Academic Integrity Report

Comprehensive Plagiarism and AI Content Analysis

File: **nearest summer training report.docx**

Generated on August 27, 2025 at 02:39 AM

Student/User: **katendeelie411-20250806085846**

Document ID: **41**

Document Information

File Name: nearest summer training report.docx

File Size: 934.5 KB

Upload Date: August 27, 2025

Analysis Date: August 27, 2025

Analysis Results

Plagiarism Detection

15.8%

Similarity Found

Total Words: 4029
Matched Words: 636

AI Content Detection

37.4%

AI Generated

Total Words: 4029
AI Words: 1506

Analysis Type	Score	Issues Found	Risk Level
Plagiarism Detection	15.8%	248 instances	**Medium Risk**
AI Content Detection	37.4%	248 instances	**High Risk**

Document Analysis with Highlighted Issues

Plagiarism Detected (248)

AI Generated Content (248)

GIRNE AMERICAN UNIVERSITY Faculty of Engineering Department of Computer Engineering AcadCheck: AI and Plagiarism Detection Tool SUMMER TRAINING REPORT Student: Elie katende kazwela Student ID: 221701001 Supervisor: Mr. GIRNE AMERICAN UNIVERSITY Faculty of Engineering Department of Computer Engineering AcadCheck: AI and Plagiarism Detection Tool SUMMER TRAINING REPORT Student: Elie katende kazwela Student ID: 221701001 Supervisor: Mr. Cliff Acknowledgements I would like to sincerely thank all those who supported me during my internship at the university and throughout the preparation of this report. Cliff Acknowledgements I would like to sincerely thank all those who supported me during my internship at the university and throughout the preparation of this report. First and foremost , I am especially grateful to my supervisor , Mr. First and foremost , I am especially grateful to my supervisor , Mr. Peter Cliff, for his guidance ,encouragement and constant support. Peter Cliff, for his guidance ,encouragement and constant support. His expertise and thoughtful feedback played a key role in the progress and success of my work. His expertise and thoughtful feedback played a key role in the progress and success of my work.

I also wish to express my appreciation to the staff and faculty of the university for welcoming me , Providing the necessary resources, and creating a positive environment that made this internship a valuable learning experience. I also wish to express my appreciation to the staff and faculty of the university for welcoming me , Providing the necessary resources, and creating a positive environment that made this internship a valuable learning experience. the knowledge and skills I gained have been highly beneficial to my academic and professional growth. the knowledge and skills I gained have been highly beneficial to my academic and professional growth.

Lastly ,I would like to thank my family and friends for their patience , encouragement , and continuous Support , which greatly helped me in completing this project. Lastly ,I would like to thank my family and friends for their patience , encouragement , and continuous Support , which greatly helped me in completing this project. Abstract This report introduces AcadCheck, a web application developed to detect both plagiarism and AI generated content in academic documents . Abstract This report introduces AcadCheck, a web application developed to detect both plagiarism and AI generated content in academic documents . the system was implemented using React, TypeScript, and Supabase, and it integrates advanced document analysis methods with a user-friendly interface to deliver precise and comprehensive results. the system was implemented using React, TypeScript, and Supabase, and it integrates advanced

document analysis methods with a user-friendly interface to deliver precise and comprehensive results. The main objectives of the project are structured around four core component : Plagiarism detection : comparing documents against a large corpus by applying N-grams, semantic similarity, and pattern-matching techniques. AI content detection – identifying AI-generated text through transformer-based models and linguistic analysis. Document analysis and reporting – producing detailed, sentence-level feedback with metrics such as lexical diversity, syntactic complexity, semantic coherence, and perplexity. Secure and scalable architecture – leveraging Supabase for database management, data protection, and Row Level Security (RLS). This report presents the overall architecture , development process ,challenges ,faced and evaluation results of the application ,while also reflecting on the ethical considerations involved .it highlights the use of modern approaches to document analysis , the integration of artificial intelligence in plagiarism detection ,and the design of scalable web applications. Chapter 1: Introduction Background In academic settings, maintaining the integrity of written work is essential. The growing use of digital tools, online resources, and AI-assisted content creation has heightened the risk of plagiarism and the improper use of external materials. AcadCheck tackles these issues by offering an automated solution that can identify both plagiarized content and text generated by artificial intelligence, helping to uphold academic honesty and the quality of scholarly work.

generated by artificial intelligence, helping to uphold academic honesty and the quality of scholarly work. Contemporary document analysis relies on advanced technologies such as machine learning, semantic evaluation, and natural language processing (NLP). By combining these methods with a straightforward and intuitive web interface, AcadCheck enables educators, students, and researchers to examine documents effectively while ensuring the security and privacy of their data.

1.2 Objectives The primary goals of this internship project are: To create a web application capable of identifying both plagiarism and AI-generated content. To integrate sophisticated analysis techniques, combining machine learning with heuristic methods. To provide comprehensive reporting, including analysis at the sentence level, confidence indicators, and clear visual cues. To guarantee secure management of user data, leveraging authentication, Row-Level Security (RLS), and protected storage. To deliver a user-friendly and responsive interface that accommodates multilingual users effectively.

1.3 Scope This project centers on the design, development, and implementation of AcadCheck as a tool for research and education. The application is compatible with PDF, DOCX, and TXT file formats, and it offers detailed sentence-level analysis for both plagiarism and AI-generated content. All testing and usage of the application are carried out ethically, within controlled environments, and in strict accordance with data protection regulations. The project does not involve

deployment in untrusted environments or the collection of unauthorized data. The project does not involve deployment in untrusted environments or the collection of unauthorized data.

1.4 Importance of the Study AcadCheck contributes to multiple domains: Academic integrity: Helps institutions maintain originality standards.

AI literacy: Educates users on distinguishing AI-generated content.

Technical education: Demonstrates the use of modern web technologies, machine learning, and database security in practical applications.

By integrating AI detection with plagiarism analysis, the project addresses modern challenges in digital content evaluation, providing insights for students, educators, and software developers.

1.5 Structure of the Report The organization of the report is: Chapter 1: Introduction – Presents background, objectives, scope, and significance.

Chapter 2: Literature Review – Reviews plagiarism detection methods, AI content detection, and ethical/legal considerations.

Chapter 3: Methodology – Details the system architecture, tools, libraries, and implementation steps.

Chapter 4: Implementation and Discussion – Covers code implementation, document processing, analysis workflow, and troubleshooting.

Chapter 5: Conclusion – Summarizes findings, limitations, educational value, and future improvements.

References – List of sources cited throughout the report.

CHAPTER 2: LITERATURE REVIEW 2.1 Overview of Plagiarism Detection Detecting plagiarism is essential for upholding academic integrity.

academic integrity.

Traditional tools mainly rely on exact text matching, which can spot direct copying but often misses paraphrased or slightly altered content. Modern systems, like AcadCheck, use a combination of syntactic, semantic, and contextual analysis, enabling more accurate detection of potential plagiarism.

N-gram Analysis: Documents are broken into sequences of words (n-grams) and compared to a reference corpus. Similar sequences indicate potential plagiarism. AcadCheck uses adaptive n-gram sizes to capture both short phrases and longer patterns.

Semantic Similarity: Using NLP techniques, AcadCheck compares the meaning of sentences rather than exact wording. This allows the system to detect paraphrased content. Contextual Patterns: Certain academic formulations are commonly reused. By identifying these patterns, the system can differentiate between general academic phrases and copied unique content. Recent research emphasizes the integration of machine learning models for more nuanced plagiarism detection, allowing systems to learn from labeled datasets and improve detection accuracy over time.

2.2 AI-Based Text Detection With the emergence of AI-generated content, particularly large language models like GPT, distinguishing between human-written and AI-generated text has become a crucial challenge.

challenge. AcadCheck implements hybrid AI detection techniques, combining transformer-based models with linguistic heuristics: Transformer Models: Pre-trained models analyze text for patterns indicative of AI generation, such as overuse of certain connectors, predictable phrasing, or low variability. Linguistic Heuristics: Metrics such as lexical diversity, syntactic complexity, and perplexity are used to evaluate the naturalness of text. AI-generated text often exhibits high perplexity uniformity and low lexical diversity. Sentence-Level Scoring: Each sentence is assigned an AI score, indicating the likelihood of machine generation. This granularity allows users to review suspicious content precisely. By integrating AI detection with plagiarism analysis, AcadCheck provides a comprehensive evaluation of document originality. 2.3 Document Analysis Techniques Effective plagiarism and AI detection require robust document analysis. AcadCheck supports multiple formats, including PDF, DOCX, and TXT, with consistent text extraction: PDF Analysis: Using PDF.js, the system extracts text while preserving document structure. Word Document Handling: Mammoth and DOCX Preview convert DOCX files to HTML, ensuring accurate rendering and sentence-level analysis. Text Preprocessing: Cleaning, tokenization, and normalization prepare the content for further analysis.

Sentence segmentation, feature extraction, and vectorization are applied before any similarity or AI detection algorithms, ensuring that results are both accurate and interpretable. Sentence segmentation, feature extraction, and vectorization are applied before any similarity or AI detection algorithms, ensuring that results are both accurate and interpretable. 2.4 Ethical and Legal Considerations Given the sensitivity of plagiarism and AI detection, ethics and legality are central to AcadCheck's design: User Consent: Only documents uploaded with explicit consent are analyzed. 2.4 Ethical and Legal Considerations Given the sensitivity of plagiarism and AI detection, ethics and legality are central to AcadCheck's design: User Consent: Only documents uploaded with explicit consent are analyzed. Data Security: Supabase provides encrypted storage and Row Level Security (RLS) to isolate user data. Data Security: Supabase provides encrypted storage and Row Level Security (RLS) to isolate user data.

Legal Compliance: The system aligns with GDPR and other data protection regulations, preventing unauthorized access or retention. Legal Compliance: The system aligns with GDPR and other data protection regulations, preventing unauthorized access or retention. Responsible Reporting: Results are presented solely for educational or institutional purposes, avoiding misuse of the analysis. Responsible Reporting: Results are presented solely for educational or institutional purposes, avoiding misuse of the analysis. Ethical considerations also guide the inclusion of AI detection. Ethical considerations also guide the inclusion of AI detection. The system is designed for transparency, providing users with interpretable metrics and avoiding punitive assumptions. The system is designed for transparency, providing users with interpretable metrics and avoiding punitive assumptions. 2.5 Use of React, TypeScript, and Supabase in Web Apps Modern web technologies play a key role in the functionality and usability of AcadCheck: React (v18.3.1): Provides a responsive, dynamic frontend capable of handling real-time document previews and interactive dashboards.

2.5 Use of React, TypeScript, and Supabase in Web Apps Modern web technologies play a key role in the functionality and usability of AcadCheck: React (v18.3.1): Provides a responsive, dynamic frontend capable of handling real-time document previews and interactive dashboards. TypeScript: Adds type safety and improves maintainability, especially important for complex components such as analysis tables and reports. TypeScript: Adds type safety and improves maintainability, especially important for complex components such as analysis tables and reports. Supabase: Serves as a backend and database solution, offering authentication, secure storage, and serverless functions. Supabase: Serves as a backend and database solution, offering authentication, secure storage, and serverless functions. RLS ensures that users only access their own

analyses. RLS ensures that users only access their own analyses. Frontend UI Components: Using Tailwind CSS and shadcn/ui, the interface is modern, clean, and intuitive. Frontend UI Components: Using Tailwind CSS and shadcn/ui, the interface is modern, clean, and intuitive. Components like DocumentViewer and HighlightedText provide detailed visual feedback to users. Components like DocumentViewer and HighlightedText provide detailed visual feedback to users. The combination of these technologies allows for a robust, scalable, and secure web application, supporting complex AI and plagiarism detection workflows while maintaining excellent user experience. The combination of these technologies allows for a robust, scalable, and secure web application, supporting complex AI and plagiarism detection workflows while maintaining excellent user experience. 2.6 Integration of Detection Algorithms in Web Applications A critical aspect of modern plagiarism detection systems is the seamless integration of backend analysis algorithms with frontend visualization: Sentence-Level Analysis: Each sentence is analyzed for plagiarism and AI content. 2.6 Integration of Detection Algorithms in Web Applications A critical aspect of modern plagiarism detection systems is the seamless integration of backend analysis algorithms with frontend visualization: Sentence-Level Analysis: Each sentence is analyzed for plagiarism and AI content. Results are transmitted from the backend to the frontend, where they are visually highlighted in the document. Results are transmitted from the backend to the frontend, where they are visually highlighted in the document.

Scoring and Metrics: Global scores, confidence levels, and detailed metrics are presented using interactive dashboards. Scoring and Metrics: Global scores, confidence levels, and detailed metrics are presented using interactive dashboards.

Real-Time Feedback: Users can upload documents and receive analysis results without page reloads, thanks to React's component architecture. Real-Time Feedback: Users can upload documents and receive analysis results without page reloads, thanks to React's component architecture.

This integration ensures that technical sophistication in analysis does not compromise usability. This integration ensures that technical sophistication in analysis does not compromise usability. 2.7 Summary Chapter 2 shows that modern plagiarism and AI detection relies on multi-layered methods that combine semantic analysis, machine learning, and secure web technologies. 2.7 Summary Chapter 2 shows that modern plagiarism and AI detection relies on multi-layered methods that combine semantic analysis, machine learning, and secure web technologies. AcadCheck applies these principles by providing a system that is easy to use, secure, and accurate, effectively tackling today's challenges in maintaining academic integrity. AcadCheck applies these

principles by providing a system that is easy to use, secure, and accurate, effectively tackling today's challenges in maintaining academic integrity. CHAPTER 3: METHODOLOGY 3.1 Design Objectives and System Architecture The main goal in designing AcadCheck is to build a web application that is secure, scalable, and easy to use for detecting plagiarism and AI-generated content. The key aspects taken into account include: Accuracy: Implement advanced algorithms for precise detection of plagiarized and AI-generated text. Security: Ensure strict user data isolation through authentication and Row Level Security (RLS). Usability: Provide an intuitive dashboard and interactive reports that can be easily interpreted by students, educators, and researchers. Scalability: Design the system to handle large volumes of document uploads and concurrent analyses.

