

# AI vs Human: Academic Essay Authenticity Challenge

Arush U Rai<sup>1</sup>, Srushti S K<sup>2</sup>, Pranitha R Shekar<sup>3</sup>, Hemanth S K<sup>4</sup>, Asst. Prof. Ms. Kayal Vizhi V<sup>5</sup>  
<sup>1,2,3,4,5</sup>Dept. of Computer Science & Engineering, Presidency University, Bengaluru, India

**Abstract**—This paper presents the development and evaluation of an AI-powered system designed to identify whether a given essay is AI generated or human-written. The growing sophistication of natural language generation models, such as OpenAI's GPT series, has led to a surge in content that closely mimics human writing styles. This rapid advancement has raised critical questions about authenticity, intellectual integrity, and the trustworthiness of textual material across various industries. Our proposed solution integrates a Flask backend framework with AWS Textract for text extraction and OpenAI's GPT for language analysis. It provides a user-friendly interface where individuals can upload images of essays—handwritten or printed—then automatically extracts and classifies their textual content.

**Index Terms**—Essay Validator, AI Detection, Flask Framework, AWS Textract, OpenAI GPT, Text Extraction, Image Upload, Natural Language Processing, Image Processing.

## I. INTRODUCTION

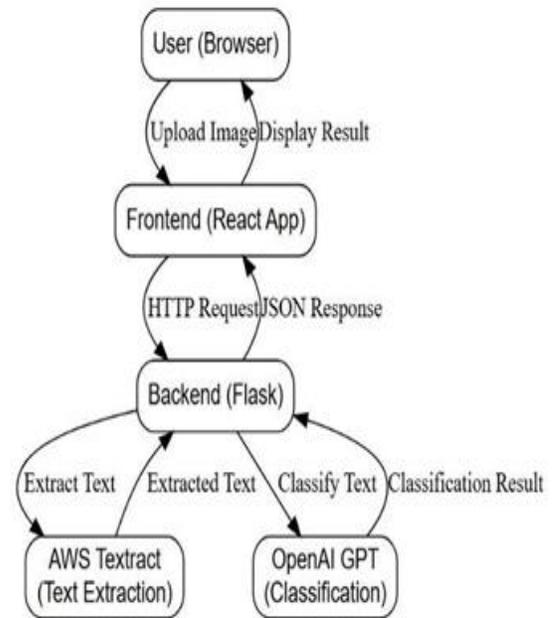
In recent years, the rise of artificial intelligence, particularly in natural language processing (NLP), has significantly impacted how we create, consume, and evaluate text-based content. Models such as GPT-3 and GPT-4 can now generate essays, articles, and creative writing pieces that closely resemble human-authored text. This advancement, while remarkable, poses a novel challenge: how do we distinguish machine-generated text from genuinely human-crafted content? The implications of this challenge span multiple domains. In educational contexts, ensuring that students produce original work is paramount to maintaining academic integrity. The infiltration of AI-generated text into essays and reports may misrepresent a student's true understanding and effort. In journalism and media, verifying the authenticity of

### 1. Backend (Flask Framework):

The backend serves as the core orchestrator.

sources and articles is vital for preserving the credibility of the press. Similarly, in corporate and governmental communications, authenticity ensures that strategic documents and public announcements genuinely reflect human decision-making processes rather than automated outputs.

## II. PROPOSED SYSTEM ARCHITECTURE



The AI vs Human: Essay Validator application consists of three primary components: the backend, the frontend, and AI integration. These components work in tandem to deliver a seamless user experience for essay validation.

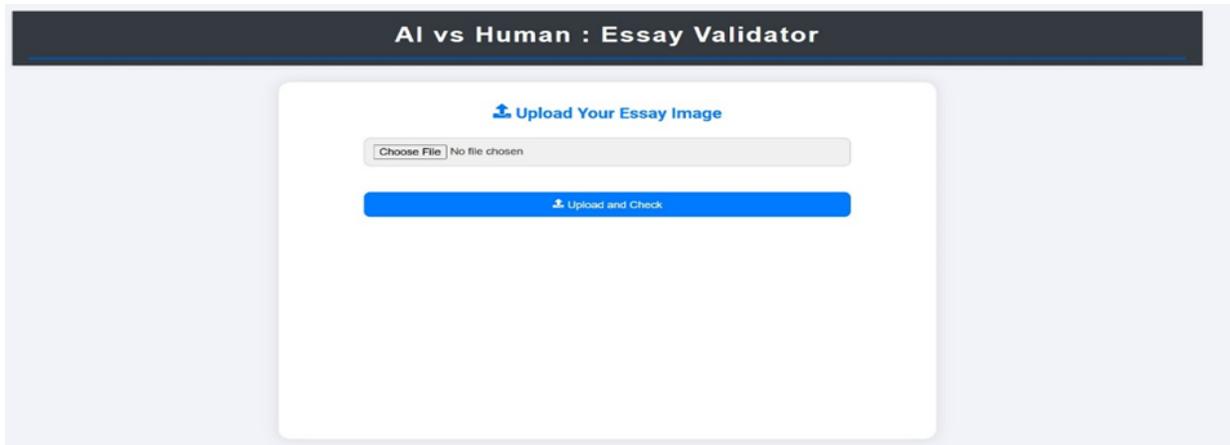
```
PS C:\Users\heasi\OneDrive\Desktop\essay-auth> python app.py
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use
a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 824-472-999
```

- Routing and API Communication: Flask routes incoming requests from the frontend, delegating tasks such as image processing, text extraction, and classification. Its lightweight footprint ensures rapid response times and easy scalability.
- Integration with AWS Textract: The backend leverages AWS Textract's APIs to transform uploaded images into machine-readable text. This

process abstracts away complexities in OCR and ensures reliable text extraction even from low-quality or complex documents.

