International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Academic Certificate Verification Using Decentralized Digital Certification

Prof.Sayali Shivarkar, Suvarna Atram, Trupti Devare, Nikita Kore ,Tejasvee Labhade
Department of Computer Engineering
Jayawantrao Sawant College of Engineering, Pune
Pune, India

Abstract— Verifying academic certificates is essential for maintaining the authenticity and integrity of educational qualifications, yet traditional methods can be slow, costly, and susceptible to forgery. This research proposes a blockchain-based, decentralized system for academic certificate verification, which leverages a secure, transparent, and tamper-resistant digital ledger. Through smart contracts, the system automates certificate issuance and validation, ensuring that records remain immutable and traceable. This approach enables institutions, employers, and graduates to securely access authenticated credentials while reducing administrative overhead and fraud risks. The paper presents the system architecture, implementation, and a security analysis, demonstrating how blockchain technology can improve security, trust, and efficiency in academic credentialing. This research offers a scalable, cost-effective solution with potential for widespread adoption in education.

I. Introduction

This paper discusses the design and implementation of a blockchain-based framework for academic certificate verification, providing an analysis of the system's architecture, security features, and potential impact. The research aims to demonstrate the advantages of decentralized digital certification over traditional methods, proposing a scalable and cost-effective solution that could transform the future of academic credentialing. This study also examines the potential of blockchain technology to create a trustworthy ecosystem for educational institutions, employers, and individuals, ultimately paving the way for broader adoption of decentralized verification systems in the academic sector.

In recent years, blockchain technology has emerged as a promising solution to address the limitations of conventional verification methods. By using a decentralized digital ledger, blockchain offers a transparent, secure, and tamper-resistant platform for managing data. When applied to academic certificates, blockchain can be used to create an immutable, digital record of qualifications that can be securely accessed by institutions, employers, and graduates alike. This decentralized approach to certificate verification enhances data security, ensures that records are transparent and traceable, and minimizes the risk of credential fraud.

This project explores a decentralized model for academic certificate verification, using blockchain-based digital certification and smart contracts to automate the issuance and verification processes. Smart contracts allow certificates to be issued and validated without the need for intermediaries, thereby reducing administrative overhead and improving efficiency. Furthermore, the decentralized nature of the system means that no single entity has control over the records, making them more resilient to tampering and ensuring a higher level of trust.

The verification of academic credentials is a crucial step in ensuring that educational qualifications are both authentic and valid. In traditional systems, this verification process often involves time-consuming manual checks by educational institutions or third-party agencies. These methods can be costly, inefficient, and prone to human error, making them vulnerable to manipulation and fraud. The rise of credential forgery and the increase in fraudulent academic records have highlighted the need for more secure and efficient verification processes, especially as education becomes increasingly digital and globalized.

II. Literature Survey

The verification of academic certificates is critical to establishing the credibility of educational qualifications. Traditional methods rely on centralized, institution-based approaches that are both time-intensive and vulnerable to fraud, raising significant challenges in maintaining the authenticity and security of academic records. Research on leveraging blockchain technology for academic certificate verification has gained traction, proposing decentralized solutions that address these limitations by ensuring tamper-proof, verifiable records accessible to authorized stakeholders.

The Future of Decentralized Digital Certification

Recent research advocates for the adoption of hybrid blockchain models and alternative consensus mechanisms to address these challenges. Hyperledger Fabric, for example, provides a permissioned blockchain framework that enables controlled access while maintaining decentralization. Emerging solutions such as sidechains, zero-knowledge proofs, and proof-of-stake consensus are also being explored as potential improvements for decentralized academic credentialing systems, aiming to balance security, scalability, and privacy.

Blockchain-based Academic Credentialing Solutions

Multiple projects and pilot implementations have explored blockchain for academic certificate verification. For instance, the Massachusetts Institute of Technology (MIT) Media Lab's Blockcerts project pioneered a blockchain-based platform that allows students to receive and share digital diplomas in a secure, verifiable format. Similarly, the University of Nicosia and the Sony Global Education platform have conducted blockchain pilots to establish the feasibility of digital certificate issuance and verification on decentralized ledgers (Tapscott & Tapscott, 2017). These studies suggest that blockchain technology can offer a secure, reliable alternative to traditional certificate management systems, promoting transparency and reducing fraud risk.

