SCTR’S PUNE INSTITUTE OF COMPUTER TECHNOLOGY

TECHFIESTA’26

INTERNATIONAL HACKATHON

CODE. CREATE. CONQUER

TITLE PAGE

Problem Statement ID : ED007

Problem Statement Title : Plagiarism & Authenticity Checker for Assignments

PS Domain : Education

Team ID : 206

Team Name (Registered on portal) : PixelCrafters

AI-Powered Plagiarism & Authenticity Checker

From Detection to Learning

Our solution is a lightweight AI assistant that checks plagiarism in text and code but more importantly, it helps students understand and fix their mistakes instead of just penalizing them.

Our goal is to shift plagiarism tools from fear-based policing to guided learning.

What We Built

- Web-based demo for Text + Code
- Real-time AI-powered analysis
- Separate Student & Teacher views
- LMS-ready (simulated LTI integration)

Why This Matters

- Encourages academic integrity
- Reduces manual review for teachers
- Scalable and cost-effective
- Built for modern, AI-assisted education

How OUR SOLUTION Works

- Semantic text similarity & paraphrase detection
- Code logic similarity (not just copy-paste)
- **Citation Suggestion Engine**
- Originality scoring with clear explanations
- Easy-to-read, explainable reports
- **Handwriting to text**

What Makes US Different

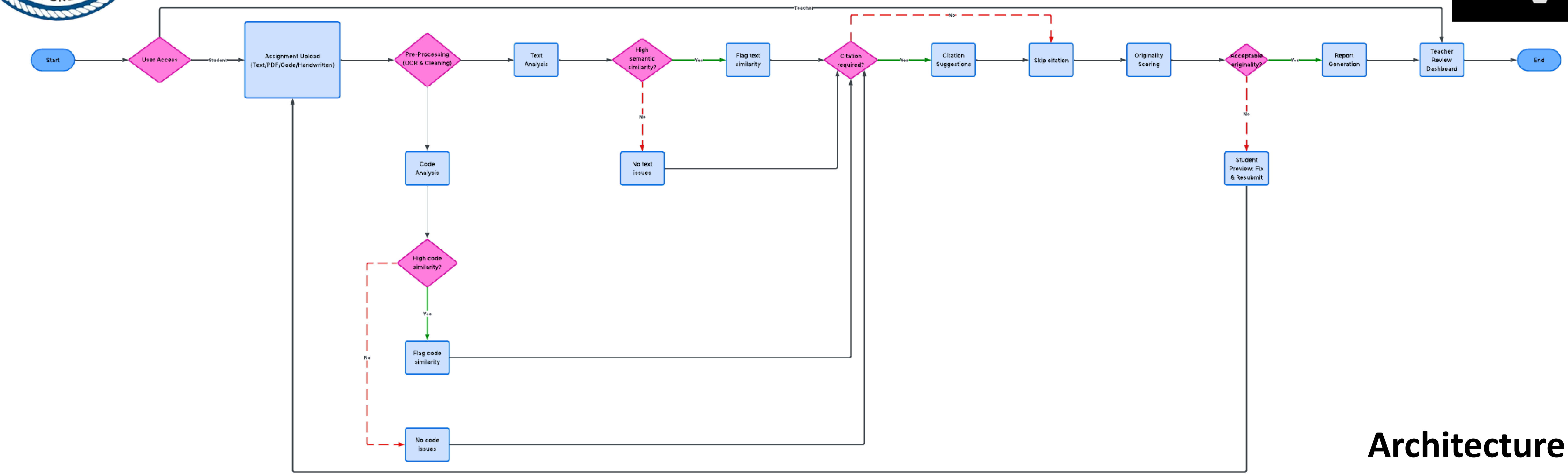
- Prevention over punishment
- Detects AI-assisted paraphrasing
- Handles both text and code
- Designed for learning, not just detection

A plagiarism checker that teaches students how to write ethically.