- OpenAI GPT-3 Classification: Once the text is available, the backend invokes the OpenAI API to classify the text. GPT-3 provides a context-aware evaluation, identifying subtle linguistic markers indicative of AI authorship.

## 2. Frontend (React App):



The frontend acts as the user's gateway, focusing on clarity, responsiveness, and usability.

- User-Friendly Interfaces: By using React, the system can handle dynamic state updates, offer drag-and-drop file uploads, and provide progress indicators during the processing pipeline.
- Display of Extracted Text & Results: After classification, users can review the extracted text and its classification (AI-generated or human-written) instantly. This direct feedback loop builds user confidence and trust in the system's results.

## 3. AI Integration (OpenAI GPT-3):

The classification model is the system's intelligent core.

- Sophisticated NLP: GPT-3's extensive training on diverse text corpora allows it to sense patterns in sentence structure, word choice, topic transitions, and stylistic features that might be unnatural or too structured JSON, enabling seamless parsing and integration into other applications or services. This makes the system's capabilities easy to extend beyond the current frontend. The entire architecture is designed to be cloud-ready, capable of running on scalable server infrastructures, and easily integrated into third-party platforms through RESTful APIs. This ensures that as the demand for authenticity checks grows, the system can scale and evolve without fundamental redesign.

### III. METHODOLOGY

The methodology underlying the AI vs Human: Essay Validator aims to streamline the user journey from initial essay upload to final classification result:

- **Image Upload:** Users begin by uploading images that may contain handwritten or typed text. The variety of input forms (scans, screenshots, phone-captured images) ensures the tool’s applicability across diverse scenarios. This step abstracts away user technicalities.
- **Text Extraction via AWS Textract:** The system calls AWS Textract to perform OCR on the uploaded image. Textract’s advanced algorithms handle skewed text, varying fonts, and complex layouts, returning clean, structured text suitable for NLP analysis. By offloading OCR complexity to a robust, cloud-based service, the methodology ensures both reliability and performance.
- **AI Classification with OpenAI GPT-3:** Once the text is extracted, it is submitted to GPT-3 for classification. Unlike simpler heuristic methods (e.g., perplexity analysis), GPT-3 leverages its deep learning backbone to identify nuanced markers of AI generation.

### IV. COMPARISON WITH RELATED RESEARCH

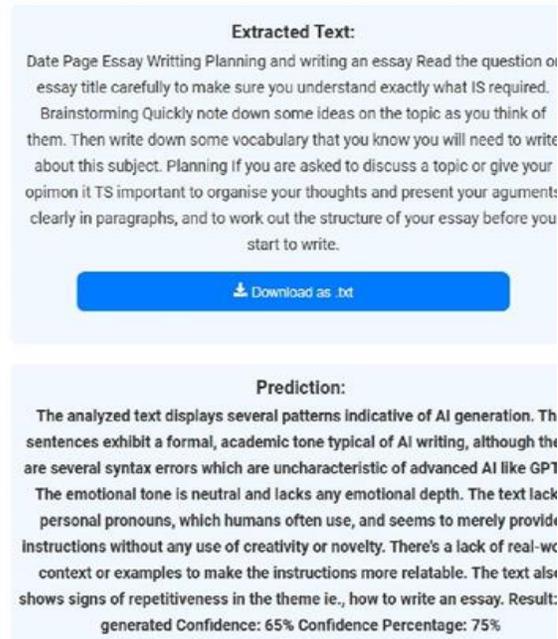
Detecting AI-generated text is an evolving research frontier. Early methods relied on simple stylistic checks—counting rare words, analyzing n-gram distributions, or measuring perplexity.

In contrast, our solution stands apart in several ways:

- **Comprehensive Input Handling:** Many existing detection tools assume clean digital text input. Our approach handles images, broadening the range of materials that can be analyzed.
- **Deep Integration of State-of-the-Art Models:** By leveraging GPT-3, a cutting-edge model, the system has an inherent advantage over methods limited to older or less capable NLP frameworks.
- **Real-Time and Scalable Infrastructure:** Traditional research prototypes often run offline and lack the scalability for real-time use.
- **Holistic User Experience:** Beyond detection accuracy, our system prioritizes user experience, providing intuitive interfaces and meaningful,

instantaneous feedback. Whereas prior research has often focused on niche language models or limited datasets, the Essay Validator’s use of commercially available, widely acknowledged services like Textract and GPT-3 ensures a solution that is not only academically interesting but also practically deployable.

