.**

- Sponsored by **Midea America Research Center**, the project aims to reduce manual effort and human error in financial cross checking tasks. The system **extracts key transaction details** such as amounts, vendors, and categories, then identifies discrepancies and triggers **alert workflows** when mismatches occur. Verified transactions are automatically **classified and exported to a Vendor Expense Claim Excel sheet**, providing an efficient, secure, and intelligent solution for financial data verification and management.

Wrong receipt uploaded

Other information for the Vendor Expense Claim
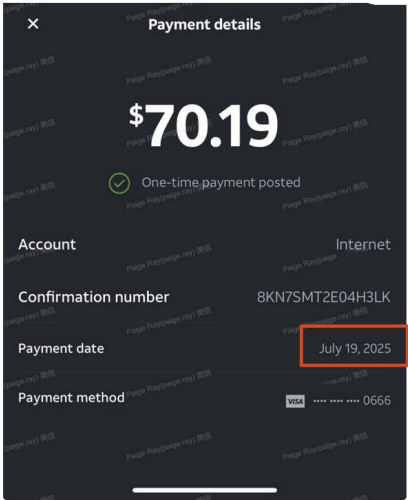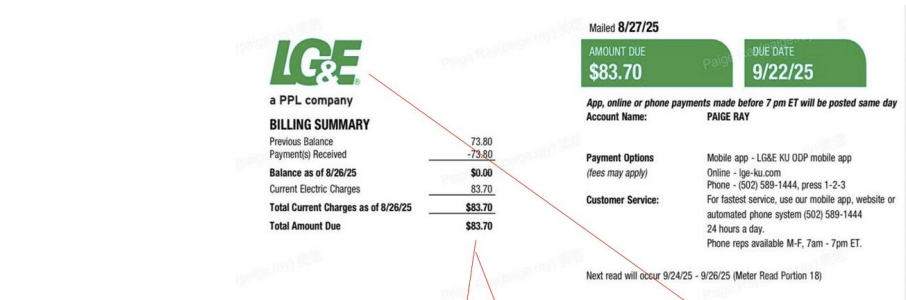
Students must determine the best way to predict the type

