AAIE (Artificial Assessment Intelligence for Educator) Platform System Design

Overview ***1***

Diagram Description – Context Diagram **3**

1. System Boundary (AAIE Platform) 3

2. External Actors 3

3. Key Functional Modules (Inside AAIE) 4

4. Major Data Flows 5

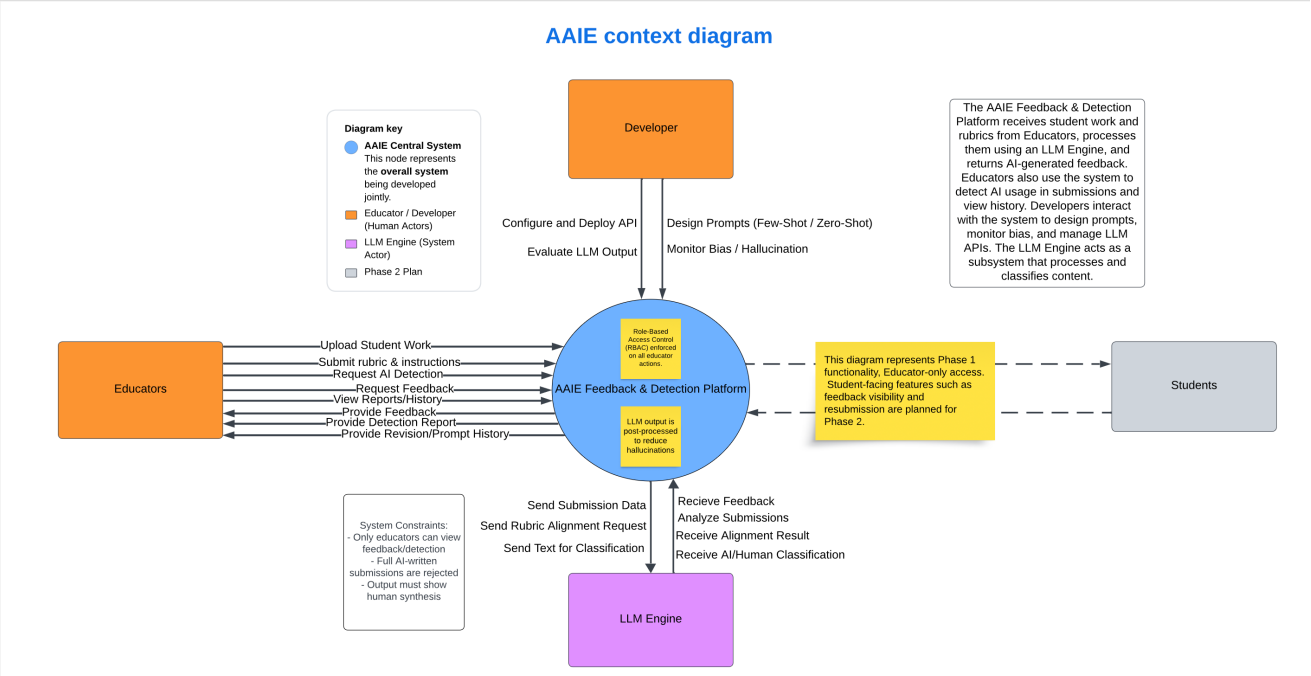
5. Scope Notes for Phase 1 (from PO) 5

System Components (Inside AAIE Platform) 6

Data Flow Description 8

Overview

The context diagram for the AAIE (Academic AI Evaluator) system outlines the high-level structure of the application by defining the interactions between the system and its external actors. It serves as a top-level view of the system's boundaries, inputs, and outputs, showing how various stakeholders interact with the platform. The context diagram ensures all development teams (Product Engineering, Model Development, and Data Curation) have a shared understanding of the system's role in the broader academic environment.



Important Notes:

Phase 1 of AAIE focuses on enabling educators to submit student work for automated AI classification, rubric-based scoring, and formative feedback. The system allows educators to review results, add their own comments, and download a combined feedback report.

In Scope:

- Educator login/register.
- Upload student work.
- LLM Engine analysis: AI usage classification (Human/AI/Hybrid), rubric scoring, and short feedback generation.
- Educator review of AI output and addition of teacher feedback.
- Download of final structured feedback report.

Deferred to Phase 2:

- Role-Based Access Control (RBAC) enforcement.
- Rubric upload and rubric alignment logic.
- Prompt history viewing/audit logs.
- Plagiarism detection and prompt similarity analysis.
- Student-facing submission workflows.