The system architecture is composed of the following layers: Frontend (React + TypeScript): Handles user interaction, document preview, and visualization of analysis results. Backend (Supabase + Serverless Functions): Processes document uploads, manages user authentication, and executes analysis algorithms. Database (PostgreSQL): Stores user data, documents, analysis results, and sentence-level metrics. AI Engine: Implements transformer-based models and linguistic heuristics for AI detection. Security Layer: Enforces RLS, secure storage, and access control to protect sensitive data. The architecture is modular, allowing future integration of additional features such as multi-language support, paraphrasing detection, and integration with LMS platforms.

features such as multi-language support, paraphrasing detection, and integration with LMS platforms. 3.2 Tools and Libraries The following tools and libraries were selected for their suitability and robustness: Frontend: React 18.3.1 for component-based UI TypeScript for type safety and maintainability Tailwind CSS and shadcn/ui for consistent, responsive styling React Router DOM for client-side routing Backend & Database: Supabase (PostgreSQL) for secure user management, storage, and database operations Serverless functions for document processing Document Processing: PDF.js for PDF extraction Mammoth and DOCX Preview for Word document conversion and rendering Custom fileToText library for preprocessing and normalization AI & Plagiarism Detection: Hugging Face Transformers for AI detection N-gram analysis, semantic similarity, and heuristic algorithms for plagiarism detection Testing & Quality: Jest and React Testing Library for unit tests Cypress for integration tests ESLint, Prettier, and Husky for code quality and consistency 3.3 Development and Implementation The development of AcadCheck was divided into several modules to ensure modularity and maintainability. 3.2 Tools and Libraries The following tools and libraries were selected for their suitability and robustness: Frontend: React 18.3.1 for component-based UI TypeScript for type safety and maintainability Tailwind CSS and shadcn/ui for consistent, responsive styling React Router DOM for client-side routing Backend & Database: Supabase (PostgreSQL) for secure user management, storage, and database operations Serverless functions for document processing Document Processing: PDF.js for PDF extraction Mammoth and DOCX Preview for Word document conversion and rendering Custom fileToText library for preprocessing and normalization AI & Plagiarism Detection: Hugging Face Transformers for AI detection N-gram analysis, semantic similarity, and heuristic algorithms for plagiarism detection Testing & Quality: Jest and React Testing Library for unit tests Cypress for integration tests ESLint, Prettier, and Husky for code quality and consistency 3.3 Development and Implementation The development of AcadCheck was divided into several modules to ensure modularity and maintainability. 3.3.1 User Authentication and Session Management Sign-up/Login: Users register with email and password. 3.3.1 User Authentication and Session Management Sign-up/Login: Users register with email and password. Supabase handles authentication securely. Supabase handles authentication securely. Session Persistence: JWT tokens are stored in browser local storage, with automatic session renewal. Session Persistence: JWT tokens are stored in browser local storage, with automatic session renewal. Route Protection: Only authenticated users can access dashboards and analysis

features. Route Protection: Only authenticated users can access dashboards and analysis features. 3.3.2 Document Upload and Preprocessing Supported Formats: PDF, DOCX, and TXT File Validation: Ensures correct format and maximum file size compliance Preprocessing Steps: Text extraction from uploaded files Sentence segmentation and tokenization Removal of unnecessary characters or formatting issues 3.3.3 Plagiarism Detection Algorithms Exact Matching: Detects copied content using n-grams (3-5 words). 3.3.2 Document Upload and Preprocessing Supported Formats: PDF, DOCX, and TXT File Validation: Ensures correct format and maximum file size compliance Preprocessing Steps: Text extraction from uploaded files Sentence segmentation and tokenization Removal of unnecessary characters or formatting issues 3.3.3 Plagiarism Detection Algorithms Exact Matching: Detects copied content using n-grams (3-5 words). Semantic Similarity: Compares sentences against reference corpus using vector embeddings. Semantic Similarity: Compares sentences against reference corpus using vector embeddings. Contextual Analysis: Identifies repeated academic phrases to minimize false positives. Contextual Analysis: Identifies repeated academic phrases to minimize false positives.

3.3.4 AI Content Detection Transformer Models: Analyzes text for patterns indicative of AI generation. 3.3.4 AI Content Detection Transformer Models: Analyzes text for patterns indicative of AI generation. Stylistic Metrics: Computes lexical diversity, syntactic complexity, semantic coherence, and perplexity. Stylistic Metrics: Computes lexical diversity, syntactic complexity, semantic coherence, and perplexity. Sentence-Level Scoring: Each sentence receives a probability score for AI authorship. Sentence-Level Scoring: Each sentence receives a probability score for AI authorship. 3.3.5 Analysis Report Generation Visual Dashboard: Displays sentence-level highlights and overall metrics. 3.3.5 Analysis Report Generation Visual Dashboard: Displays sentence-level highlights and overall metrics. Export Options: Users can download PDF reports with annotated content. Export Options: Users can download PDF reports with annotated content. Confidence Scores: Metrics such as AI score, plagiarism score, and confidence levels guide interpretation. Confidence Scores: Metrics such as AI score, plagiarism score, and confidence levels guide interpretation. 3.3.6 Internationalization Languages Supported: English and French Dynamic Switching: Users can switch languages without refreshing the page Persistence: Language preferences saved in user profile 3.3.7 Security and RLS Policies Row Level Security (RLS): Ensures users can only access their own analyses Encrypted Storage: All uploaded documents are stored securely in Supabase buckets Access Control: Temporary signed URLs prevent unauthorized access to stored files 3.4 Packaging and Deployment Frontend Build: Using

Vite for optimized bundle creation Backend Deployment: Supabase handles serverless functions and database management Environment Variables: Secured for production, including Supabase API keys and AI model endpoints CI/CD Pipeline: Automated deployment with GitHub Actions and version control 3.5 Testing and Quality Assurance Unit Tests: Validate individual components and algorithm functions Integration Tests: Ensure end-to-end functionality from document upload to report generation Performance Tests: Analyze large document handling and real-time AI scoring Code Quality: ESLint, Prettier, and Husky ensure consistent, maintainable code 3.6 Ethical and Legal Considerations All analyses performed on user-uploaded documents with explicit consent. 3.3.6 Internationalization Languages Supported: English and French Dynamic Switching: Users can switch languages without refreshing the page Persistence: Language preferences saved in user profile 3.3.7 Security and RLS Policies Row Level Security (RLS): Ensures users can only access their own analyses Encrypted Storage: All uploaded documents are stored securely in Supabase buckets Access Control: Temporary signed URLs prevent unauthorized access to stored files 3.4 Packaging and Deployment Frontend Build: Using Vite for optimized bundle creation Backend Deployment: Supabase handles serverless functions and database management Environment Variables: Secured for production, including Supabase API keys and AI model endpoints CI/CD Pipeline: Automated deployment with GitHub Actions and version control 3.5 Testing and Quality Assurance Unit Tests: Validate individual components and algorithm functions Integration Tests: Ensure end-to-end functionality from document upload to report generation Performance Tests: Analyze large document handling and real-time AI scoring Code Quality: ESLint, Prettier, and Husky ensure consistent, maintainable code 3.6 Ethical and Legal Considerations All analyses performed on user-uploaded documents with explicit consent. Compliance with GDPR and local privacy laws. Compliance with GDPR and local privacy laws. System designed for educational and research purposes only. System designed for educational and research purposes only. Users receive interpretable results, avoiding punitive or misrepresentative conclusions. Users receive interpretable results, avoiding punitive or misrepresentative conclusions. 3.7 Summary Chapter 3 provides a detailed account of AcadCheck's methodology, covering system architecture, tools, implementation, security, testing, and ethical considerations. 3.7 Summary Chapter 3 provides a detailed account of AcadCheck's methodology, covering system architecture, tools, implementation, security, testing, and ethical considerations. This methodology ensures accuracy, reliability, and usability, forming the foundation for Chapters 4 and 5, which present the actual implementation and results. This methodology ensures accuracy, reliability, and

usability, forming the foundation for Chapters 4 and 5, which present the actual implementation and results. CHAPTER 4: IMPLEMENTATION AND DISCUSSION 4.1 Development Environment Setup The development of AcadCheck was performed using a modern web stack on a local and cloud-based environment: Operating System: Windows 11 / Ubuntu 22.04 Frontend: Node.js 20, Vite, React 18.3.1, TypeScript Backend & Database: Supabase (PostgreSQL), serverless functions for analysis processing IDE: Visual Studio Code with ESLint, Prettier, and Git integration Version Control: GitHub for source code management and CI/CD pipelines All dependencies were installed using npm or yarn. CHAPTER 4: IMPLEMENTATION AND DISCUSSION 4.1 Development Environment Setup The development of AcadCheck was performed using a modern web stack on a local and cloud-based environment: Operating System: Windows 11 / Ubuntu 22.04 Frontend: Node.js 20, Vite, React 18.3.1, TypeScript Backend & Database: Supabase (PostgreSQL), serverless functions for analysis processing IDE: Visual Studio Code with ESLint, Prettier, and Git integration Version Control: GitHub for source code management and CI/CD pipelines All dependencies were installed using npm or yarn.

Environment variables were securely configured for API keys, database URLs, and AI model endpoints. Environment variables were securely configured for API keys, database URLs, and AI model endpoints. 4.2 Frontend Implementation The frontend was implemented using React and TypeScript, emphasizing responsive design and real-time feedback. 4.2 Frontend Implementation The frontend was implemented using React and TypeScript, emphasizing responsive design and real-time feedback.

4.2.1 Dashboard and Document Upload Drag-and-Drop Upload: Users can upload PDF, DOCX, or TXT files. 4.2.1 Dashboard and Document Upload Drag-and-Drop Upload: Users can upload PDF, DOCX, or TXT files. Live Preview: Using DocumentViewer, the uploaded file is displayed with high fidelity. Live Preview: Using DocumentViewer, the uploaded file is displayed with high fidelity. Upload Validation: The system checks file type, size, and integrity before analysis. Upload Validation: The system checks file type, size, and integrity before analysis. const handleFileUpload = async (file: File) => { if (! isValidFile(file)) return; const text = await fileToText(file); setUploadedText(text); }; 4.2.2 Interactive Analysis Report Sentence-level highlights indicate plagiarism and AI-generated content. const handleFileUpload = async (file: File) => { if (!isValidFile(file)) return; const text = await fileToText(file); setUploadedText(text); }; 4.2.2 Interactive Analysis Report Sentence-level highlights indicate plagiarism and AI-generated content.

Users can hover over sentences to view metrics such as confidence scores and AI probability. Users can hover over sentences to view metrics such as confidence scores

and AI probability. Results are dynamically updated without page reloads using React hooks. 4.3 Backend Implementation The backend is built with Supabase, providing secure storage, authentication, and serverless analysis functions. 4.3.1 Document Storage Uploaded documents are stored in encrypted buckets with temporary signed URLs. Metadata, including document name, user ID, and timestamp, is stored in the analyses table. 4.3.2 Analysis Workflow Text Extraction: Using PDF.js and Mammoth for PDFs and DOCX respectively.

Preprocessing: Tokenization, sentence segmentation, and cleaning.

Plagiarism Detection: N-grams and semantic similarity comparison with corpus documents. AI Detection: Transformer models assign a probability score per sentence.

const analysisResult = await analyzeText(extractedText, { corpus: referenceDocuments, useAI: true }); saveAnalysis(userId, analysisResult); 4.4 Document Analysis Workflow User uploads document → Frontend validates → Sends to backend.

Backend extracts text and segments into sentences. Each sentence is processed: Plagiarism score calculated AI detection score calculated Results are stored in analysis_sentences table. Frontend renders interactive report with color-coded highlights. Workflow Diagram: [User Upload] → [Text Extraction] → [Sentence Segmentation] → [AI & Plagiarism Analysis] → [Database Storage] → [Interactive Report] 4.5 AI and Plagiarism Detection Results Sentence-Level Scores: Each sentence is scored 0-100% for plagiarism and AI probability. Workflow

Diagram: [User Upload] → [Text Extraction] → [Sentence Segmentation] → [AI & Plagiarism Analysis] → [Database Storage] → [Interactive Report] 4.5 AI and Plagiarism Detection Results Sentence-Level Scores: Each sentence is scored 0-100% for plagiarism and AI probability. Global Metrics: Average plagiarism score, average AI score, lexical diversity, syntactic complexity. Visual Feedback: Red highlight → High plagiarism Yellow highlight → Medium plagiarism or AI probability Green → Safe or low risk Example Table: 4.6 Challenges and Troubleshooting During implementation, several challenges were encountered: Document Rendering Issues: Some DOCX formatting was lost; resolved using Mammoth and DOCX Preview.

Large File Handling: PDFs > 50MB caused slow extraction; solved by batch processing sentences. AI Model Latency: Transformer-based scoring was time-consuming; optimized using caching and async functions. Data Security: Careful attention was given to applying Row-Level Security (RLS) correctly across all tables. To ensure proper functionality, thorough testing was conducted, including unit tests, integration tests, and detailed logging. 4.7 Ethical and Legal Considerations All analyses respect user privacy. Temporary URLs prevent unauthorized document access. System designed for demonstration and educational purposes, avoiding legal infringement. Detailed logs and transparency ensure users can trust results. 4.8 Summary Chapter 4 presents the hands-on implementation of AcadCheck, covering the frontend and backend development, document processing workflows, AI and plagiarism detection features, and strategies for troubleshooting.

features, and strategies for troubleshooting. It also discusses the real-world challenges encountered when integrating complex algorithms into a modern web application, while ensuring security, scalability, and a smooth user experience. It also discusses the real-world challenges encountered when integrating complex algorithms into a modern web application, while ensuring security, scalability, and a smooth user experience.

CHAPTER 5: CONCLUSION 5.1 Key Findings The development and deployment of AcadCheck led to several key insights: Accurate Plagiarism Detection: By combining n-gram techniques, semantic comparisons, and contextual pattern recognition, AcadCheck can reliably detect both copied and paraphrased content. CHAPTER 5: CONCLUSION 5.1 Key Findings The development and deployment of AcadCheck led to several key insights: Accurate Plagiarism Detection: By combining n-gram techniques, semantic comparisons, and contextual pattern recognition, AcadCheck can reliably detect both copied and paraphrased content.