### V. RESULT ANALYSIS



Preliminary testing on a diverse dataset of essays—some generated by well-known AI models, others written by students and educators demonstrated promising results:

- **Classification Accuracy:** Early evaluations show a high true-positive rate for AI-generated content and a low false-positive rate for human-written texts.
- **Robustness to Noise and Variability:** Even when given low-quality images or texts with unusual layouts, AWS Textract extracted text reliably, enabling the downstream classifier to work effectively.
- **User Studies and Feedback:** Informal surveys with educators and editorial staff indicated a positive reception. Users appreciated the transparency of being able to see the extracted text and the immediate classification result.

## VI. FUTURE WORK AND IMPLICATIONS

The challenges of detecting AI-generated text are not static; they evolve as language models improve and their outputs become ever more human-like. Hence, our roadmap includes:

- **Multi-Language Support:**  
Adding support for languages beyond English would broaden the tool's global applicability. This expansion involves training or selecting language specific models and integrating language detection capabilities.
- **Hybrid Approaches:**  
Combining the deep-learning-based approach with traditional linguistic theory could yield a hybrid model that is more interpretable. For instance, examining narrative coherence, logical argumentation patterns, and cultural references might offer richer insights.

- **Explainability and Transparency:** Future versions may include explainable AI (XAI) features. By highlighting specific phrases or linguistic markers that informed the classification, we would enhance user trust and educational value, helping teachers guide students in their writing improvement.
- **Integration with Integrity and Plagiarism Tools:** Incorporating plagiarism checks would create a one-stop solution for educators and content verifiers. Authenticity checks could also be melded with style analysis, authorship attribution, and sentiment evaluation for more comprehensive textual insights.
- **Adapting to Future AI Models:** As large language models become more sophisticated, continuous updates are required. Periodic fine-tuning against newly released generative models will keep the tool relevant

## VII. OUTPUTS

The screenshot displays the 'AI vs Human : Essay Validator' web application. At the top, the title 'AI vs Human : Essay Validator' is centered in a dark blue header. Below the header, the main interface features a central white box with a blue 'Upload Your Essay Image' button. A file selection area shows 'Choose File 1.jpg'. A large blue 'Upload and Check' button is positioned below the file selection. Below the main interface, the prediction results are shown in two light blue boxes. The left box, titled 'Extracted Text:', contains the text of an essay about AI in healthcare. The right box, titled 'Prediction:', provides a detailed analysis of the text, identifying formal language usage, error-free syntax, and a consistent logical flow of ideas, concluding that the text is AI-generated with a confidence of 85% and a confidence percentage of 75%.

**AI vs Human : Essay Validator**

Upload Your Essay Image

Choose File 1.jpg

Upload and Check

Prediction completed

**Extracted Text:**

Essay 1: The Impact of Artificial Intelligence on Healthcare Artificial Intelligence (AI) has significantly transformed the healthcare industry, offering innovative solutions to various challenges. AI-powered systems can analyze vast amounts of medical data and detect patterns that may go unnoticed by human doctors. These systems assist in diagnosing diseases, predicting patient outcomes, and recommending personalized treatments. Additionally, AI-driven robots and automated machines have revolutionized surgery, allowing for more precise and less invasive procedures. However, the integration of AI in healthcare raises concerns about data privacy and the potential for machine errors. Despite these challenges, AI holds immense promise in enhancing healthcare efficiency and accessibility, paving the way for improved patient care. As AI continues to evolve, it will likely play an even more prominent role in the healthcare sector, further optimizing treatments and patient management.

Download as .txt

**Prediction:**

Analyzing the given text for the enlisted factors, I see formal language usage, error-free syntax, neutral emotional tone, and a consistent and logical flow of ideas. The lack of personal pronouns, stylistic inconsistency, and real-world specific references further suggest machine-generated content. A theme of repetitive exploration of AI's impact on healthcare – an AI-centric topic written from an objective standpoint also hints at AI origin. The argument does not present deep emotional resonance or personal bias, and it includes well-structured, complex sentences, which is typical of AI text generation rather than human writing. Result: AI-generated, Confidence: 85%. Confidence Percentage: 75%

REFERENCES

- [1] OpenAI, "GPT-3 API Documentation", OpenAI, accessed December 2024.
- [2] Amazon Web Services, "Amazon Textract", AWS Textract, accessed December 2024.
- [3] Pallets Projects, "Flask Web Framework", Flask, accessed December 2024.
- [4] React Team, "React – A JavaScript library for building user interfaces", React, accessed December 2024.
- [5] Amazon Web Services, "AWS Command Line Interface", AWS CLI, accessed December 2024.
- [6] Armin Ronacher, "python-dotenv", Python Dotenv, accessed December 2024.
- [7] Pallets Projects, "Werkzeug - The Python WSGI Utility Library", Werkzeug, accessed December 2024.
- [8] Axios, "Axios – Promise-based HTTP client for the browser and node.js", Axios, accessed December 2024.
- [9] Fonticons, Inc., "Font Awesome - The iconic font and toolkit", Font Awesome, accessed December 2024