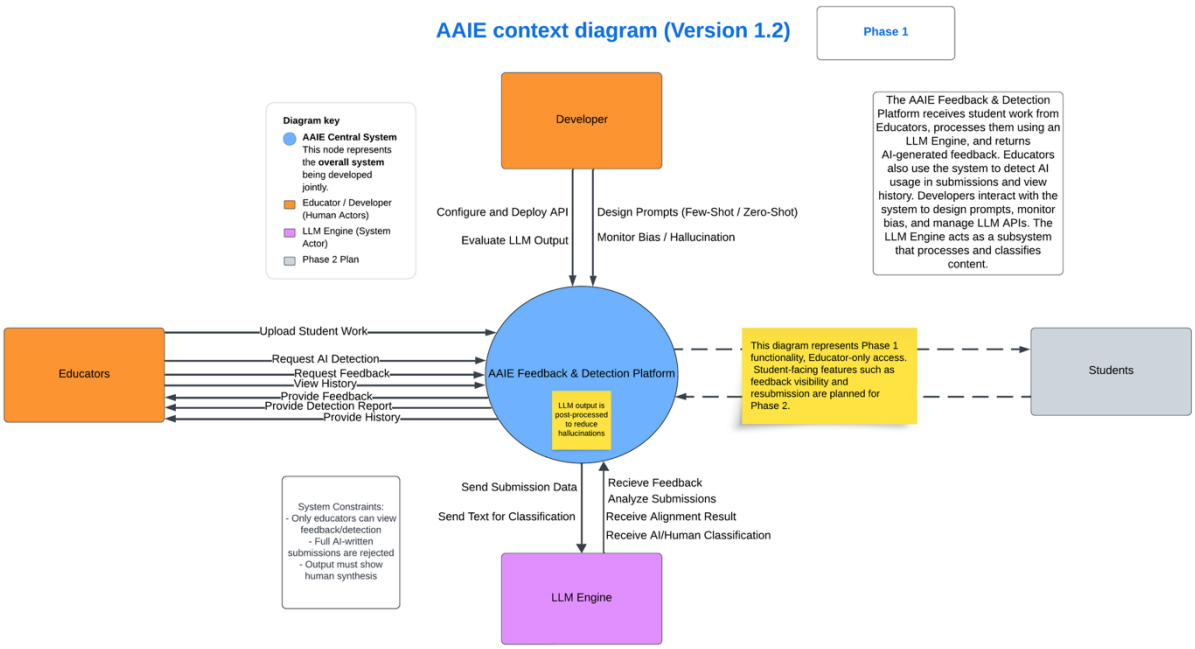


Diagram Description – Context Diagram

Diagram Name: *AAIE Prompt Feedback & Detection System – Context Diagram*
Version: v1.2
Stakeholders: Educators, Students (phase 2), System Admins, PO, LLM & Model Team
Purpose: This context diagram illustrates how the core AAIE platform interacts with external actors and subsystems. It defines the system boundary and identifies all the major data flows and integration points with external entities.

1. System Boundary (AAIE Platform)

The AAIE platform is the core system providing functionalities for prompt evaluation, feedback generation, AI usage detection, and academic integrity checking. Inside this boundary are several tightly coupled modules that together form the feedback and detection workflow.

2. External Actors

A. Educator (Primary User)

- Logs in and uploads prompts and student responses (real or simulated)
- Selects rubrics or defines new evaluation criteria (Deferred to Phase 2)
- Reviews AI feedback, similarity results, and rubric scores
- View structured feedback reports

Interactions:

- Initiates prompt submission and feedback generation. Interact with the system by uploading student assessments, viewing AI-generated feedback, and similarity or rubric alignment reports.
- Views/downloads outputs and evaluation reports. Receive academic integrity and AI misuse analysis.

B. System Admin / Developer

Currently assigned to the Product engineering Team

- Oversees platform operations and simulation processes
- Uploads synthetic responses using GenAI for testing
- Connects backend models (APIs, LLM endpoints)
- Deploys and monitors model updates and system health

Interactions:

- Deploys or updates LLM APIs
- Monitors hallucinations, biases, and system logs

C. Model Developers

Currently assigned to Model Design Team

- Provide backend LLM models (e.g., fine-tuned or zero-shot)
- Provides and maintains API endpoints for prompt detection, rubric feedback, and similarity scoring.
- Receive requirements and output feedback from the platform
- Help evaluate authenticity, hallucination, and rubric alignment

Interactions:

- Model API integration and result evaluation
- Back-and-forth on test performance and prompt design
- Designs few-shot/zero-shot prompts
- Consumes input prompts and returns structured evaluation data

Internal Data Curation Team:

Currently assigned to Data Curation Team

- Supplies curated and synthetic datasets for training and benchmarking.
- Ensures prompt-response pairs are realistic and representative of actual assessment types.

D. Students

- Indirectly represented (not active users in Phase 1) but are the subject of evaluations. Their submissions are the core data processed by the system.
- In Phase 2 or future expansion, students receive feedback reports and resubmit corrected responses.