AI Content Recognition: The use of transformer-based models alongside heuristic methods allows for precise identification of AI-generated text, with detailed analysis at the sentence level. AI Content Recognition: The use of transformer-based models alongside heuristic methods allows for precise identification of AI-generated text, with detailed analysis at the sentence level. Intuitive User Interface: The interactive dashboard and real-time feedback make the results easy to understand, enhancing the overall user experience. Intuitive User Interface: The interactive dashboard and real-time feedback make the results easy to understand, enhancing the overall user experience.

Secure and Scalable System: Built with Supabase, React, and TypeScript, the platform ensures data privacy, user authentication, and the ability to handle large volumes of documents efficiently. Secure and Scalable System: Built with Supabase, React, and TypeScript, the platform ensures data privacy, user authentication, and the ability to handle large volumes of documents efficiently.

Support for Multiple File Formats: PDFs, DOCX, and TXT files are processed while maintaining formatting, ensuring accurate and consistent analysis. Support for Multiple File Formats: PDFs, DOCX, and TXT files are processed while maintaining formatting, ensuring accurate and consistent analysis. These observations highlight that tools for maintaining academic integrity must combine technical sophistication with user-centered design to be truly effective. These observations highlight that tools for maintaining academic integrity must combine technical sophistication with user-centered design to be truly effective. 5.2 Educational Value AcadCheck functions as a practical tool for students, researchers, and educators

alike: Learning Resource: Students can check their own work for accidental plagiarism or AI-generated content, helping them improve their writing habits. 5.2 Educational Value AcadCheck functions as a practical tool for students, researchers, and educators alike: Learning Resource: Students can check their own work for accidental plagiarism or AI-generated content, helping them improve their writing habits. Teaching Aid: Educators can use the platform to illustrate proper citation methods and maintain academic writing standards. Teaching Aid: Educators can use the platform to illustrate proper citation methods and maintain academic writing standards. Research Support: The application provides detailed metrics and visualizations that facilitate studies on academic integrity. Research Support: The application provides detailed metrics and visualizations that facilitate studies on academic integrity. Additionally, the development of AcadCheck offered valuable hands-on experience in full-stack web development, AI integration, and implementing secure data management practices. Additionally, the development of AcadCheck offered valuable hands-on experience in full-stack web development, AI integration, and implementing secure data management practices. 5.3 Limitations Despite its achievements, AcadCheck has some limitations: AI Detection Accuracy: Although generally reliable, the system may occasionally flag highly formal or technical writing as AI-generated. 5.3 Limitations Despite its achievements, AcadCheck has some limitations: AI Detection Accuracy: Although generally reliable, the system may occasionally flag highly formal or technical writing as AI-generated. Corpus Dependency: The accuracy of plagiarism detection depends heavily on the size and variety of the reference corpus. Corpus Dependency: The accuracy of plagiarism detection depends heavily on the size and variety of the reference corpus. Processing Time: Analyzing large documents or multiple files at once can take longer, even with performance optimizations in place. Processing Time: Analyzing large documents or multiple files at once can take longer, even with performance optimizations in place. Language Support: Currently, the application only handles English and French; supporting additional languages would require retraining the models. Language Support: Currently, the application only handles English and French; supporting additional languages would require retraining the models. Being aware of these limitations helps guide improvements and refinements for future versions of AcadCheck. Being aware of these limitations helps guide improvements and refinements for future versions of AcadCheck. 5.4 Recommendations for Future Work Future development of AcadCheck could focus on several key areas: Expanded Multilingual Support: Incorporating additional languages such as Spanish and German to make the tool accessible to a wider audience. 5.4 Recommendations for Future Work Future

development of AcadCheck could focus on several key areas: Expanded Multilingual Support: Incorporating additional languages such as Spanish and German to make the tool accessible to a wider audience. Integration with Learning Management Systems (LMS): Enabling direct connections to platforms like Moodle, Canvas, and Blackboard for seamless automated analysis. Enhanced Paraphrasing Detection: Implementing more advanced semantic analysis techniques to identify subtle paraphrasing in student submissions. Mobile Application: Creating a React Native version to allow users to access AcadCheck on smartphones and tablets. API for External Tools: Providing a public API so third-party applications can utilize AcadCheck's detection algorithms. Performance Improvements: Adopting microservices architecture and queue-based processing to efficiently handle large volumes of documents.

These improvements would further enhance AcadCheck's value as a reliable and professional tool for academic integrity and research support.

5.5 Final Thoughts AcadCheck showcases how advanced AI, plagiarism detection, and secure web technologies can be combined into a practical, user-friendly platform. It demonstrates that technology can support academic integrity in a way that is both ethical and transparent. The development of this project offered valuable hands-on learning experiences, covering everything from system architecture and algorithm implementation to frontend and backend integration, as well as data security management.

management. The resulting application not only meets its intended goals but also provides a solid foundation for further research and innovation in academic integrity tools. In conclusion, AcadCheck stands as a modern, dependable, and scalable solution for detecting plagiarism and AI-generated content, striking a balance between technical sophistication, usability, security, and ethical responsibility.

REFERENCES Books / Monographs Bishop, C. M. M. (2006). (2006). Pattern recognition and machine learning. Pattern recognition and machine learning. Springer. Springer. Bird, S., Klein, E., & Loper, E. Bird, S., Klein, E., & Loper, E. (2009). (2009). Natural language processing with Python. Natural language processing with Python. O'Reilly Media. O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. Goodfellow, I., Bengio, Y., & Courville, A. (2016). (2016). Deep learning. Deep learning. MIT Press. MIT Press. Jurafsky, D., & Martin, J. Jurafsky, D., & Martin, J. H. H. (2021). (2021). Speech and language processing (3rd ed.). Speech and language processing (3rd ed.). Pearson. Pearson. Journal Articles / Conference Papers Alzahrani, S. Journal Articles / Conference Papers Alzahrani, S. M., Salim, N., & Abraham, A. M., Salim, N., & Abraham, A. (2012). (2012).

Understanding plagiarism: Linguistic patterns, textual features, and detection techniques. Understanding plagiarism: Linguistic patterns, textual features, and detection techniques. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 42(5), 1339–1352. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 42(5), 1339–1352. https://doi.org/10.1109/TSMCC.2012.2184410 Potthast, M., Stein, B., Barrón-Cedeño, A., & Rosso, P. https://doi.org/10.1109/TSMCC.2012.2184410 Potthast, M., Stein, B., Barrón-Cedeño, A., & Rosso, P. (2010). (2010). An evaluation framework for plagiarism detection. An evaluation framework for plagiarism detection. Proceedings of the 23rd International Conference on Computational Linguistics (COLING), 997–1005. Proceedings of the 23rd International Conference on Computational Linguistics (COLING), 997–1005. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. N., Kaiser, Ł., & Polosukhin, I. (2017). (2017).

Attention is all you need. Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008. Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008. Gao, C., Wang, X., & He, D. Gao, C., Wang, X., & He, D. (2023). (2023). Detecting AI-generated text using linguistic features and transformer models. Detecting AI-generated text using linguistic features and transformer models. Journal of Artificial Intelligence Research, 76, 1201–1220. Journal of Artificial Intelligence Research, 76, 1201–1220. https://doi.org/10.1613/jair.1.13345 Web / Technical Documentation Hugging Face. https://doi.org/10.1613/jair.1.13345 Web / Technical Documentation Hugging Face. (2025). (2025). Transformers documentation. Transformers documentation.

Retrieved August 2025, from https://huggingface.co/docs/transformers PDF.js. Retrieved August 2025, from https://huggingface.co/docs/transformers PDF.js. (2025). (2025). Mozilla PDF.js project. Mozilla PDF.js project. Retrieved August 2025, from https://mozilla.github.io/pdf.js/ Supabase. Retrieved August 2025, from https://mozilla.github.io/pdf.js/ Supabase. (2025). (2025). Supabase documentation. Supabase documentation. Retrieved August 2025, from https://supabase.com/docs

Vite. Retrieved August 2025, from https://supabase.com/docs Vite. (2025). (2025). Vite documentation. Vite documentation. Retrieved August 2025, from https://vitejs.dev/ Martin, J., & Ha, D. Retrieved August 2025, from https://vitejs.dev/ Martin, J., & Ha, D. (2022). (2022). Detecting AI-generated content: Challenges and strategies. Detecting AI-generated content: Challenges and strategies.

Medium. Medium. Retrieved from https://medium.com Standards / Guidelines European Union. Retrieved from https://medium.com Standards / Guidelines European Union. (2018). (2018). General Data Protection Regulation (GDPR). General Data Protection Regulation (GDPR). Official Journal of the European Union. Official Journal of the European Union. Association for Computing Machinery (ACM). Association for Computing Machinery (ACM). (2021). (2021). ACM code of ethics and professional conduct. ACM code of ethics and professional conduct.

Retrieved from https://www.acm.org/code-of-ethics APENDIX This appendix presents visual documentation of the main features of the AcadCheck application developed during the internship. Retrieved from https://www.acm.org/code-of-ethics APENDIX This appendix presents visual documentation of the main features of the AcadCheck application developed during the internship. The figures illustrate the user interface, examples of plagiarism and AI detection results, and the report generation

functionality. The figures illustrate the user interface, examples of plagiarism and AI detection results, and the report generation functionality. All displayed content was created within a controlled academic setting and is intended solely for educational purposes. All displayed content was created within a controlled academic setting and is intended solely for educational purposes. Figure A.1 – Login and Authentication Interface The screenshot displays AcadCheck's secure login page. Figure A.1 – Login and Authentication Interface The screenshot displays AcadCheck's secure login page. Users can sign in using their email and password to access the dashboard. Users can sign in using their email and password to access the dashboard. The page also provides options for account registration, password recovery, and selecting the interface language (English or French). The page also provides options for account registration, password recovery, and selecting the interface language (English or French). Figure A.2 – Main Dashboard Overview This figure shows the main dashboard of AcadCheck, where users can access the core features, upload documents, and view summaries of plagiarism and AI detection results. Figure A.2 – Main Dashboard Overview This figure shows the main dashboard of AcadCheck, where users can access the core features, upload documents, and view summaries of plagiarism and AI detection results. Figure A.3 – Document Preview (PDF/DOCX/TXT) This screenshot displays the DocumentViewer feature, which lets users preview uploaded documents in real time. Figure A.3 – Document Preview (PDF/DOCX/TXT) This screenshot displays the DocumentViewer feature, which lets users preview uploaded documents in real time. The component maintains the original formatting and offers zoom and navigation options for a thorough review. The component maintains the original formatting and offers zoom and navigation options for a thorough review. Figure A.4 – Plagiarism and AI Detection Results The figure shows sentence-by-sentence analysis, with plagiarism and AI-generated content scores. Figure A.4 – Plagiarism and AI Detection Results The figure shows sentence-by-sentence analysis, with plagiarism and AI-generated content scores. Color-coded highlights indicate suspicious sections, while confidence metrics provide additional insight for each sentence. Color-coded highlights indicate suspicious sections, while confidence metrics provide additional insight for each sentence. Figure A.5 – Detailed Report Generation This screenshot presents the PDF report generated by AcadCheck. Figure A.5 – Detailed Report Generation This screenshot presents the PDF report generated by AcadCheck. It includes global metrics, sentence-level scores, and highlighted passages. It includes global metrics, sentence-level scores, and highlighted passages. Users can download or share the report directly from the dashboard. Users

can download or share the report directly from the dashboard. Figure A.6 – Analysis History and Filtering This figure shows the historical analysis interface, listing previous uploads with dates, document names, and scores. Figure A.6 – Analysis History and Filtering This figure shows the historical analysis interface, listing previous uploads with dates, document names, and scores. Users can filter, search, and re-open previous reports for reference or export. Users can filter, search, and re-open previous reports for reference or export.

Detailed Issue Analysis

Plagiarism Detected   15.8% Match

"GIRNE AMERICAN UNIVERSITY Faculty of Engineering Department of Computer Engineering AcadCheck: AI and Plagiarism Detection Tool SUMMER TRAINING REPORT Student: Elie katende kazwela Student ID: 221701001 Supervisor: Mr."

Plagiarism Detected   15.8% Match

"Cliff Acknowledgements I would like to sincerely thank all those who supported me during my internship at the university and throughout the preparation of this report."

Plagiarism Detected   15.8% Match

"First and foremost , I am especially grateful to my supervisor , Mr."

Plagiarism Detected   15.8% Match

"Peter Cliff, for his guidance ,encouragement and constant support."

Plagiarism Detected   15.8% Match

"His expertise and thoughtful feedback played a key role in the progress and success of my work."

Plagiarism Detected `15.8% Match`

"I also wish to express my appreciation to the staff and faculty of the university for welcoming me , Providing the necessary resources, and creating a positive environment that made this internship a valuable learning experience."

Plagiarism Detected `15.8% Match`

"the knowledge and skills I gained have been highly beneficial to my academic and professional growth."

Plagiarism Detected `15.8% Match`

"Lastly ,I would like to thank my family and friends for their patience , encouragement , and continuous Support , which greatly helped me in completing this project."

Plagiarism Detected `15.8% Match`

"Abstract This report introduces AcadCheck, a web application developed to detect both plagiarism and AI generated content in academic documents ."

Plagiarism Detected `15.8% Match`

"the system was implemented using React, TypeScript, and Supabase, and it integrates advanced document analysis methods with a user-friendly interface to deliver precise and comprehensive results."

AcadCheck - Academic Integrity Report

Plagiarism Detected   15.8% Match

"The main objectives of the project are structured around four core component : Plagiarism detection : comparing documents against a large corpus by applying N-grams, semantic similarity, and pattern-matching techniques."

Plagiarism Detected   15.8% Match

"AI content detection – identifying AI-generated text through transformer-based models and linguistic analysis."

Plagiarism Detected   15.8% Match

"Document analysis and reporting – producing detailed, sentence-level feedback with metrics such as lexical diversity, syntactic complexity, semantic coherence, and perplexity."

Plagiarism Detected   15.8% Match

"Secure and scalable architecture – leveraging Supabase for database management, data protection, and Row Level Security (RLS)."

Plagiarism Detected   15.8% Match

"This report presents the overall architecture , development process ,challenges ,faced and evaluation results of the application ,while also reflecting on the ethical considerations involved .it highlights the use of modern approaches to document analysis , the integration of artificial intelligence in plagiarism detection ,and the design of scalable web applications."