User Stories

- As a user, I want to upload receipts and bank statements, and the system can automatically cross-check them for accuracy.
- As a user, I want to be alerted if there are missing or mismatched receipts.
- As a user, I want the system to automatically categorize the purchases and export them into a structured Vendor Expense Claim Excel sheet.
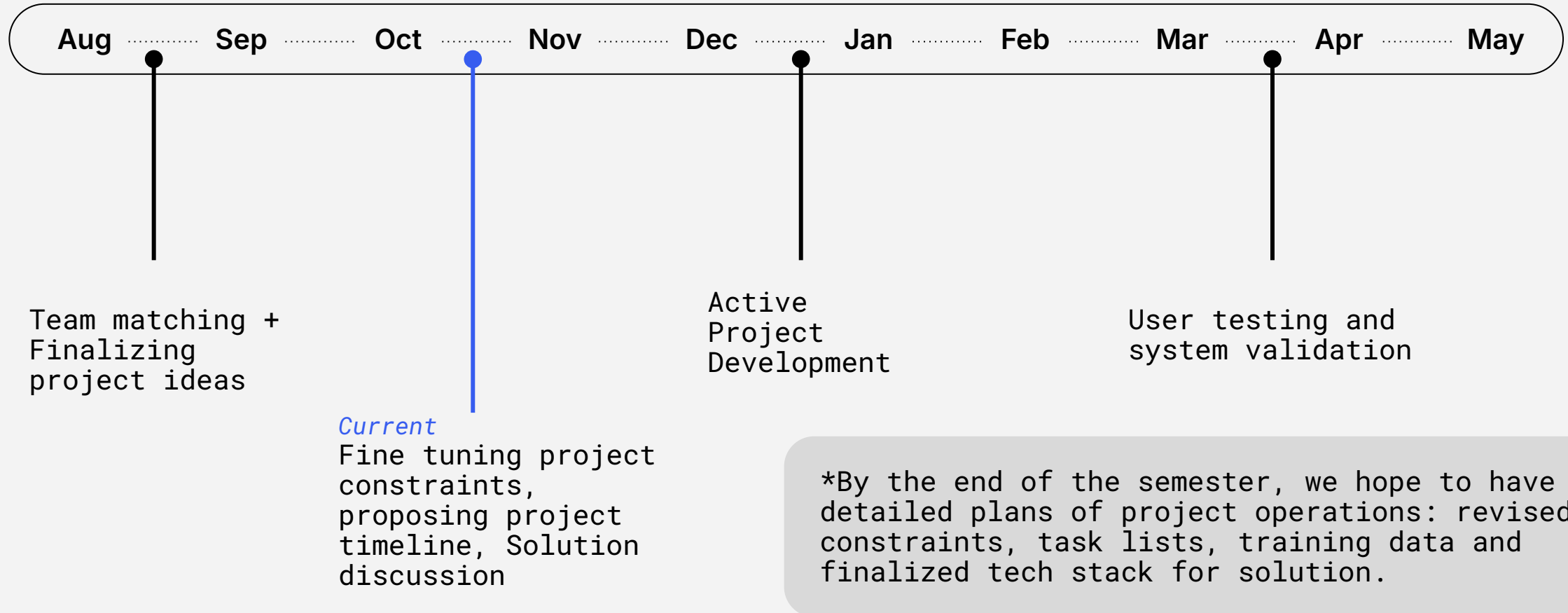
Design Diagram
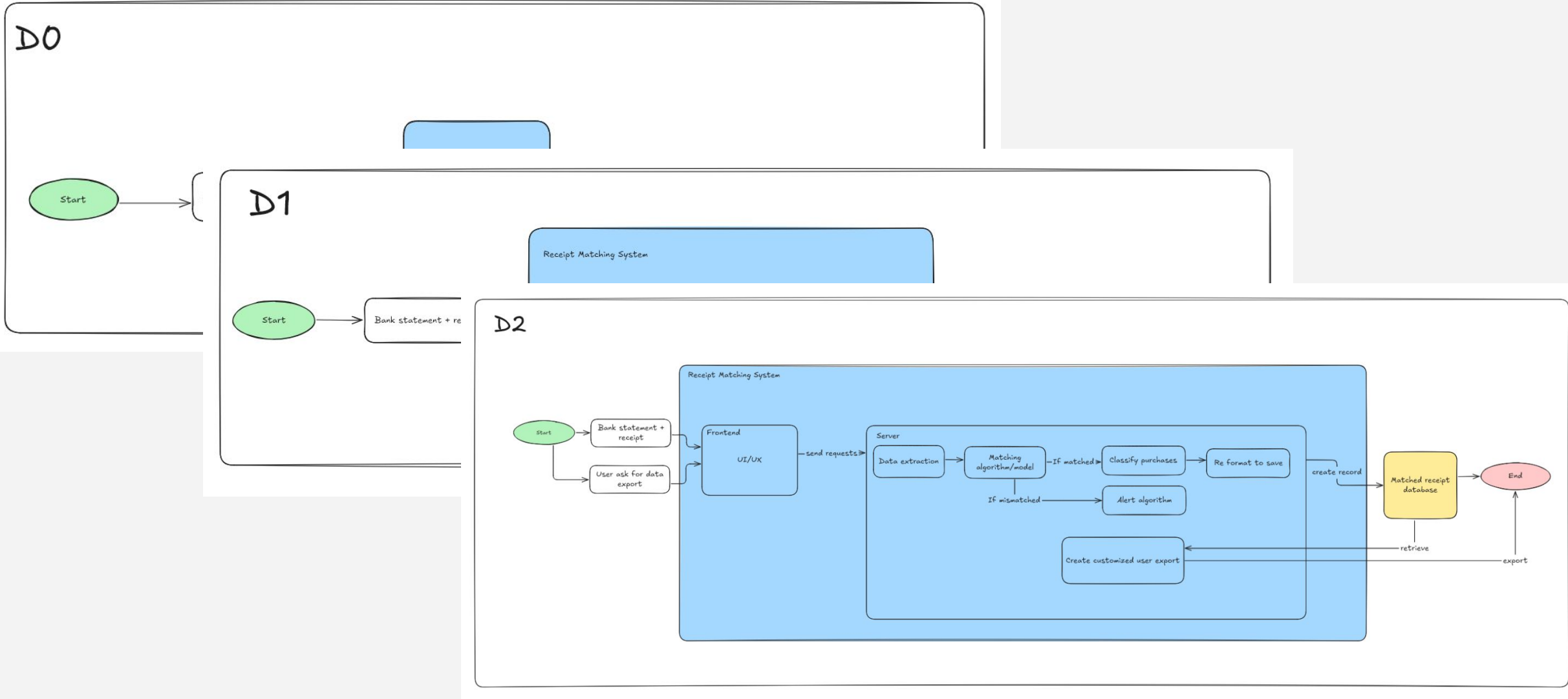
Project Constraints