Optional Actors:

- **University Stakeholders (Optional / Governance Layer)**
  - Review reports and dashboards generated from educator usage.
  - Use analytics and summaries to assess compliance and AI usage trends.
- **External Services (e.g., FastAPI / Hosting / GitHub)**
  - Backend infrastructure such as the API server, version control, deployment environments, etc.
  - Supports frontend/backend integration and service reliability.

3. Key Functional Modules (Inside AAIE)

1. Prompt Submission

- Collects prompt, student response, rubric, and metadata
- Interfaces with role permissions via RBAC (Deferred to Phase 2)
- Hands over data for LLM processing and feedback

2. Feedback Generation

- Calls LLM API to generate natural-language feedback
- Aligns with the educator's rubric and submission intent
- Connects with detection pipeline

3. AI Usage Detection

- Identifies possible AI-generated text via heuristic or ML model (Deferred to Phase 2)
- Flags suspicious submissions with a score or label
- Integrated with authenticity scoring

4. Rubric Evaluation

- Matches feedback and student response against rubric (Deferred to Phase 2)
- Outputs clarity, structure, and relevance scores
- Supports similarity reports and feedback breakdown (Deferred to Phase 2)

5. Report Generator (Deferred to Phase 2)

- Aggregates all outputs into downloadable reports
- Allows PDF download and UI viewing

6. Model Interaction / Monitoring

- Allows Devs to deploy/update LLMs
- Logs hallucinations, biases, and low-confidence output
- Provides telemetry to GitHub or system log

4. Major Data Flows

- **Prompt/Response Data** → Flows from Educator → Feedback Engine
- **Generated Feedback** → Sent to UI for viewing/download (Deferred to Phase 2)
- **Rubric Scoring + Detection Results** → Flow into Final Report
- **API Results and Metrics** → Monitored by Admin and Model Team
- **Synthetic Data** → Generated by Admin → Fed into system

5. Scope Notes for Phase 1 (from PO)

- Real student data is not required; synthetic GenAI data can be used
- LLM should pause fine-tuning and focus on few-shot/zero-shot testing
- Prompt outlining, alignment evaluation, and authenticity scoring are priorities for later phases
- Report delivery summary and output chain tracking are required

System Components (Inside AAIE Platform)

1. Prompt Submission & Feedback Interface

- **Purpose**: Allows educators to submit student work (prompts + responses).
- **Functionality**: Supports file/text input, metadata tagging (e.g., subject, rubric ID), and triggers the evaluation process.
- **Links with**: User authentication in phase 1 and role-based access control in phase 2.

2. Enforce Role-Based Access Control (RBAC) → Deferred to phase 2

- **Purpose**: Ensures that actions within the system are authorized based on user roles (educator, system admin, student in future phases).
- **Functionality**: Limits access to specific functionalities like viewing analytics or uploading assessments based on assigned privileges.
- **Important For**: Academic integrity, data privacy, traceability.

3. Generate Feedback Module

- **Purpose**: Interacts with the LLM API to generate detailed feedback based on educator-submitted prompts.
- **Functionality**: Converts structured input into evaluation requests and formats the feedback into readable summaries.

4. Rubric Alignment Engine (Deferred to Phase 2)

- **Purpose**: Aligns LLM-generated feedback to specific rubric categories provided by educators.
- **Functionality**: Compares AI feedback with expected learning outcomes and highlights areas of alignment/misalignment.

5. AI Usage Detection (Academic Integrity Module)

- **Purpose**: Identifies if a student submission was potentially generated by AI.
- **Functionality**: Uses linguistic features, AI usage scoring, and cross-checks with known patterns to determine authenticity.
- **Flagging Options**: “AI,” “Hybrid,” “Human”

6. Feedback Report Generator (Deferred to Phase 2)

- **Purpose**: Combines outputs from rubric alignment, feedback generation, and AI usage detection into a single downloadable document.
- **Export Options**: PDF, structured JSON (for integration with LMS), or internal dashboard display.

7. Synthetic Data Generator (GenAI)

- **Purpose**: Replaces real student data with simulated responses using generative AI for testing and validation.

- **Benefits**: Bypasses ethical/data privacy issues during testing phases.

8. LLM Output Evaluator

- **Purpose**: Evaluates and validates the performance of the LLM (Large Language Model) for accuracy, clarity, bias, and hallucination.
- **Workflow**:
  - Test with few-shot prompts.
  - Monitor model performance with benchmark rubrics and curated response samples.
- **Key Insight from PO**: Focus should shift from fine-tuning to zero-shot/few-shot for quick iteration and higher adaptability.

9. Prompt-to-Output Chain Tracker

- **Purpose**: Monitors the entire flow from prompt input to final feedback to ensure transparency and auditability.
- **Includes**:
  - Input prompt structure
  - Rubric references
  - AI output logs
  - Educator revisions (if any)

10. Model Deployment & API Infrastructure

- **Purpose**: Manages how the backend models (e.g., GPT) are integrated via APIs.
- **Tools**: Uses FastAPI or similar backend frameworks.
- **Controls**: Versioning, access controls, health checks.