Plagiarism Detected `15.8% Match`

"Chapter 1: Introduction Background In academic settings, maintaining the integrity of written work is essential."

Plagiarism Detected `15.8% Match`

"The growing use of digital tools, online resources, and AI-assisted content creation has heightened the risk of plagiarism and the improper use of external materials."

Plagiarism Detected `15.8% Match`

"AcadCheck tackles these issues by offering an automated solution that can identify both plagiarized content and text generated by artificial intelligence, helping to uphold academic honesty and the quality of scholarly work."

Plagiarism Detected `15.8% Match`

"Contemporary document analysis relies on advanced technologies such as machine learning, semantic evaluation, and natural language processing (NLP)."

Plagiarism Detected `15.8% Match`

"By combining these methods with a straightforward and intuitive web interface, AcadCheck enables educators, students, and researchers to examine documents effectively while ensuring the security and privacy of their data."

Plagiarism Detected | 15.8% Match

"1.2 Objectives The primary goals of this internship project are: To create a web application capable of identifying both plagiarism and AI-generated content."

Plagiarism Detected | 15.8% Match

"To integrate sophisticated analysis techniques, combining machine learning with heuristic methods."

Plagiarism Detected | 15.8% Match

"To provide comprehensive reporting, including analysis at the sentence level, confidence indicators, and clear visual cues."

Plagiarism Detected | 15.8% Match

"To guarantee secure management of user data, leveraging authentication, Row-Level Security (RLS), and protected storage."

Plagiarism Detected | 15.8% Match

"To deliver a user-friendly and responsive interface that accommodates multilingual users effectively."

Plagiarism Detected | 15.8% Match

"1.3 Scope This project centers on the design, development, and implementation of AcadCheck as a tool for research and education."

Plagiarism Detected   15.8% Match

"The application is compatible with PDF, DOCX, and TXT file formats, and it offers detailed sentence-level analysis for both plagiarism and AI-generated content."

Plagiarism Detected   15.8% Match

"All testing and usage of the application are carried out ethically, within controlled environments, and in strict accordance with data protection regulations."

Plagiarism Detected   15.8% Match

"The project does not involve deployment in untrusted environments or the collection of unauthorized data."

Plagiarism Detected   15.8% Match

"1.4 Importance of the Study AcadCheck contributes to multiple domains: Academic integrity: Helps institutions maintain originality standards."

Plagiarism Detected   15.8% Match

"AI literacy: Educates users on distinguishing AI-generated content."

Plagiarism Detected `15.8% Match`

"Technical education: Demonstrates the use of modern web technologies, machine learning, and database security in practical applications."

Plagiarism Detected `15.8% Match`

"By integrating AI detection with plagiarism analysis, the project addresses modern challenges in digital content evaluation, providing insights for students, educators, and software developers."

Plagiarism Detected `15.8% Match`

"1.5 Structure of the Report The organization of the report is: Chapter 1: Introduction – Presents background, objectives, scope, and significance."

Plagiarism Detected `15.8% Match`

"Chapter 2: Literature Review – Reviews plagiarism detection methods, AI content detection, and ethical/legal considerations."

Plagiarism Detected `15.8% Match`

"Chapter 3: Methodology – Details the system architecture, tools, libraries, and implementation steps."

AcadCheck - Academic Integrity Report

Plagiarism Detected **15.8% Match**

"Chapter 4: Implementation and Discussion – Covers code implementation, document processing, analysis workflow, and troubleshooting."

Plagiarism Detected **15.8% Match**

"Chapter 5: Conclusion – Summarizes findings, limitations, educational value, and future improvements."

Plagiarism Detected **15.8% Match**

"References – List of sources cited throughout the report."

Plagiarism Detected **15.8% Match**

"CHAPTER 2: LITERATURE REVIEW 2.1 Overview of Plagiarism Detection Detecting plagiarism is essential for upholding academic integrity."

Plagiarism Detected **15.8% Match**

"Traditional tools mainly rely on exact text matching, which can spot direct copying but often misses paraphrased or slightly altered content."

Plagiarism Detected **15.8% Match**

"Modern systems, like AcadCheck, use a combination of syntactic, semantic, and contextual analysis, enabling more accurate detection of potential plagiarism."

Plagiarism Detected **15.8% Match**

"N-gram Analysis: Documents are broken into sequences of words (n-grams) and compared to a reference corpus."

Plagiarism Detected **15.8% Match**

"Similar sequences indicate potential plagiarism."

Plagiarism Detected **15.8% Match**

"AcadCheck uses adaptive n-gram sizes to capture both short phrases and longer patterns."

Plagiarism Detected **15.8% Match**

"Semantic Similarity: Using NLP techniques, AcadCheck compares the meaning of sentences rather than exact wording."

Plagiarism Detected **15.8% Match**

"This allows the system to detect paraphrased content."

Plagiarism Detected **15.8% Match**

"Contextual Patterns: Certain academic formulations are commonly reused."

Plagiarism Detected   15.8% Match

> *"By identifying these patterns, the system can differentiate between general academic phrases and copied unique content."*

Plagiarism Detected   15.8% Match

> *"Recent research emphasizes the integration of machine learning models for more nuanced plagiarism detection, allowing systems to learn from labeled datasets and improve detection accuracy over time."*

Plagiarism Detected   15.8% Match

> *"2.2 AI-Based Text Detection With the emergence of AI-generated content, particularly large language models like GPT, distinguishing between human-written and AI-generated text has become a crucial challenge."*

Plagiarism Detected   15.8% Match

> *"AcadCheck implements hybrid AI detection techniques, combining transformer-based models with linguistic heuristics: Transformer Models: Pre-trained models analyze text for patterns indicative of AI generation, such as overuse of certain connectors, predictable phrasing, or low variability."*

Plagiarism Detected   15.8% Match

> *"Linguistic Heuristics: Metrics such as lexical diversity, syntactic complexity, and perplexity are used to evaluate the naturalness of text."*

Plagiarism Detected

"AI-generated text often exhibits high perplexity uniformity and low lexical diversity."

Plagiarism Detected

"Sentence-Level Scoring: Each sentence is assigned an AI score, indicating the likelihood of machine generation."

Plagiarism Detected

"This granularity allows users to review suspicious content precisely."

Plagiarism Detected

"By integrating AI detection with plagiarism analysis, AcadCheck provides a comprehensive evaluation of document originality."

Plagiarism Detected

"2.3 Document Analysis Techniques Effective plagiarism and AI detection require robust document analysis."

Plagiarism Detected

"AcadCheck supports multiple formats, including PDF, DOCX, and TXT, with consistent text extraction: PDF Analysis: Using PDF.js, the system extracts text while preserving document structure."

Plagiarism Detected `15.8% Match`

"Word Document Handling: Mammoth and DOCX Preview convert DOCX files to HTML, ensuring accurate rendering and sentence-level analysis."

Plagiarism Detected `15.8% Match`

"Text Preprocessing: Cleaning, tokenization, and normalization prepare the content for further analysis."

Plagiarism Detected `15.8% Match`

"Sentence segmentation, feature extraction, and vectorization are applied before any similarity or AI detection algorithms, ensuring that results are both accurate and interpretable."

Plagiarism Detected `15.8% Match`

"2.4 Ethical and Legal Considerations Given the sensitivity of plagiarism and AI detection, ethics and legality are central to AcadCheck's design: User Consent: Only documents uploaded with explicit consent are analyzed."

Plagiarism Detected `15.8% Match`

"Data Security: Supabase provides encrypted storage and Row Level Security (RLS) to isolate user data."

Plagiarism Detected
15.8% Match

"Legal Compliance: The system aligns with GDPR and other data protection regulations, preventing unauthorized access or retention."

Plagiarism Detected
15.8% Match

"Responsible Reporting: Results are presented solely for educational or institutional purposes, avoiding misuse of the analysis."

Plagiarism Detected
15.8% Match

"Ethical considerations also guide the inclusion of AI detection."

Plagiarism Detected
15.8% Match

"The system is designed for transparency, providing users with interpretable metrics and avoiding punitive assumptions."

Plagiarism Detected
15.8% Match

"2.5 Use of React, TypeScript, and Supabase in Web Apps Modern web technologies play a key role in the functionality and usability of AcadCheck: React (v18.3.1): Provides a responsive, dynamic frontend capable of handling real-time document previews and interactive dashboards."

Plagiarism Detected

"TypeScript: Adds type safety and improves maintainability, especially important for complex components such as analysis tables and reports."

Plagiarism Detected

15.8% Match

"Supabase: Serves as a backend and database solution, offering authentication, secure storage, and serverless functions."

Plagiarism Detected

15.8% Match

"RLS ensures that users only access their own analyses."

Plagiarism Detected

15.8% Match

"Frontend UI Components: Using Tailwind CSS and shadcn/ui, the interface is modern, clean, and intuitive."

Plagiarism Detected

15.8% Match

"Components like DocumentViewer and HighlightedText provide detailed visual feedback to users."

Plagiarism Detected

15.8% Match

"The combination of these technologies allows for a robust, scalable, and secure web application, supporting complex AI and plagiarism detection workflows while maintaining excellent user experience."

Plagiarism Detected

"2.6 Integration of Detection Algorithms in Web Applications A critical aspect of modern plagiarism detection systems is the seamless integration of backend analysis algorithms with frontend visualization: Sentence-Level Analysis: Each sentence is analyzed for plagiarism and AI content."

Plagiarism Detected

15.8% Match

"Results are transmitted from the backend to the frontend, where they are visually highlighted in the document."

Plagiarism Detected

15.8% Match

"Scoring and Metrics: Global scores, confidence levels, and detailed metrics are presented using interactive dashboards."

Plagiarism Detected

15.8% Match

"Real-Time Feedback: Users can upload documents and receive analysis results without page reloads, thanks to React's component architecture."

Plagiarism Detected

15.8% Match

"This integration ensures that technical sophistication in analysis does not compromise usability."

Plagiarism Detected   `15.8% Match`

> *"2.7 Summary Chapter 2 shows that modern plagiarism and AI detection relies on multi-layered methods that combine semantic analysis, machine learning, and secure web technologies."*

Plagiarism Detected   `15.8% Match`

> *"AcadCheck applies these principles by providing a system that is easy to use, secure, and accurate, effectively tackling today's challenges in maintaining academic integrity."*

Plagiarism Detected   `15.8% Match`

> *"CHAPTER 3: METHODOLOGY 3.1 Design Objectives and System Architecture The main goal in designing AcadCheck is to build a web application that is secure, scalable, and easy to use for detecting plagiarism and AI-generated content."*

Plagiarism Detected   `15.8% Match`

> *"The key aspects taken into account include: Accuracy: Implement advanced algorithms for precise detection of plagiarized and AI-generated text."*

Plagiarism Detected   `15.8% Match`

> *"Security: Ensure strict user data isolation through authentication and Row Level Security (RLS)."*

Plagiarism Detected `15.8% Match`

"Usability: Provide an intuitive dashboard and interactive reports that can be easily interpreted by students, educators, and researchers."

Plagiarism Detected `15.8% Match`

"Scalability: Design the system to handle large volumes of document uploads and concurrent analyses."

Plagiarism Detected `15.8% Match`

"The system architecture is composed of the following layers: Frontend (React + TypeScript): Handles user interaction, document preview, and visualization of analysis results."

Plagiarism Detected `15.8% Match`

"Backend (Supabase + Serverless Functions): Processes document uploads, manages user authentication, and executes analysis algorithms."

Plagiarism Detected `15.8% Match`

"Database (PostgreSQL): Stores user data, documents, analysis results, and sentence-level metrics."

Plagiarism Detected   15.8% Match

> *"AI Engine: Implements transformer-based models and linguistic heuristics for AI detection."*

Plagiarism Detected   15.8% Match

> *"Security Layer: Enforces RLS, secure storage, and access control to protect sensitive data."*

Plagiarism Detected   15.8% Match

> *"The architecture is modular, allowing future integration of additional features such as multi-language support, paraphrasing detection, and integration with LMS platforms."*

AcadCheck - Academic Integrity Report

Plagiarism Detected

15.8% Match

"3.2 Tools and Libraries The following tools and libraries were selected for their suitability and robustness: Frontend: React 18.3.1 for component-based UI TypeScript for type safety and maintainability Tailwind CSS and shadcn/ui for consistent, responsive styling React Router DOM for client-side routing Backend & Database: Supabase (PostgreSQL) for secure user management, storage, and database operations Serverless functions for document processing Document Processing: PDF.js for PDF extraction Mammoth and DOCX Preview for Word document conversion and rendering Custom fileToText library for preprocessing and normalization AI & Plagiarism Detection: Hugging Face Transformers for AI detection N-gram analysis, semantic similarity, and heuristic algorithms for plagiarism detection Testing & Quality: Jest and React Testing Library for unit tests Cypress for integration tests ESLint, Prettier, and Husky for code quality and consistency 3.3 Development and Implementation The development of AcadCheck was divided into several modules to ensure modularity and maintainability."

Plagiarism Detected

15.8% Match

"3.3.1 User Authentication and Session Management Sign-up/Login: Users register with email and password."

Plagiarism Detected

15.8% Match

"Supabase handles authentication securely."

Plagiarism Detected

15.8% Match

"Session Persistence: JWT tokens are stored in browser local storage, with automatic session renewal."

Plagiarism Detected `15.8% Match`

"Route Protection: Only authenticated users can access dashboards and analysis features."

Plagiarism Detected `15.8% Match`

"3.3.2 Document Upload and Preprocessing Supported Formats: PDF, DOCX, and TXT File Validation: Ensures correct format and maximum file size compliance Preprocessing Steps: Text extraction from uploaded files Sentence segmentation and tokenization Removal of unnecessary characters or formatting issues 3.3.3 Plagiarism Detection Algorithms Exact Matching: Detects copied content using n-grams (3-5 words)."

Plagiarism Detected `15.8% Match`

"Semantic Similarity: Compares sentences against reference corpus using vector embeddings."

Plagiarism Detected `15.8% Match`

"Contextual Analysis: Identifies repeated academic phrases to minimize false positives."

Plagiarism Detected `15.8% Match`

"3.3.4 AI Content Detection Transformer Models: Analyzes text for patterns indicative of AI generation."

