# AI Based Plagarism Checker

Plagarism:-It is basically when someone tries to use someone else’s work without any proper permission or Attribution,credits or citation.

* In Plagarism Detector App we usually **give input in terms of text** and using AI based app it checks for plagiarism and **displays output as plagiarism** **detected if content is not original**.

### ****How AI-Based Plagiarism Checkers Work****

1. **Upload the Text**
   * You enter or upload the document you want to check.
2. **Text Analysis**
   * The AI **breaks down the text** into words and sentences.
   * It **removes extra spaces and symbols** to clean the text.
3. **Compare with Other Sources**
   * The AI checks the text against **millions of online sources** like books, websites, and research papers.
   * It looks for **exact matches and reworded (paraphrased) content**.
4. **Detect Plagiarism**
   * The tool **highlights copied parts** and shows where they came from.
   * It also finds **paraphrased text** (when words are changed but the meaning stays the same).
5. **Generate a Report**
   * The AI **shows a plagiarism percentage** (e.g., 20% copied).
   * It provides **source links** and suggests how to fix copied content.

### ****Why AI Plagiarism Checkers Are Useful?****

✅ **Detects exact and reworded copies**  
✅ **Compares with millions of sources**  
✅ **Gives a detailed plagiarism report**  
✅ **Helps in proper citation**

Algorithm and its uses

### ****1. BLEU Score (Bilingual Evaluation Understudy)****

🔹 **What it does**: Compares text based on small word sequences (**n-grams**) to measure similarity.  
🔹 **Use case**: Detects direct copying of phrases.  
🔹 **Example**:

* **Original**: "The cat is sleeping."
* **Copied**: "The cat is resting."
* BLEU will **give a high score** because "The cat is" matches.

### ****2. ROUGE Score (Recall-Oriented Understudy for Gisting Evaluation)****

🔹 **What it does**: Counts overlapping words between two texts.  
🔹 **Use case**: Checks similarity in summaries or paraphrased content.  
🔹 **Example**:

* **Original**: "AI is transforming technology."
* **Paraphrased**: "Technology is being transformed by AI."
* ROUGE detects **word matches** and gives a similarity score

### ****3. Perplexity Score****

🔹 **What it does**: Measures how "natural" a sentence is based on a language model.  
🔹 **Use case**: Helps detect AI-generated or rewritten text.  
🔹 **Example**:

* **Human-written**: "The weather is pleasant today."
* **AI-generated**: "Climatic conditions exhibit favorable characteristics”

### ****4. MRR (Mean Reciprocal Rank)****

🔹 **What it does**: Ranks search results based on relevance.  
🔹 **Use case**: Helps find the **most likely source** of plagiarism.

Libraries

### ****1. For Text Similarity (Detecting Copy-Paste & Paraphrasing)****

📌 **NLTK** – Basic text processing (tokenization, stemming)  
📌 **spaCy** – Advanced NLP processing (word embeddings)  
📌 **Scikit-learn** – Cosine similarity, TF-IDF for checking word overlap  
📌 **Sentence Transformers (BERT/SBERT)** – Deep learning-based text similarity

### ****2. For Plagiarism Metrics (BLEU, ROUGE, etc.)****

📌 **nltk.translate.bleu\_score** – For **BLEU score** (copy-paste detection)  
📌 **rouge-score** – For **ROUGE score** (paraphrase detection)  
📌 **textstat** – For readability and perplexity scoring  
📌 **Hugging Face Transformers** – For GPT-based perplexity scoring

### ****3. For Searching Plagiarized Content Online****

📌 **SerpAPI / Google Search API** – To find **exact matches on the web**  
📌 **Elasticsearch** – If you want to create your own **plagiarism search engine**