

# Assignment Web Similarity Analysis

Generated on 2025-05-08 20:42:26

## Executive Summary

**Overall Web Similarity Score:** 15%

**Assessment:** The overall similarity between the student assignment and the provided web sources is low. While the assignment mentions concepts and techniques related to AI-generated text detection and plagiarism detection that are discussed in the web sources, there's no evidence of direct copying of substantial portions of text. The overlap primarily consists of common terminology and concepts within the field.

**Conclusion:** The assignment demonstrates a basic understanding of plagiarism detection and AI-generated content identification. While some concepts and terminology overlap with the web sources, there is no significant evidence of plagiarism. The project description appears to be the students' own formulation of the project scope, leveraging publicly available knowledge about the field. Further investigation might be warranted if the actual implementation shows a higher degree of similarity to specific sources. However, based on the project description alone, the likelihood of plagiarism is low.

## Web Sources Analyzed

Source URL	Similarity Score
https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-text/	41.7%
https://www.eastcentral.edu/free/ai-faculty-resources/detecting-ai-generated-text/	42.22%
https://www.pluralsight.com/resources/blog/ai-and-data/ai-generated-text-detection	50.65%
https://www.youtube.com/watch?v=FiFAefkHFaA&pp=0gcJCdgAo7VqN5tD	10.69%

## Detailed Content Matches

### Match 1 - Similar Content (20%)

**Assignment:** The AI Plagiarism Detector Tool aims to develop a system that uses advanced machine learning techniques to identify and prevent plagiarism and detect AI-generated content in student assignments.

**Source:** https://www.pluralsight.com/resources/blog/ai-and-data/ai-generated-text-detection

**Source Text:** As generative artificial intelligence (genAI) becomes more common in content creation, its ability to produce text that mirrors human writing is both exciting and unsettling.

### Match 2 - Common Knowledge (0%)

**Assignment:** Use techniques such as cosine similarity, word embedding (Word2Vec, GloVe), perplexity, and burstiness.

**Source:** None

**Source Text:** None

### Match 3 - Similar Content (25%)

**Assignment:** Create techniques to identify AI-generated assignments. Use proper measuring techniques to identify AI-generated text.

**Source:** https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-text/

**Source Text:** The internet is increasingly awash with text written by AI software. We need new tools to detect it.

## **Match 4 - Common Knowledge (0%)**

**Assignment:** Develop a user-friendly platform for students to upload documents. Offer fast feedback on plagiarism and AI-generated content scores.

**Source:** None

**Source Text:** None

---

# Full Assignment with Highlighted Plagiarism

*Sections highlighted in yellow with red text indicate potential plagiarism.*

Project Title

AI Plagiarism Detector Tool

Group Members

Sewwandi BTI

EG/2020/4210 ( Frontend developer )

Balasooriya JM

EG/2021/4424 ( UI/UX Designer )

Bandara KMTON

EG/2021/4432 ( Mobile APP Developer )

Bandara LRTD

EG/2021/4433 ( Backend Developer, Project Manager )

Contact Person

Dr. Kushan Sudheera (kushan@eie.ruh.ac.lk)

Desired Number of Group Members

: 4

Description

The AI Plagiarism Detector Tool aims to develop a system that uses advanced machine learning techniques to identify and prevent plagiarism and detect AI-generated content in student assignments. The tool will serve educational institutions by ensuring the originality and authenticity of submitted work.

Functional Requirements

1. Plagiarism Detection: Check for plagiarism on one or two-page papers. (Expand if feasible) ■ Use departmental assignment submission examples to test. Create algorithms for comparing and evaluating text similarities. ■ Modify any pre-built models as needed. To successfully compare, use techniques such as cosine similarity, word embedding (Word2Vec, GloVe), perplexity, and burstiness.
2. AI-Generated Content Detection: ■ ■ Create techniques to identify AI-generated assignments. ■ Use proper measuring techniques to identify AI-generated text. Use a publicly accessible large language model (LLM) for this purpose if necessary.
3. Submission Portal Development: ■ ■ Develop a user-friendly platform for students to upload documents. ■ Offer fast feedback on plagiarism and AI-generated content scores. Ensure the site displays both results for students following submission.
4. Report Generation: ■ ■ Create separate papers for AI and plagiarism detections. Ensure documents are downloadable for students and instructors.
5. Database Integration: ■ Create a database to hold all scores and related data. Optimize information retrieval and management. This is our project, and our primary functions are assignment to assignment similarity score and assignment to internet similarity score. Please advise me on how to complete that project and provide me with all necessary procedures, as well as this platform similar to Turnity.

# Analysis Methodology

---

**Web Similarity Analysis Method:** This report analyzes the similarity between a student assignment and web content using multiple approaches:

1. **Basic similarity analysis** using TF-IDF vectorization and cosine similarity metrics to calculate statistical similarity between texts.
2. **Advanced semantic analysis** using Google's Gemini AI to identify conceptual similarities, common phrases, and potential plagiarism patterns.
3. **Source verification** by analyzing multiple sources to distinguish between common knowledge and unique content.

## Interpretation Guide:

- 0-15%: Very low similarity - Likely original content
- 16-30%: Low similarity - Contains common phrases but largely original
- 31-50%: Moderate similarity - May contain some paraphrased content
- 51-70%: High similarity - Contains substantial similar content
- 71-100%: Very high similarity - Significant portions may be unoriginal

---

*Disclaimer: This automated similarity analysis provides an approximation of content similarity against web sources. Results should be interpreted by a human reviewer for context-appropriate assessment. Common knowledge, standard phrases, and coincidental matches may be flagged and require human judgment.*