Plagiarism Detected `15.8% Match`

"Stylistic Metrics: Computes lexical diversity, syntactic complexity, semantic coherence, and perplexity."

Plagiarism Detected `15.8% Match`

"Sentence-Level Scoring: Each sentence receives a probability score for AI authorship."

Plagiarism Detected `15.8% Match`

"3.3.5 Analysis Report Generation Visual Dashboard: Displays sentence-level highlights and overall metrics."

Plagiarism Detected `15.8% Match`

"Export Options: Users can download PDF reports with annotated content."

Plagiarism Detected `15.8% Match`

"Confidence Scores: Metrics such as AI score, plagiarism score, and confidence levels guide interpretation."

Plagiarism Detected

15.8% Match

"3.3.6 Internationalization Languages Supported: English and French Dynamic Switching: Users can switch languages without refreshing the page Persistence: Language preferences saved in user profile 3.3.7 Security and RLS Policies Row Level Security (RLS): Ensures users can only access their own analyses Encrypted Storage: All uploaded documents are stored securely in Supabase buckets Access Control: Temporary signed URLs prevent unauthorized access to stored files 3.4 Packaging and Deployment Frontend Build: Using Vite for optimized bundle creation Backend Deployment: Supabase handles serverless functions and database management Environment Variables: Secured for production, including Supabase API keys and AI model endpoints CI/CD Pipeline: Automated deployment with GitHub Actions and version control 3.5 Testing and Quality Assurance Unit Tests: Validate individual components and algorithm functions Integration Tests: Ensure end-to-end functionality from document upload to report generation Performance Tests: Analyze large document handling and real-time AI scoring Code Quality: ESLint, Prettier, and Husky ensure consistent, maintainable code 3.6 Ethical and Legal Considerations All analyses performed on user-uploaded documents with explicit consent."

Plagiarism Detected

15.8% Match

"Compliance with GDPR and local privacy laws."

Plagiarism Detected

15.8% Match

"System designed for educational and research purposes only."

Plagiarism Detected

"Users receive interpretable results, avoiding punitive or misrepresentative conclusions."

Plagiarism Detected

"3.7 Summary Chapter 3 provides a detailed account of AcadCheck's methodology, covering system architecture, tools, implementation, security, testing, and ethical considerations."

Plagiarism Detected

"This methodology ensures accuracy, reliability, and usability, forming the foundation for Chapters 4 and 5, which present the actual implementation and results."

Plagiarism Detected

"CHAPTER 4: IMPLEMENTATION AND DISCUSSION 4.1 Development Environment Setup The development of AcadCheck was performed using a modern web stack on a local and cloud-based environment: Operating System: Windows 11 / Ubuntu 22.04 Frontend: Node.js 20, Vite, React 18.3.1, TypeScript Backend & Database: Supabase (PostgreSQL), serverless functions for analysis processing IDE: Visual Studio Code with ESLint, Prettier, and Git integration Version Control: GitHub for source code management and CI/CD pipelines All dependencies were installed using npm or yarn."

Plagiarism Detected

15.8% Match

"Environment variables were securely configured for API keys, database URLs, and AI model endpoints."

Plagiarism Detected

15.8% Match

"4.2 Frontend Implementation The frontend was implemented using React and TypeScript, emphasizing responsive design and real-time feedback."

Plagiarism Detected

15.8% Match

"4.2.1 Dashboard and Document Upload Drag-and-Drop Upload: Users can upload PDF, DOCX, or TXT files."

Plagiarism Detected

15.8% Match

"Live Preview: Using DocumentViewer, the uploaded file is displayed with high fidelity."

Plagiarism Detected

15.8% Match

"Upload Validation: The system checks file type, size, and integrity before analysis."

Plagiarism Detected `15.8% Match`

"const handleFileUpload = async (file: File) => { if (!isValidFile(file)) return; const text = await fileToText(file); setUploadedText(text); }; 4.2.2 Interactive Analysis Report Sentence-level highlights indicate plagiarism and AI-generated content."

Plagiarism Detected `15.8% Match`

"Users can hover over sentences to view metrics such as confidence scores and AI probability."

Plagiarism Detected `15.8% Match`

"Results are dynamically updated without page reloads using React hooks."

Plagiarism Detected `15.8% Match`

"4.3 Backend Implementation The backend is built with Supabase, providing secure storage, authentication, and serverless analysis functions."

Plagiarism Detected `15.8% Match`

"4.3.1 Document Storage Uploaded documents are stored in encrypted buckets with temporary signed URLs."

AcadCheck - Academic Integrity Report

Plagiarism Detected 15.8% Match

"Metadata, including document name, user ID, and timestamp, is stored in the analyses table."

Plagiarism Detected 15.8% Match

"4.3.2 Analysis Workflow Text Extraction: Using PDF.js and Mammoth for PDFs and DOCX respectively."

Plagiarism Detected 15.8% Match

"Preprocessing: Tokenization, sentence segmentation, and cleaning."

Plagiarism Detected 15.8% Match

"Plagiarism Detection: N-grams and semantic similarity comparison with corpus documents."

Plagiarism Detected 15.8% Match

"AI Detection: Transformer models assign a probability score per sentence."

Plagiarism Detected 15.8% Match

"const analysisResult = await analyzeText(extractedText, { corpus: referenceDocuments, useAI: true }); saveAnalysis(userId, analysisResult); 4.4 Document Analysis Workflow User uploads document → Frontend validates → Sends to backend."

Plagiarism Detected   `15.8% Match`

"Backend extracts text and segments into sentences."

Plagiarism Detected   `15.8% Match`

"Each sentence is processed: Plagiarism score calculated AI detection score calculated Results are stored in analysis_sentences table."

Plagiarism Detected   `15.8% Match`

"Frontend renders interactive report with color-coded highlights."

Plagiarism Detected   `15.8% Match`

"Workflow Diagram: [User Upload] → [Text Extraction] → [Sentence Segmentation] → [AI & Plagiarism Analysis] → [Database Storage] → [Interactive Report] 4.5 AI and Plagiarism Detection Results Sentence-Level Scores: Each sentence is scored 0-100% for plagiarism and AI probability."

Plagiarism Detected   `15.8% Match`

"Global Metrics: Average plagiarism score, average AI score, lexical diversity, syntactic complexity."

Plagiarism Detected　　　15.8% Match

"Visual Feedback: Red highlight → High plagiarism Yellow highlight → Medium plagiarism or AI probability Green → Safe or low risk Example Table: 4.6 Challenges and Troubleshooting During implementation, several challenges were encountered: Document Rendering Issues: Some DOCX formatting was lost; resolved using Mammoth and DOCX Preview."

Plagiarism Detected　　　15.8% Match

"Large File Handling: PDFs > 50MB caused slow extraction; solved by batch processing sentences."

Plagiarism Detected　　　15.8% Match

"AI Model Latency: Transformer-based scoring was time-consuming; optimized using caching and async functions."

Plagiarism Detected　　　15.8% Match

"Data Security: Careful attention was given to applying Row-Level Security (RLS) correctly across all tables."

Plagiarism Detected　　　15.8% Match

"To ensure proper functionality, thorough testing was conducted, including unit tests, integration tests, and detailed logging."

AcadCheck - Academic Integrity Report

Plagiarism Detected   15.8% Match

"4.7 Ethical and Legal Considerations All analyses respect user privacy."

Plagiarism Detected   15.8% Match

"Temporary URLs prevent unauthorized document access."

Plagiarism Detected   15.8% Match

"System designed for demonstration and educational purposes, avoiding legal infringement."

Plagiarism Detected   15.8% Match

"Detailed logs and transparency ensure users can trust results."

Plagiarism Detected   15.8% Match

"4.8 Summary Chapter 4 presents the hands-on implementation of AcadCheck, covering the frontend and backend development, document processing workflows, AI and plagiarism detection features, and strategies for troubleshooting."

Plagiarism Detected   15.8% Match

"It also discusses the real-world challenges encountered when integrating complex algorithms into a modern web application, while ensuring security, scalability, and a smooth user experience."

AcadCheck - Academic Integrity Report

Plagiarism Detected

15.8% Match

"CHAPTER 5: CONCLUSION 5.1 Key Findings The development and deployment of AcadCheck led to several key insights: Accurate Plagiarism Detection: By combining n-gram techniques, semantic comparisons, and contextual pattern recognition, AcadCheck can reliably detect both copied and paraphrased content."

Plagiarism Detected

15.8% Match

"AI Content Recognition: The use of transformer-based models alongside heuristic methods allows for precise identification of AI-generated text, with detailed analysis at the sentence level."

Plagiarism Detected

15.8% Match

"Intuitive User Interface: The interactive dashboard and real-time feedback make the results easy to understand, enhancing the overall user experience."

Plagiarism Detected

15.8% Match

"Secure and Scalable System: Built with Supabase, React, and TypeScript, the platform ensures data privacy, user authentication, and the ability to handle large volumes of documents efficiently."

Plagiarism Detected

15.8% Match

"Support for Multiple File Formats: PDFs, DOCX, and TXT files are processed while maintaining formatting, ensuring accurate and consistent analysis."

Plagiarism Detected

<div style="text-align:right">**15.8% Match**</div>

"These observations highlight that tools for maintaining academic integrity must combine technical sophistication with user-centered design to be truly effective."

Plagiarism Detected

<div style="text-align:right">**15.8% Match**</div>

"5.2 Educational Value AcadCheck functions as a practical tool for students, researchers, and educators alike: Learning Resource: Students can check their own work for accidental plagiarism or AI-generated content, helping them improve their writing habits."

Plagiarism Detected

<div style="text-align:right">**15.8% Match**</div>

"Teaching Aid: Educators can use the platform to illustrate proper citation methods and maintain academic writing standards."

Plagiarism Detected

<div style="text-align:right">**15.8% Match**</div>

"Research Support: The application provides detailed metrics and visualizations that facilitate studies on academic integrity."

Plagiarism Detected

<div style="text-align:right">**15.8% Match**</div>

"Additionally, the development of AcadCheck offered valuable hands-on experience in full-stack web development, AI integration, and implementing secure data management practices."

Plagiarism Detected　　　　　15.8% Match

"5.3 Limitations Despite its achievements, AcadCheck has some limitations: AI Detection Accuracy: Although generally reliable, the system may occasionally flag highly formal or technical writing as AI-generated."

Plagiarism Detected　　　　　15.8% Match

"Corpus Dependency: The accuracy of plagiarism detection depends heavily on the size and variety of the reference corpus."

Plagiarism Detected　　　　　15.8% Match

"Processing Time: Analyzing large documents or multiple files at once can take longer, even with performance optimizations in place."

Plagiarism Detected　　　　　15.8% Match

"Language Support: Currently, the application only handles English and French; supporting additional languages would require retraining the models."

Plagiarism Detected　　　　　15.8% Match

"Being aware of these limitations helps guide improvements and refinements for future versions of AcadCheck."

Plagiarism Detected   15.8% Match

"5.4 Recommendations for Future Work Future development of AcadCheck could focus on several key areas: Expanded Multilingual Support: Incorporating additional languages such as Spanish and German to make the tool accessible to a wider audience."

Plagiarism Detected   15.8% Match

"Integration with Learning Management Systems (LMS): Enabling direct connections to platforms like Moodle, Canvas, and Blackboard for seamless automated analysis."

Plagiarism Detected   15.8% Match

"Enhanced Paraphrasing Detection: Implementing more advanced semantic analysis techniques to identify subtle paraphrasing in student submissions."

Plagiarism Detected   15.8% Match

"Mobile Application: Creating a React Native version to allow users to access AcadCheck on smartphones and tablets."

Plagiarism Detected   15.8% Match

"API for External Tools: Providing a public API so third-party applications can utilize AcadCheck's detection algorithms."

Plagiarism Detected   15.8% Match

"Performance Improvements: Adopting microservices architecture and queue-based processing to efficiently handle large volumes of documents."

Plagiarism Detected   15.8% Match

"These improvements would further enhance AcadCheck's value as a reliable and professional tool for academic integrity and research support."

Plagiarism Detected   15.8% Match

"5.5 Final Thoughts AcadCheck showcases how advanced AI, plagiarism detection, and secure web technologies can be combined into a practical, user-friendly platform."

Plagiarism Detected   15.8% Match

"It demonstrates that technology can support academic integrity in a way that is both ethical and transparent."

Plagiarism Detected   15.8% Match

"The development of this project offered valuable hands-on learning experiences, covering everything from system architecture and algorithm implementation to frontend and backend integration, as well as data security management."

Plagiarism Detected

15.8% Match

"The resulting application not only meets its intended goals but also provides a solid foundation for further research and innovation in academic integrity tools."

Plagiarism Detected

15.8% Match

"In conclusion, AcadCheck stands as a modern, dependable, and scalable solution for detecting plagiarism and AI-generated content, striking a balance between technical sophistication, usability, security, and ethical responsibility."

Plagiarism Detected

15.8% Match

"REFERENCES Books / Monographs Bishop, C."

Plagiarism Detected

15.8% Match

"M."

Plagiarism Detected

15.8% Match

"(2006)."

Plagiarism Detected

15.8% Match

"Pattern recognition and machine learning."

Plagiarism Detected `15.8% Match`

"Springer."

Plagiarism Detected `15.8% Match`

"Bird, S., Klein, E., & Loper, E."

Plagiarism Detected `15.8% Match`

"(2009)."

Plagiarism Detected `15.8% Match`

"Natural language processing with Python."

Plagiarism Detected `15.8% Match`

"O'Reilly Media."

Plagiarism Detected `15.8% Match`

"Goodfellow, I., Bengio, Y., & Courville, A."

Plagiarism Detected `15.8% Match`

"(2016)."

AcadCheck - Academic Integrity Report

Plagiarism Detected 15.8% Match

"Deep learning."

Plagiarism Detected 15.8% Match

"MIT Press."

Plagiarism Detected 15.8% Match

"Jurafsky, D., & Martin, J."

Plagiarism Detected 15.8% Match

"H."

Plagiarism Detected 15.8% Match

"(2021)."

Plagiarism Detected 15.8% Match

"Speech and language processing (3rd ed.)."

Plagiarism Detected 15.8% Match

"Pearson."

Plagiarism Detected 15.8% Match

"Journal Articles / Conference Papers Alzahrani, S."

Plagiarism Detected 15.8% Match

"M., Salim, N., & Abraham, A."

