

How the DocuVerify System Works

The DocuVerify system is a comprehensive web application designed to detect plagiarism and AI-generated content in uploaded documents.

Below is a detailed explanation of how the system works—from account creation to document analysis—including logic behind detection and the React-based frontend.

Links

deployed: <https://plagiarismaidetector-nine.vercel.app>

Youtube video link: https://youtu.be/eT_C8uvWqHY

1. Document Upload & Analysis

Uploading a Document

Endpoint: POST /analyze/

File content extracted using:

```
py
extract_text_from_file() // in /documents/utlis.py
```

2. Plagiarism Detection

Logic

Detection is handled by:

```
analyze_text() in documents/utlis.py
```

Steps:

- Vectorization: TFIDF with 5gram character analysis.
- Cosine Similarity: Compared against other documents.
- Highlighting: Texts exceeding threshold (e.g., 0.3) are marked.

3. AI Content Detection

Logic

Detection via:

```
py  
check_ai_probability() // in /documents/utils.py
```

Steps:

Chunking: Splits text into 512char chunks.

Model: Uses HelloSimpleAI/chatgptdetectorroberta (Hugging Face).

Scoring: AI confidence score per chunk.

Highlighting: Flags AIgenerated segments.

The DocuVerify system combines:

A secure backend for authentication and analysis.

A userfriendly frontend built in React.

Advanced techniques:



TFIDF + Cosine Similarity for plagiarism.



Transformer models for AI content detection.