11. Bias and Hallucination Monitor

- **Purpose**: A safeguard system that flags responses that deviate from expected patterns or introduce factual inaccuracies.
- **Used for**: Validating feedback quality and ensuring fair academic assessment.

12. Admin Tools and Analytics Dashboard

- **Purpose**: Allows system administrators and educational governance users to:
  - Track usage
  - Access evaluation metrics
  - Review flagged submissions
- **Planned Features**: Aggregated AI usage statistics, rubric effectiveness heatmaps.

Data Flow Description

This section outlines how data flows within the AAIE system across key functional stages. The flow supports both the Educator's user journey and AI evaluation pipeline, ensuring feedback is personalized, authentic, and rubric-aligned.

1. Input Stage: Prompt Submission

- **Actors**: Educator (primary), System Admin (optional)
- **Actions**:
  - Educator logs into the platform
  - Submits prompt and sample student response (either real or synthetic)
  - Chooses or uploads a rubric for evaluation (Deferred to Phase 2)
  - Sets evaluation parameters (e.g., feedback type, strictness)
- **Data Created**:
  - Prompt Text
  - Student Response
  - Rubric Reference (Deferred to Phase 2)
  - Evaluation Metadata

Data Flow:
Input → Prompt Management Module → Stored in Prompt Database

2. Preprocessing & Role Enforcement → Transferred to Phase 2

- **Component**: RBAC (Role-Based Access Control)
- **Function**:
  - Validates user permissions
  - Filters and routes the request to the appropriate services
- **Data Transformed**:
  - Request metadata is tagged with user role
  - Certain options may be enabled/disabled based on access

Data Flow:
Prompt Data → RBAC Filter → Forwarded to Feedback & Detection Engine

3. Feedback Generation + (Rubric Alignment Deferred to Phase 2)

- **Components**:
  - Feedback Engine (calls LLM via API)
  - Rubric Alignment Engine Deferred to Phase 2
- **Function**:
  - LLM generates initial feedback
  - Feedback is parsed and matched against uploaded rubric (Deferred to Phase 2)
  - Highlights missing or misaligned points (Deferred to Phase 2)

Data Flow:
Prompt & Response + Rubric → LLM API → Generated Feedback → Alignment Module → Aligned Feedback Output

4. AI Usage Detection

- **Components**:
  - AI Detector Module
  - History Cross-Check (optional for Phase 2)
- **Function**:
  - Scans language style, tokens, and prompt-response structure
  - Compares with known AI response patterns
- **Outcome**:
  - AI-generated likelihood score (e.g., 0.85 likelihood AI) → optional for phase 2
  - Classification: Human, AI, or Hybrid

Data Flow:
Response Text → AI Detection Engine → Detection Label + Explanation

5. Report Generation & Download

- **Component**: Report Formatter / (Export Module Deferred to Phase 2)
- **Function**:
  - Aggregates feedback, rubric alignment (Deferred to Phase 2), and detection label
  - Formats it into human-readable report (PDF / dashboard) → (Deferred to Phase 2)
  - Allows educator to view in dashboard
- **Data Generated**:
  - Feedback Summary
  - Rubric Score Table (Deferred to Phase 2)
  - AI Usage Indicator
  - Authorship Likelihood

Data Flow:
Feedback + Rubric (Phase 2) + Detection → Report Generator → Downloadable Report (Phase 2)

6. Backend Monitoring and Logging

- **Components**:
  - API Call Tracker
  - Bias/Hallucination Monitor
  - System Logs / GitHub Integration
- **Function**:
  - Monitors API latency and LLM response anomalies
  - Logs submission history, review comments, role changes
  - Sends logs to GitHub / DevOps pipeline for peer review

Data Flow:
All Outputs → Log Aggregator → Dev Analytics Dashboard / GitHub

7. GenAI Simulation (Optional Data Source)

- **Actor**: System Admin / Developer
- **Function**:

    - Generates simulated student responses to seed the system
    - Enables testing of rubric alignment and LLM evaluation

Data Flow:
Rubric → GenAI → Synthetic Prompt/Response → Feedback Pipeline

8. Output Archival (Phase 2/3)

- Stores:
    - Submission records
    - Reports for audit trails
    - AI-detection logs
- Compliance with privacy laws (if real student data used in the future)

Data Flow:
Final Reports → Secure Storage (Encrypted Database)

These descriptions help clarify which users interact with what parts of the AAIE platform and how data moves in and out of the system. It sets the stage for detailed component design, API documentation, and testing plans by clearly visualizing platform scope and responsibilities.