Plagiarism Detected 15.8% Match

"(2012)."

Plagiarism Detected 15.8% Match

"Understanding plagiarism: Linguistic patterns, textual features, and detection techniques."

Plagiarism Detected 15.8% Match

"IEEE Transactions on Systems, Man, and Cybernetics, Part C, 42(5), 1339– 1352."

Plagiarism Detected 15.8% Match

"https://doi.org/10.1109/TSMCC.2012.2184410 Potthast, M., Stein, B., Barrón-Cedeño, A., & Rosso, P."

AcadCheck - Academic Integrity Report

Plagiarism Detected 15.8% Match

"(2010)."

Plagiarism Detected 15.8% Match

"An evaluation framework for plagiarism detection."

Plagiarism Detected 15.8% Match

"Proceedings of the 23rd International Conference on Computational Linguistics (COLING), 997–1005."

Plagiarism Detected 15.8% Match

"Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A."

Plagiarism Detected 15.8% Match

"N., Kaiser, Ł., & Polosukhin, I."

Plagiarism Detected 15.8% Match

"(2017)."

Plagiarism Detected 15.8% Match

"Attention is all you need."

Plagiarism Detected　　　　　15.8% Match

"Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008."

Plagiarism Detected　　　　　15.8% Match

"Gao, C., Wang, X., & He, D."

Plagiarism Detected　　　　　15.8% Match

"(2023)."

Plagiarism Detected　　　　　15.8% Match

"Detecting AI-generated text using linguistic features and transformer models."

Plagiarism Detected　　　　　15.8% Match

"Journal of Artificial Intelligence Research, 76, 1201–1220."

Plagiarism Detected　　　　　15.8% Match

"https://doi.org/10.1613/jair.1.13345 Web / Technical Documentation Hugging Face."

Plagiarism Detected | 15.8% Match

"(2025)."

Plagiarism Detected | 15.8% Match

"Transformers documentation."

Plagiarism Detected | 15.8% Match

"Retrieved August 2025, from https://huggingface.co/docs/transformers PDF.js."

Plagiarism Detected | 15.8% Match

"(2025)."

Plagiarism Detected | 15.8% Match

"Mozilla PDF.js project."

Plagiarism Detected | 15.8% Match

"Retrieved August 2025, from https://mozilla.github.io/pdf.js/ Supabase."

Plagiarism Detected | 15.8% Match

"(2025)."

AcadCheck - Academic Integrity Report

Plagiarism Detected   15.8% Match

"Supabase documentation."

Plagiarism Detected   15.8% Match

"Retrieved August 2025, from https://supabase.com/docs Vite."

Plagiarism Detected   15.8% Match

"(2025)."

Plagiarism Detected   15.8% Match

"Vite documentation."

Plagiarism Detected   15.8% Match

"Retrieved August 2025, from https://vitejs.dev/ Martin, J., & Ha, D."

Plagiarism Detected   15.8% Match

"(2022)."

Plagiarism Detected   15.8% Match

"Detecting AI-generated content: Challenges and strategies."

Plagiarism Detected | 15.8% Match

"Medium."

Plagiarism Detected | 15.8% Match

"Retrieved from https://medium.com Standards / Guidelines European Union."

Plagiarism Detected | 15.8% Match

"(2018)."

Plagiarism Detected | 15.8% Match

"General Data Protection Regulation (GDPR)."

Plagiarism Detected | 15.8% Match

"Official Journal of the European Union."

Plagiarism Detected | 15.8% Match

"Association for Computing Machinery (ACM)."

Plagiarism Detected | 15.8% Match

"(2021)."

AcadCheck - Academic Integrity Report

Plagiarism Detected

"ACM code of ethics and professional conduct."

Plagiarism Detected

"Retrieved from https://www.acm.org/code-of-ethics APENDIX This appendix presents visual documentation of the main features of the AcadCheck application developed during the internship."

Plagiarism Detected

"The figures illustrate the user interface, examples of plagiarism and AI detection results, and the report generation functionality."

Plagiarism Detected

"All displayed content was created within a controlled academic setting and is intended solely for educational purposes."

Plagiarism Detected

"Figure A.1 – Login and Authentication Interface The screenshot displays AcadCheck's secure login page."

Plagiarism Detected

"Users can sign in using their email and password to access the dashboard."

Plagiarism Detected   15.8% Match

"The page also provides options for account registration, password recovery, and selecting the interface language (English or French)."

Plagiarism Detected   15.8% Match

"Figure A.2 – Main Dashboard Overview This figure shows the main dashboard of AcadCheck, where users can access the core features, upload documents, and view summaries of plagiarism and AI detection results."

Plagiarism Detected   15.8% Match

"Figure A.3 – Document Preview (PDF/DOCX/TXT) This screenshot displays the DocumentViewer feature, which lets users preview uploaded documents in real time."

Plagiarism Detected   15.8% Match

"The component maintains the original formatting and offers zoom and navigation options for a thorough review."

Plagiarism Detected   15.8% Match

"Figure A.4 – Plagiarism and AI Detection Results The figure shows sentence-by-sentence analysis, with plagiarism and AI-generated content scores."

Plagiarism Detected　　　　　　　　　　`15.8% Match`

"Color-coded highlights indicate suspicious sections, while confidence metrics provide additional insight for each sentence."

Plagiarism Detected　　　　　　　　　　`15.8% Match`

"Figure A.5 – Detailed Report Generation This screenshot presents the PDF report generated by AcadCheck."

Plagiarism Detected　　　　　　　　　　`15.8% Match`

"It includes global metrics, sentence-level scores, and highlighted passages."

Plagiarism Detected　　　　　　　　　　`15.8% Match`

"Users can download or share the report directly from the dashboard."

Plagiarism Detected　　　　　　　　　　`15.8% Match`

"Figure A.6 – Analysis History and Filtering This figure shows the historical analysis interface, listing previous uploads with dates, document names, and scores."

Plagiarism Detected　　　　　　　　　　`15.8% Match`

"Users can filter, search, and re-open previous reports for reference or export."

AcadCheck - Academic Integrity Report

AI-Generated Content

37.4% Confidence

"GIRNE AMERICAN UNIVERSITY Faculty of Engineering Department of Computer Engineering AcadCheck: AI and Plagiarism Detection Tool SUMMER TRAINING REPORT Student: Elie katende kazwela Student ID: 221701001 Supervisor: Mr."

AI-Generated Content

37.4% Confidence

"Cliff Acknowledgements I would like to sincerely thank all those who supported me during my internship at the university and throughout the preparation of this report."

AI-Generated Content

37.4% Confidence

"First and foremost , I am especially grateful to my supervisor , Mr."

AI-Generated Content

37.4% Confidence

"Peter Cliff, for his guidance ,encouragement and constant support."

AI-Generated Content

37.4% Confidence

"His expertise and thoughtful feedback played a key role in the progress and success of my work."

AI-Generated Content 37.4% Confidence

"I also wish to express my appreciation to the staff and faculty of the university for welcoming me , Providing the necessary resources, and creating a positive environment that made this internship a valuable learning experience."

AI-Generated Content 37.4% Confidence

"the knowledge and skills I gained have been highly beneficial to my academic and professional growth."

AI-Generated Content 37.4% Confidence

"Lastly ,I would like to thank my family and friends for their patience , encouragement , and continuous Support , which greatly helped me in completing this project."

AI-Generated Content 37.4% Confidence

"Abstract This report introduces AcadCheck, a web application developed to detect both plagiarism and AI generated content in academic documents ."

AI-Generated Content 37.4% Confidence

"the system was implemented using React, TypeScript, and Supabase, and it integrates advanced document analysis methods with a user-friendly interface to deliver precise and comprehensive results."

AcadCheck - Academic Integrity Report

AI-Generated Content   37.4% Confidence

"The main objectives of the project are structured around four core component : Plagiarism detection : comparing documents against a large corpus by applying N-grams, semantic similarity, and pattern-matching techniques."

AI-Generated Content   37.4% Confidence

"AI content detection – identifying AI-generated text through transformer-based models and linguistic analysis."

AI-Generated Content   37.4% Confidence

"Document analysis and reporting – producing detailed, sentence-level feedback with metrics such as lexical diversity, syntactic complexity, semantic coherence, and perplexity."

AI-Generated Content   37.4% Confidence

"Secure and scalable architecture – leveraging Supabase for database management, data protection, and Row Level Security (RLS)."

AI-Generated Content   37.4% Confidence

"This report presents the overall architecture , development process ,challenges ,faced and evaluation results of the application ,while also reflecting on the ethical considerations involved .it highlights the use of modern approaches to document analysis , the integration of artificial intelligence in plagiarism detection ,and the design of scalable web applications."

AI-Generated Content

37.4% Confidence

"Chapter 1: Introduction Background In academic settings, maintaining the integrity of written work is essential."

AI-Generated Content

37.4% Confidence

"The growing use of digital tools, online resources, and AI-assisted content creation has heightened the risk of plagiarism and the improper use of external materials."

AI-Generated Content

37.4% Confidence

"AcadCheck tackles these issues by offering an automated solution that can identify both plagiarized content and text generated by artificial intelligence, helping to uphold academic honesty and the quality of scholarly work."

AI-Generated Content

37.4% Confidence

"Contemporary document analysis relies on advanced technologies such as machine learning, semantic evaluation, and natural language processing (NLP)."

AI-Generated Content

37.4% Confidence

"By combining these methods with a straightforward and intuitive web interface, AcadCheck enables educators, students, and researchers to examine documents effectively while ensuring the security and privacy of their data."

AI-Generated Content
37.4% Confidence

"1.2 Objectives The primary goals of this internship project are: To create a web application capable of identifying both plagiarism and AI-generated content."

AI-Generated Content
37.4% Confidence

"To integrate sophisticated analysis techniques, combining machine learning with heuristic methods."

AI-Generated Content
37.4% Confidence

"To provide comprehensive reporting, including analysis at the sentence level, confidence indicators, and clear visual cues."

AI-Generated Content
37.4% Confidence

"To guarantee secure management of user data, leveraging authentication, Row-Level Security (RLS), and protected storage."

AI-Generated Content
37.4% Confidence

"To deliver a user-friendly and responsive interface that accommodates multilingual users effectively."

AI-Generated Content
37.4% Confidence

"1.3 Scope This project centers on the design, development, and implementation of AcadCheck as a tool for research and education."

AI-Generated Content
`37.4% Confidence`

"The application is compatible with PDF, DOCX, and TXT file formats, and it offers detailed sentence-level analysis for both plagiarism and AI-generated content."

AI-Generated Content
`37.4% Confidence`

"All testing and usage of the application are carried out ethically, within controlled environments, and in strict accordance with data protection regulations."

AI-Generated Content
`37.4% Confidence`

"The project does not involve deployment in untrusted environments or the collection of unauthorized data."

AI-Generated Content
`37.4% Confidence`

"1.4 Importance of the Study AcadCheck contributes to multiple domains: Academic integrity: Helps institutions maintain originality standards."

AI-Generated Content
`37.4% Confidence`

"AI literacy: Educates users on distinguishing AI-generated content."

AI-Generated Content

37.4% Confidence

"Technical education: Demonstrates the use of modern web technologies, machine learning, and database security in practical applications."

AI-Generated Content

37.4% Confidence

"By integrating AI detection with plagiarism analysis, the project addresses modern challenges in digital content evaluation, providing insights for students, educators, and software developers."

AI-Generated Content

37.4% Confidence

"1.5 Structure of the Report The organization of the report is: Chapter 1: Introduction – Presents background, objectives, scope, and significance."

AI-Generated Content

37.4% Confidence

"Chapter 2: Literature Review – Reviews plagiarism detection methods, AI content detection, and ethical/legal considerations."

AI-Generated Content

37.4% Confidence

"Chapter 3: Methodology – Details the system architecture, tools, libraries, and implementation steps."

AcadCheck - Academic Integrity Report

AI-Generated Content

37.4% Confidence

"Chapter 4: Implementation and Discussion – Covers code implementation, document processing, analysis workflow, and troubleshooting."

AI-Generated Content

37.4% Confidence

"Chapter 5: Conclusion – Summarizes findings, limitations, educational value, and future improvements."

AI-Generated Content

37.4% Confidence

"References – List of sources cited throughout the report."

AI-Generated Content

37.4% Confidence

"CHAPTER 2: LITERATURE REVIEW 2.1 Overview of Plagiarism Detection Detecting plagiarism is essential for upholding academic integrity."

AI-Generated Content

37.4% Confidence

"Traditional tools mainly rely on exact text matching, which can spot direct copying but often misses paraphrased or slightly altered content."

AI-Generated Content

37.4% Confidence

"Modern systems, like AcadCheck, use a combination of syntactic, semantic, and contextual analysis, enabling more accurate detection of potential plagiarism."

AcadCheck - Academic Integrity Report

AI-Generated Content 37.4% Confidence

"N-gram Analysis: Documents are broken into sequences of words (n-grams) and compared to a reference corpus."

AI-Generated Content 37.4% Confidence

"Similar sequences indicate potential plagiarism."

AI-Generated Content 37.4% Confidence

"AcadCheck uses adaptive n-gram sizes to capture both short phrases and longer patterns."

AI-Generated Content 37.4% Confidence

"Semantic Similarity: Using NLP techniques, AcadCheck compares the meaning of sentences rather than exact wording."

AI-Generated Content 37.4% Confidence

"This allows the system to detect paraphrased content."

AI-Generated Content 37.4% Confidence

"Contextual Patterns: Certain academic formulations are commonly reused."

AI-Generated Content

"By identifying these patterns, the system can differentiate between general academic phrases and copied unique content."

AI-Generated Content

"Recent research emphasizes the integration of machine learning models for more nuanced plagiarism detection, allowing systems to learn from labeled datasets and improve detection accuracy over time."

AI-Generated Content

"2.2 AI-Based Text Detection With the emergence of AI-generated content, particularly large language models like GPT, distinguishing between human-written and AI-generated text has become a crucial challenge."

AI-Generated Content

"AcadCheck implements hybrid AI detection techniques, combining transformer-based models with linguistic heuristics: Transformer Models: Pre-trained models analyze text for patterns indicative of AI generation, such as overuse of certain connectors, predictable phrasing, or low variability."