Problem	Our Fix
Paraphrased cheating	Semantic AI detection
Code plagiarism	Logic-level comparison
Fear-based tools	Fix-before-submit approach
No student guidance	Smart citation suggestions

TECHNICAL APPROACH

WorkFlow Diagram(Link)



Architecture Diagram

(Link)

Technologies Involved

Frontend

- Languages: JavaScript, HTML, CSS
- Frameworks: React.js / Streamlit

Backend

- Language: Python
- Frameworks: FastAPI / Flask

Database & Storage

- Tools: ChromaDB, JSON

AI / ML Engine

- NLP Models: Sentence-Transformers (semantic embeddings)
- Techniques: Cosine similarity, paraphrase detection
- Code Analysis: Tree-sitter (AST-based structural comparison)

FEASIBILITY AND VIABILITY

Feasibility Analysis

- Text similarity: sentence-transformers (BERT)
- Code similarity: AST + tokenization (MOSS, JPlag validated)
- Paraphrase detection: Transformer embeddings
- AI detection: Multiple statistical signals
- LMS integration: Standard LTI 1.3, OAuth 2.0
- Infrastructure: Free tools (HuggingFace, CrossRef APIs)
- Compute: 2-core CPU, no GPU needed
- Data: Text based Public corpora available

Challenges & Risks

- Universities process 1000s of assignments per semester
- Traditional tools miss 40–60% of paraphrased content
- Instructors spend 5–10 hours per course on plagiarism review
- False Positives : Overly strict paraphrase detection flags legit student work
- AI Detection Accuracy: Can't perfectly distinguish human vs AI content
- Source Coverage: Prototype misses web plagiarism

Mitigation Strategies

- AI accuracy → Use multiple signals + explainability
- Privacy concerns → On-premise option from day 1
- Instructor distrust → Show evidence, not just percentages
- Integration delays → Mock LMS in MVP, real plugins in Phase 2
- Latency → Cache results, optimize AST processing

IMPACT AND BENEFITS

Impact on Different Users

Institutions

At an institutional level, it supports a culture of academic integrity.

It offers a cost-effective alternative to traditional plagiarism tools and is designed to integrate smoothly with existing LMS platforms for easy scaling.

Educators

For educators, it significantly reduces the time spent on manual plagiarism checks.

It provides transparent and explainable reports, making evaluations fair, consistent, and easy to justify.

Students

Our solutions helps students understand when and how to cite sources correctly.

It reduces unintentional plagiarism and misuse of AI tools by giving clear, actionable feedback before submission.

Students gain confidence knowing their work is original and ethically written.

Benefits

Social Benefits

- Promotes academic honesty and originality
- Encourages ethical use of AI tools
- Reduces plagiarism-related stress for students
- Builds trust between students and educators

Economic Benefits

- Low-cost alternative to premium plagiarism software
- Reduces faculty time spent on manual checking
- Scalable for institutions of all sizes
- Minimizes operational and licensing costs

Environmental Benefits

- Fully digital, paperless assignment evaluation
- Reduces physical submissions and printing
- Supports sustainable, eco-friendly education

Community & Educational Benefits

- Improves quality of learning outcomes
- Strengthens academic integrity culture
- Prepares students for ethical professional practices
- Supports fair and transparent evaluation systems

Improved User Experience & Accessibility

The interface is intuitive, submissions are quick, and results are delivered instantly.

Visual highlights make reports easy to interpret, and the system works reliably even on low-end devices and browsers.

Overall Impact

From punishment to prevention, plagiarism detection into a learning experience.

RESEARCH AND REFERENCES

OUR RESEARCH (LINK)

1. Autograding and Detecting Plagiarism in Student Programming Assignments :- (LINK)
2. A Source Code Similarity System for Plagiarism Detection (Đurić & Gašević, 2013) :- (LINK)
3. A systematic literature review on source code similarity measurement and clone detection :- (LINK)
4. Identifying Plagiarised Programming Assignments with Detection Tool Consensus (2023) :- (LINK)