- Economically: we will be using free and open-sourced tools provided to us (Python libraries, AWS…)

- Security: ensure that any financial or personal information remains confidential and only limited to our uses within the project. In detail, the project sponsor also provides constraints to LLM usage that prohibit training/sending data to a cloud hosted LLM service provider

- Legally: comply with data protection standards and avoid using copyrighted or proprietary models without proper licensing. All the data we'd use must also be anonymized to prevent unintentional disclosure of sensitive information during training or testing.

Project Progress

P1 2025 ──────────────────────────────── **P2 2026** ──────────────────

| Aug ···· Sep ···· Oct ···· Nov ···· Dec ···· Jan ···· Feb ···· Mar ···· Apr ···· May |

Team matching +
Finalizing
project ideas

Current
Fine tuning project
constraints,
proposing project
timeline, Solution
discussion

Active
Project
Development

User testing and
system validation

*By the end of the semester, we hope to have a detailed plans of project operations: revised constraints, task lists, training data and finalized tech stack for solution.

Work Distribution

- Given the ideal project timeline below, our team will all work together on major features, ensuring cohesion and coordination throughout the whole project.
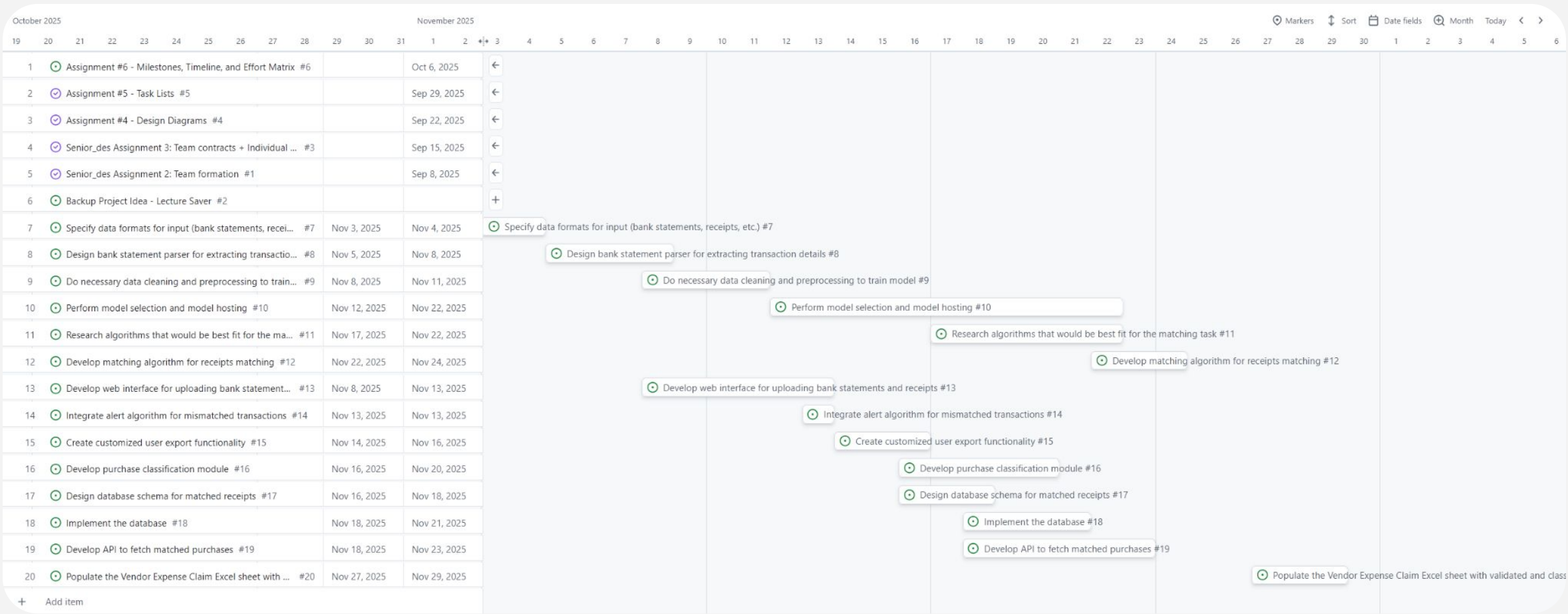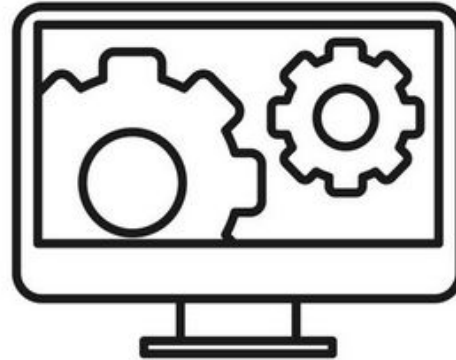
Expo Demo

Model Showcase

DATA CLASSIFICATION

- Showcase our **classification model performance**
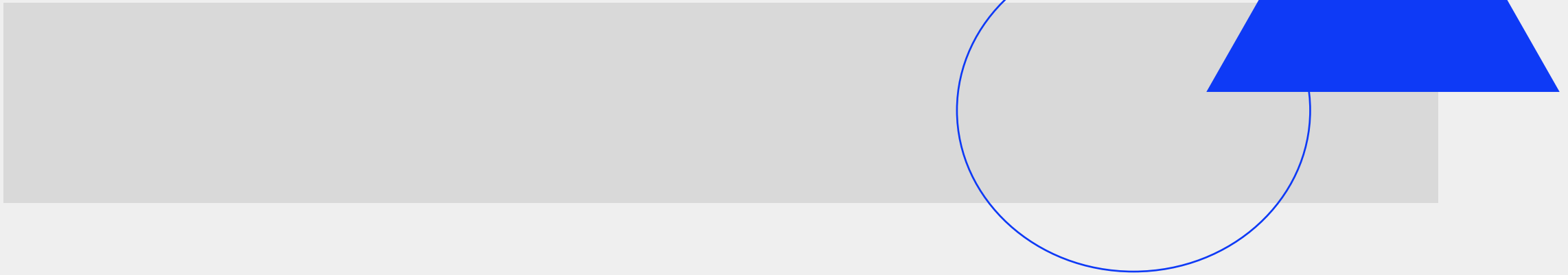- Model overview, training techniques, innovations, …

System Showcase

Showcase the flow of the system, including how the system will **flag mismatch receipts** and how user can **export a customized report of the system**

THANK YOU

We look forward to hearing your feedbacks.