AI-Generated Content

"Linguistic Heuristics: Metrics such as lexical diversity, syntactic complexity, and perplexity are used to evaluate the naturalness of text."

AcadCheck - Academic Integrity Report

AI-Generated Content `37.4% Confidence`

"AI-generated text often exhibits high perplexity uniformity and low lexical diversity."

AI-Generated Content `37.4% Confidence`

"Sentence-Level Scoring: Each sentence is assigned an AI score, indicating the likelihood of machine generation."

AI-Generated Content `37.4% Confidence`

"This granularity allows users to review suspicious content precisely."

AI-Generated Content `37.4% Confidence`

"By integrating AI detection with plagiarism analysis, AcadCheck provides a comprehensive evaluation of document originality."

AI-Generated Content `37.4% Confidence`

"2.3 Document Analysis Techniques Effective plagiarism and AI detection require robust document analysis."

AI-Generated Content `37.4% Confidence`

"AcadCheck supports multiple formats, including PDF, DOCX, and TXT, with consistent text extraction: PDF Analysis: Using PDF.js, the system extracts text while preserving document structure."

AI-Generated Content   37.4% Confidence

"Word Document Handling: Mammoth and DOCX Preview convert DOCX files to HTML, ensuring accurate rendering and sentence-level analysis."

AI-Generated Content   37.4% Confidence

"Text Preprocessing: Cleaning, tokenization, and normalization prepare the content for further analysis."

AI-Generated Content   37.4% Confidence

"Sentence segmentation, feature extraction, and vectorization are applied before any similarity or AI detection algorithms, ensuring that results are both accurate and interpretable."

AI-Generated Content   37.4% Confidence

"2.4 Ethical and Legal Considerations Given the sensitivity of plagiarism and AI detection, ethics and legality are central to AcadCheck's design: User Consent: Only documents uploaded with explicit consent are analyzed."

AI-Generated Content   37.4% Confidence

"Data Security: Supabase provides encrypted storage and Row Level Security (RLS) to isolate user data."

AI-Generated Content

37.4% Confidence

"Legal Compliance: The system aligns with GDPR and other data protection regulations, preventing unauthorized access or retention."

AI-Generated Content

37.4% Confidence

"Responsible Reporting: Results are presented solely for educational or institutional purposes, avoiding misuse of the analysis."

AI-Generated Content

37.4% Confidence

"Ethical considerations also guide the inclusion of AI detection."

AI-Generated Content

37.4% Confidence

"The system is designed for transparency, providing users with interpretable metrics and avoiding punitive assumptions."

AI-Generated Content

37.4% Confidence

"2.5 Use of React, TypeScript, and Supabase in Web Apps Modern web technologies play a key role in the functionality and usability of AcadCheck: React (v18.3.1): Provides a responsive, dynamic frontend capable of handling real-time document previews and interactive dashboards."

AI-Generated Content

37.4% Confidence

"TypeScript: Adds type safety and improves maintainability, especially important for complex components such as analysis tables and reports."

AI-Generated Content

37.4% Confidence

"Supabase: Serves as a backend and database solution, offering authentication, secure storage, and serverless functions."

AI-Generated Content

37.4% Confidence

"RLS ensures that users only access their own analyses."

AI-Generated Content

37.4% Confidence

"Frontend UI Components: Using Tailwind CSS and shadcn/ui, the interface is modern, clean, and intuitive."

AI-Generated Content

37.4% Confidence

"Components like DocumentViewer and HighlightedText provide detailed visual feedback to users."

AI-Generated Content

37.4% Confidence

"The combination of these technologies allows for a robust, scalable, and secure web application, supporting complex AI and plagiarism detection workflows while maintaining excellent user experience."

AI-Generated Content

37.4% Confidence

"2.6 Integration of Detection Algorithms in Web Applications A critical aspect of modern plagiarism detection systems is the seamless integration of backend analysis algorithms with frontend visualization: Sentence-Level Analysis: Each sentence is analyzed for plagiarism and AI content."

AI-Generated Content

37.4% Confidence

"Results are transmitted from the backend to the frontend, where they are visually highlighted in the document."

AI-Generated Content

37.4% Confidence

"Scoring and Metrics: Global scores, confidence levels, and detailed metrics are presented using interactive dashboards."

AI-Generated Content

37.4% Confidence

"Real-Time Feedback: Users can upload documents and receive analysis results without page reloads, thanks to React's component architecture."

AI-Generated Content

37.4% Confidence

"This integration ensures that technical sophistication in analysis does not compromise usability."

AI-Generated Content | 37.4% Confidence

"2.7 Summary Chapter 2 shows that modern plagiarism and AI detection relies on multi-layered methods that combine semantic analysis, machine learning, and secure web technologies."

AI-Generated Content | 37.4% Confidence

"AcadCheck applies these principles by providing a system that is easy to use, secure, and accurate, effectively tackling today's challenges in maintaining academic integrity."

AI-Generated Content | 37.4% Confidence

"CHAPTER 3: METHODOLOGY 3.1 Design Objectives and System Architecture The main goal in designing AcadCheck is to build a web application that is secure, scalable, and easy to use for detecting plagiarism and AI-generated content."

AI-Generated Content | 37.4% Confidence

"The key aspects taken into account include: Accuracy: Implement advanced algorithms for precise detection of plagiarized and AI-generated text."

AI-Generated Content | 37.4% Confidence

"Security: Ensure strict user data isolation through authentication and Row Level Security (RLS)."

AI-Generated Content

"Usability: Provide an intuitive dashboard and interactive reports that can be easily interpreted by students, educators, and researchers."

AI-Generated Content

"Scalability: Design the system to handle large volumes of document uploads and concurrent analyses."

AI-Generated Content

"The system architecture is composed of the following layers: Frontend (React + TypeScript): Handles user interaction, document preview, and visualization of analysis results."

AI-Generated Content

"Backend (Supabase + Serverless Functions): Processes document uploads, manages user authentication, and executes analysis algorithms."

AI-Generated Content

"Database (PostgreSQL): Stores user data, documents, analysis results, and sentence-level metrics."

AI-Generated Content   37.4% Confidence

"AI Engine: Implements transformer-based models and linguistic heuristics for AI detection."

AI-Generated Content   37.4% Confidence

"Security Layer: Enforces RLS, secure storage, and access control to protect sensitive data."

AI-Generated Content   37.4% Confidence

"The architecture is modular, allowing future integration of additional features such as multi-language support, paraphrasing detection, and integration with LMS platforms."

AI-Generated Content

37.4% Confidence

"3.2 Tools and Libraries The following tools and libraries were selected for their suitability and robustness: Frontend: React 18.3.1 for component-based UI TypeScript for type safety and maintainability Tailwind CSS and shadcn/ui for consistent, responsive styling React Router DOM for client-side routing Backend & Database: Supabase (PostgreSQL) for secure user management, storage, and database operations Serverless functions for document processing Document Processing: PDF.js for PDF extraction Mammoth and DOCX Preview for Word document conversion and rendering Custom fileToText library for preprocessing and normalization AI & Plagiarism Detection: Hugging Face Transformers for AI detection N-gram analysis, semantic similarity, and heuristic algorithms for plagiarism detection Testing & Quality: Jest and React Testing Library for unit tests Cypress for integration tests ESLint, Prettier, and Husky for code quality and consistency 3.3 Development and Implementation The development of AcadCheck was divided into several modules to ensure modularity and maintainability."

AI-Generated Content

37.4% Confidence

"3.3.1 User Authentication and Session Management Sign-up/Login: Users register with email and password."

AI-Generated Content

37.4% Confidence

"Supabase handles authentication securely."

AI-Generated Content

37.4% Confidence

"Session Persistence: JWT tokens are stored in browser local storage, with automatic session renewal."

AI-Generated Content   37.4% Confidence

"Route Protection: Only authenticated users can access dashboards and analysis features."

AI-Generated Content   37.4% Confidence

"3.3.2 Document Upload and Preprocessing Supported Formats: PDF, DOCX, and TXT File Validation: Ensures correct format and maximum file size compliance Preprocessing Steps: Text extraction from uploaded files Sentence segmentation and tokenization Removal of unnecessary characters or formatting issues 3.3.3 Plagiarism Detection Algorithms Exact Matching: Detects copied content using n-grams (3-5 words)."

AI-Generated Content   37.4% Confidence

"Semantic Similarity: Compares sentences against reference corpus using vector embeddings."

AI-Generated Content   37.4% Confidence

"Contextual Analysis: Identifies repeated academic phrases to minimize false positives."

AI-Generated Content   37.4% Confidence

"3.3.4 AI Content Detection Transformer Models: Analyzes text for patterns indicative of AI generation."

AI-Generated Content

"Stylistic Metrics: Computes lexical diversity, syntactic complexity, semantic coherence, and perplexity."

AI-Generated Content

"Sentence-Level Scoring: Each sentence receives a probability score for AI authorship."

AI-Generated Content

"3.3.5 Analysis Report Generation Visual Dashboard: Displays sentence-level highlights and overall metrics."

AI-Generated Content

"Export Options: Users can download PDF reports with annotated content."

AI-Generated Content

"Confidence Scores: Metrics such as AI score, plagiarism score, and confidence levels guide interpretation."

AI-Generated Content

"3.3.6 Internationalization Languages Supported: English and French Dynamic Switching: Users can switch languages without refreshing the page Persistence: Language preferences saved in user profile 3.3.7 Security and RLS Policies Row Level Security (RLS): Ensures users can only access their own analyses Encrypted Storage: All uploaded documents are stored securely in Supabase buckets Access Control: Temporary signed URLs prevent unauthorized access to stored files 3.4 Packaging and Deployment Frontend Build: Using Vite for optimized bundle creation Backend Deployment: Supabase handles serverless functions and database management Environment Variables: Secured for production, including Supabase API keys and AI model endpoints CI/CD Pipeline: Automated deployment with GitHub Actions and version control 3.5 Testing and Quality Assurance Unit Tests: Validate individual components and algorithm functions Integration Tests: Ensure end-to-end functionality from document upload to report generation Performance Tests: Analyze large document handling and real-time AI scoring Code Quality: ESLint, Prettier, and Husky ensure consistent, maintainable code 3.6 Ethical and Legal Considerations All analyses performed on user-uploaded documents with explicit consent."

AI-Generated Content

"Compliance with GDPR and local privacy laws."

AI-Generated Content

"System designed for educational and research purposes only."

AI-Generated Content   37.4% Confidence

"Users receive interpretable results, avoiding punitive or misrepresentative conclusions."

AI-Generated Content   37.4% Confidence

"3.7 Summary Chapter 3 provides a detailed account of AcadCheck's methodology, covering system architecture, tools, implementation, security, testing, and ethical considerations."

AI-Generated Content   37.4% Confidence

"This methodology ensures accuracy, reliability, and usability, forming the foundation for Chapters 4 and 5, which present the actual implementation and results."

AI-Generated Content   37.4% Confidence

"CHAPTER 4: IMPLEMENTATION AND DISCUSSION 4.1 Development Environment Setup The development of AcadCheck was performed using a modern web stack on a local and cloud-based environment: Operating System: Windows 11 / Ubuntu 22.04 Frontend: Node.js 20, Vite, React 18.3.1, TypeScript Backend & Database: Supabase (PostgreSQL), serverless functions for analysis processing IDE: Visual Studio Code with ESLint, Prettier, and Git integration Version Control: GitHub for source code management and CI/CD pipelines All dependencies were installed using npm or yarn."

AI-Generated Content `37.4% Confidence`

"Environment variables were securely configured for API keys, database URLs, and AI model endpoints."

AI-Generated Content `37.4% Confidence`

"4.2 Frontend Implementation The frontend was implemented using React and TypeScript, emphasizing responsive design and real-time feedback."

AI-Generated Content `37.4% Confidence`

"4.2.1 Dashboard and Document Upload Drag-and-Drop Upload: Users can upload PDF, DOCX, or TXT files."

AI-Generated Content `37.4% Confidence`

"Live Preview: Using DocumentViewer, the uploaded file is displayed with high fidelity."

AI-Generated Content `37.4% Confidence`

"Upload Validation: The system checks file type, size, and integrity before analysis."

AI-Generated Content

37.4% Confidence

"const handleFileUpload = async (file: File) => { if (!isValidFile(file)) return; const text = await fileToText(file); setUploadedText(text); }; 4.2.2 Interactive Analysis Report Sentence-level highlights indicate plagiarism and AI-generated content."

AI-Generated Content

37.4% Confidence

"Users can hover over sentences to view metrics such as confidence scores and AI probability."

AI-Generated Content

37.4% Confidence

"Results are dynamically updated without page reloads using React hooks."

AI-Generated Content

37.4% Confidence

"4.3 Backend Implementation The backend is built with Supabase, providing secure storage, authentication, and serverless analysis functions."

AI-Generated Content

37.4% Confidence

"4.3.1 Document Storage Uploaded documents are stored in encrypted buckets with temporary signed URLs."

AcadCheck - Academic Integrity Report

AI-Generated Content

"Metadata, including document name, user ID, and timestamp, is stored in the analyses table."

AI-Generated Content

"4.3.2 Analysis Workflow Text Extraction: Using PDF.js and Mammoth for PDFs and DOCX respectively."

AI-Generated Content

"Preprocessing: Tokenization, sentence segmentation, and cleaning."

AI-Generated Content

"Plagiarism Detection: N-grams and semantic similarity comparison with corpus documents."

AI-Generated Content

"AI Detection: Transformer models assign a probability score per sentence."

AI-Generated Content

"const analysisResult = await analyzeText(extractedText, { corpus: referenceDocuments, useAI: true }); saveAnalysis(userId, analysisResult); 4.4 Document Analysis Workflow User uploads document → Frontend validates → Sends to backend."

AI-Generated Content `37.4% Confidence`

"Backend extracts text and segments into sentences."

AI-Generated Content `37.4% Confidence`

"Each sentence is processed: Plagiarism score calculated AI detection score calculated Results are stored in analysis_sentences table."

AI-Generated Content `37.4% Confidence`

"Frontend renders interactive report with color-coded highlights."

AI-Generated Content `37.4% Confidence`

"Workflow Diagram: [User Upload] → [Text Extraction] → [Sentence Segmentation] → [AI & Plagiarism Analysis] → [Database Storage] → [Interactive Report] 4.5 AI and Plagiarism Detection Results Sentence-Level Scores: Each sentence is scored 0-100% for plagiarism and AI probability."

AI-Generated Content `37.4% Confidence`

"Global Metrics: Average plagiarism score, average AI score, lexical diversity, syntactic complexity."

AI-Generated Content

"Visual Feedback: Red highlight → High plagiarism Yellow highlight → Medium plagiarism or AI probability Green → Safe or low risk Example Table: 4.6 Challenges and Troubleshooting During implementation, several challenges were encountered: Document Rendering Issues: Some DOCX formatting was lost; resolved using Mammoth and DOCX Preview."

AI-Generated Content

"Large File Handling: PDFs > 50MB caused slow extraction; solved by batch processing sentences."

AI-Generated Content

"AI Model Latency: Transformer-based scoring was time-consuming; optimized using caching and async functions."

AI-Generated Content

"Data Security: Careful attention was given to applying Row-Level Security (RLS) correctly across all tables."

AI-Generated Content

"To ensure proper functionality, thorough testing was conducted, including unit tests, integration tests, and detailed logging."

AI-Generated Content   `37.4% Confidence`

> *"4.7 Ethical and Legal Considerations All analyses respect user privacy."*

AI-Generated Content   `37.4% Confidence`

> *"Temporary URLs prevent unauthorized document access."*

AI-Generated Content   `37.4% Confidence`

> *"System designed for demonstration and educational purposes, avoiding legal infringement."*

AI-Generated Content   `37.4% Confidence`

> *"Detailed logs and transparency ensure users can trust results."*

AI-Generated Content   `37.4% Confidence`

> *"4.8 Summary Chapter 4 presents the hands-on implementation of AcadCheck, covering the frontend and backend development, document processing workflows, AI and plagiarism detection features, and strategies for troubleshooting."*

AI-Generated Content   `37.4% Confidence`

> *"It also discusses the real-world challenges encountered when integrating complex algorithms into a modern web application, while ensuring security, scalability, and a smooth user experience."*

AcadCheck - Academic Integrity Report

AI-Generated Content

"CHAPTER 5: CONCLUSION 5.1 Key Findings The development and deployment of AcadCheck led to several key insights: Accurate Plagiarism Detection: By combining n-gram techniques, semantic comparisons, and contextual pattern recognition, AcadCheck can reliably detect both copied and paraphrased content."

AI-Generated Content

"AI Content Recognition: The use of transformer-based models alongside heuristic methods allows for precise identification of AI-generated text, with detailed analysis at the sentence level."

AI-Generated Content

"Intuitive User Interface: The interactive dashboard and real-time feedback make the results easy to understand, enhancing the overall user experience."

AI-Generated Content

"Secure and Scalable System: Built with Supabase, React, and TypeScript, the platform ensures data privacy, user authentication, and the ability to handle large volumes of documents efficiently."

AI-Generated Content

"Support for Multiple File Formats: PDFs, DOCX, and TXT files are processed while maintaining formatting, ensuring accurate and consistent analysis."

AI-Generated Content `37.4% Confidence`

"These observations highlight that tools for maintaining academic integrity must combine technical sophistication with user-centered design to be truly effective."

AI-Generated Content `37.4% Confidence`

"5.2 Educational Value AcadCheck functions as a practical tool for students, researchers, and educators alike: Learning Resource: Students can check their own work for accidental plagiarism or AI-generated content, helping them improve their writing habits."

AI-Generated Content `37.4% Confidence`

"Teaching Aid: Educators can use the platform to illustrate proper citation methods and maintain academic writing standards."

AI-Generated Content `37.4% Confidence`

"Research Support: The application provides detailed metrics and visualizations that facilitate studies on academic integrity."

AI-Generated Content `37.4% Confidence`

"Additionally, the development of AcadCheck offered valuable hands-on experience in full-stack web development, AI integration, and implementing secure data management practices."

AI-Generated Content   37.4% Confidence

"5.3 Limitations Despite its achievements, AcadCheck has some limitations: AI Detection Accuracy: Although generally reliable, the system may occasionally flag highly formal or technical writing as AI-generated."

AI-Generated Content   37.4% Confidence

"Corpus Dependency: The accuracy of plagiarism detection depends heavily on the size and variety of the reference corpus."

AI-Generated Content   37.4% Confidence

"Processing Time: Analyzing large documents or multiple files at once can take longer, even with performance optimizations in place."

AI-Generated Content   37.4% Confidence

"Language Support: Currently, the application only handles English and French; supporting additional languages would require retraining the models."

AI-Generated Content   37.4% Confidence

"Being aware of these limitations helps guide improvements and refinements for future versions of AcadCheck."

AI-Generated Content 37.4% Confidence

"5.4 Recommendations for Future Work Future development of AcadCheck could focus on several key areas: Expanded Multilingual Support: Incorporating additional languages such as Spanish and German to make the tool accessible to a wider audience."

AI-Generated Content 37.4% Confidence

"Integration with Learning Management Systems (LMS): Enabling direct connections to platforms like Moodle, Canvas, and Blackboard for seamless automated analysis."

AI-Generated Content 37.4% Confidence

"Enhanced Paraphrasing Detection: Implementing more advanced semantic analysis techniques to identify subtle paraphrasing in student submissions."

AI-Generated Content 37.4% Confidence

"Mobile Application: Creating a React Native version to allow users to access AcadCheck on smartphones and tablets."

AI-Generated Content 37.4% Confidence

"API for External Tools: Providing a public API so third-party applications can utilize AcadCheck's detection algorithms."

AI-Generated Content
37.4% Confidence

"Performance Improvements: Adopting microservices architecture and queue-based processing to efficiently handle large volumes of documents."

AI-Generated Content
37.4% Confidence

"These improvements would further enhance AcadCheck's value as a reliable and professional tool for academic integrity and research support."

AI-Generated Content
37.4% Confidence

"5.5 Final Thoughts AcadCheck showcases how advanced AI, plagiarism detection, and secure web technologies can be combined into a practical, user-friendly platform."

AI-Generated Content
37.4% Confidence

"It demonstrates that technology can support academic integrity in a way that is both ethical and transparent."

AI-Generated Content
37.4% Confidence

"The development of this project offered valuable hands-on learning experiences, covering everything from system architecture and algorithm implementation to frontend and backend integration, as well as data security management."

AI-Generated Content

37.4% Confidence

"The resulting application not only meets its intended goals but also provides a solid foundation for further research and innovation in academic integrity tools."

AI-Generated Content

37.4% Confidence

"In conclusion, AcadCheck stands as a modern, dependable, and scalable solution for detecting plagiarism and AI-generated content, striking a balance between technical sophistication, usability, security, and ethical responsibility."

AI-Generated Content

37.4% Confidence

"REFERENCES Books / Monographs Bishop, C."

AI-Generated Content

37.4% Confidence

"M."

AI-Generated Content

37.4% Confidence

"(2006)."

AI-Generated Content

37.4% Confidence

"Pattern recognition and machine learning."

AI-Generated Content `37.4% Confidence`

"Springer."

AI-Generated Content `37.4% Confidence`

"Bird, S., Klein, E., & Loper, E."

AI-Generated Content `37.4% Confidence`

"(2009)."

AI-Generated Content `37.4% Confidence`

"Natural language processing with Python."

AI-Generated Content `37.4% Confidence`

"O'Reilly Media."

AI-Generated Content `37.4% Confidence`

"Goodfellow, I., Bengio, Y., & Courville, A."

AI-Generated Content `37.4% Confidence`

"(2016)."

AI-Generated Content   37.4% Confidence

"Deep learning."

AI-Generated Content   37.4% Confidence

"MIT Press."

AI-Generated Content   37.4% Confidence

"Jurafsky, D., & Martin, J."

AI-Generated Content   37.4% Confidence

"H."

AI-Generated Content   37.4% Confidence

"(2021)."

AI-Generated Content   37.4% Confidence

"Speech and language processing (3rd ed.)."

AI-Generated Content   37.4% Confidence

"Pearson."

AI-Generated Content   37.4% Confidence

"Journal Articles / Conference Papers Alzahrani, S."

AI-Generated Content   37.4% Confidence

"M., Salim, N., & Abraham, A."

AI-Generated Content   37.4% Confidence

"(2012)."

AI-Generated Content   37.4% Confidence

"Understanding plagiarism: Linguistic patterns, textual features, and detection techniques."

AI-Generated Content   37.4% Confidence

"IEEE Transactions on Systems, Man, and Cybernetics, Part C, 42(5), 1339–1352."

AI-Generated Content   37.4% Confidence

"https://doi.org/10.1109/TSMCC.2012.2184410 Potthast, M., Stein, B., Barrón-Cedeño, A., & Rosso, P."

AcadCheck - Academic Integrity Report

AI-Generated Content — 37.4% Confidence

"(2010)."

AI-Generated Content — 37.4% Confidence

"An evaluation framework for plagiarism detection."

AI-Generated Content — 37.4% Confidence

"Proceedings of the 23rd International Conference on Computational Linguistics (COLING), 997–1005."

AI-Generated Content — 37.4% Confidence

"Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A."

AI-Generated Content — 37.4% Confidence

"N., Kaiser, Ł., & Polosukhin, I."

AI-Generated Content — 37.4% Confidence

"(2017)."

AI-Generated Content — 37.4% Confidence

"Attention is all you need."

AcadCheck - Academic Integrity Report

AI-Generated Content

"Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008."

37.4% Confidence

AI-Generated Content

"Gao, C., Wang, X., & He, D."

37.4% Confidence

AI-Generated Content

"(2023)."

37.4% Confidence

AI-Generated Content

"Detecting AI-generated text using linguistic features and transformer models."

37.4% Confidence

AI-Generated Content

"Journal of Artificial Intelligence Research, 76, 1201–1220."

37.4% Confidence

AI-Generated Content

"https://doi.org/10.1613/jair.1.13345 Web / Technical Documentation Hugging Face."

37.4% Confidence

AcadCheck - Academic Integrity Report

AI-Generated Content 37.4% Confidence

"(2025)."

AI-Generated Content 37.4% Confidence

"Transformers documentation."

AI-Generated Content 37.4% Confidence

"Retrieved August 2025, from https://huggingface.co/docs/transformers PDF.js."

AI-Generated Content 37.4% Confidence

"(2025)."

AI-Generated Content 37.4% Confidence

"Mozilla PDF.js project."

AI-Generated Content 37.4% Confidence

"Retrieved August 2025, from https://mozilla.github.io/pdf.js/ Supabase."

AI-Generated Content 37.4% Confidence

"(2025)."

AcadCheck - Academic Integrity Report

AI-Generated Content · 37.4% Confidence

"Supabase documentation."

AI-Generated Content · 37.4% Confidence

"Retrieved August 2025, from https://supabase.com/docs Vite."

AI-Generated Content · 37.4% Confidence

"(2025)."

AI-Generated Content · 37.4% Confidence

"Vite documentation."

AI-Generated Content · 37.4% Confidence

"Retrieved August 2025, from https://vitejs.dev/ Martin, J., & Ha, D."

AI-Generated Content · 37.4% Confidence

"(2022)."

AI-Generated Content · 37.4% Confidence

"Detecting AI-generated content: Challenges and strategies."

AcadCheck - Academic Integrity Report

AI-Generated Content

37.4% Confidence

"Medium."

AI-Generated Content

37.4% Confidence

"Retrieved from https://medium.com Standards / Guidelines European Union."

AI-Generated Content

37.4% Confidence

"(2018)."

AI-Generated Content

37.4% Confidence

"General Data Protection Regulation (GDPR)."

AI-Generated Content

37.4% Confidence

"Official Journal of the European Union."

AI-Generated Content

37.4% Confidence

"Association for Computing Machinery (ACM)."

AI-Generated Content

37.4% Confidence

"(2021)."

AcadCheck - Academic Integrity Report

AI-Generated Content

"ACM code of ethics and professional conduct."

AI-Generated Content

"Retrieved from https://www.acm.org/code-of-ethics APENDIX This appendix presents visual documentation of the main features of the AcadCheck application developed during the internship."

AI-Generated Content

"The figures illustrate the user interface, examples of plagiarism and AI detection results, and the report generation functionality."

AI-Generated Content

"All displayed content was created within a controlled academic setting and is intended solely for educational purposes."

AI-Generated Content

"Figure A.1 – Login and Authentication Interface The screenshot displays AcadCheck's secure login page."

AI-Generated Content

"Users can sign in using their email and password to access the dashboard."

AI-Generated Content

37.4% Confidence

"The page also provides options for account registration, password recovery, and selecting the interface language (English or French)."

AI-Generated Content

37.4% Confidence

"Figure A.2 – Main Dashboard Overview This figure shows the main dashboard of AcadCheck, where users can access the core features, upload documents, and view summaries of plagiarism and AI detection results."

AI-Generated Content

37.4% Confidence

"Figure A.3 – Document Preview (PDF/DOCX/TXT) This screenshot displays the DocumentViewer feature, which lets users preview uploaded documents in real time."

AI-Generated Content

37.4% Confidence

"The component maintains the original formatting and offers zoom and navigation options for a thorough review."

AI-Generated Content

37.4% Confidence

"Figure A.4 – Plagiarism and AI Detection Results The figure shows sentence-by-sentence analysis, with plagiarism and AI-generated content scores."

AI-Generated Content 37.4% Confidence

"Color-coded highlights indicate suspicious sections, while confidence metrics provide additional insight for each sentence."

AI-Generated Content 37.4% Confidence

"Figure A.5 – Detailed Report Generation This screenshot presents the PDF report generated by AcadCheck."

AI-Generated Content 37.4% Confidence

"It includes global metrics, sentence-level scores, and highlighted passages."

AI-Generated Content 37.4% Confidence

"Users can download or share the report directly from the dashboard."

AI-Generated Content 37.4% Confidence

"Figure A.6 – Analysis History and Filtering This figure shows the historical analysis interface, listing previous uploads with dates, document names, and scores."

AI-Generated Content 37.4% Confidence

"Users can filter, search, and re-open previous reports for reference or export."

AcadCheck Academic Integrity Platform

This report was generated using advanced AI-powered plagiarism and content detection technology.

For questions about this analysis, please contact your academic institution.

AcadCheck Academic Integrity Platform