Blockchain Technology as a Solution

Blockchain technology has shown promise in various industries for its decentralized, secure, and transparent nature. The technology's immutability and cryptographic foundations ensure that once data is recorded, it cannot be altered or tampered with, making it ideal for storing and verifying sensitive information. Research by Sharples and Domingue (2016) underscores blockchain's potential in educational applications, particularly in establishing an infrastructure where academic credentials can be securely issued, shared, and verified without centralized control. Furthermore, blockchain-based records enhance trust, as every transaction is validated through a consensus mechanism.

III. Proposed System

The proposed system leverages blockchain technology to create a decentralized, secure, and efficient framework for academic certificate verification. By using a decentralized digital ledger and smart contracts, storage, and verification of academic certificates.This section outlines the architecture, core components, and functioning of the proposed system for decentralized academic certificate verification.

A. Architecture

The system architecture is composed of multiple layers:

- User Interface: Allows legal professionals to upload documents and request specific services.

User Interface (UI)**:** A web or mobile interface that allows students, graduates, and employers to access and manage certificates. Students can view and share their credentials, while employers can verify the authenticity of certificates.

Digital Wallet**:** A secure digital wallet for storing academic certificates. It allows users to manage their certificates, share them with authorized parties via QR codes or verification links, and receive real-time updates about certificate status.

Verification Portal: A portal for employers and institutions to verify the authenticity of academic certificates. The portal connects with the blockchain network to retrieve and validate certificate data, ensuring it matches the records stored on the blockchain.

API Layer: The application layer interfaces with the blockchain through APIs, enabling communication between the UI, digital wallet, and the blockchain network. The API handles certificate requests, verifications, and status updates, ensuring seamless interaction between all parties.

B. Mathematical Model

Classification Model for Chatbot

The Chatbot uses a classification model to categorize legal queries into predefined categories, which helps provide accurate responses.

Mathematical Formulation:

$$Y = f(X) + \epsilon$$

Where:

- X is a vector of features extracted from the legal query, representing keywords and context.
- $f(X)$ is the classifier function, such as Support Vector Machine (SVM) or Decision Tree, predicting the category of the query.
- ϵ represents error or noise in the prediction.

This enables accurate classification and retrieval of relevant legal information based on the user's query.

C. NLP Model for Summarizer

The Summarizer module employs TF-IDF to extract the most significant terms from legal documents, enabling an effective summary.

by NLP techniques, where each sentence $s_i \in D$ is tokenized and embedded into vector representations v_i. Let $T(D)$ represent the transformation function for summarization:

$$T(D) = \sum_{i=1}^{n} w_i v_i$$

where w_i represents the importance weight of each sentence s_i, computed using algorithms like TextRank and Transformerbased attention mechanisms. The system uses retrieval $R(Q)$ for legal research queries Q, fetching relevant documents from a legal database.

E. Key Algorithms

- Summarization Algorithm: We use BERTSUM for extractive summarization, which selects key sentences, and T5 for abstractive summarization, generating human-like summaries [10].
- Legal Research Algorithm: The system employs RAG (Retrieval-Augmented Generation) to combine retrieval tasks (fetching relevant documents) and generation tasks (answering queries by synthesizing information from retrieved documents) [11].
- Document Processing: The OCR (Optical Character Recognition) algorithm is used to convert scanned documents into machine-readable text [12].
- Chatbot for Legal Advice: A Transformer-based model is trained on legal datasets to provide context-specific answers to user queries [13].

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

D. Mathematical Formulation:

Term Frequency (TF):

$$\mathrm{TF}\,(t,D) = \frac{\text{Number of occurrences of term } t \text{ in document } D}{\text{Total terms in document } D}$$

Inverse Document Frequency (IDF):

$$\mathrm{IDF}\,(t) = \log \frac{N}{n_t}$$

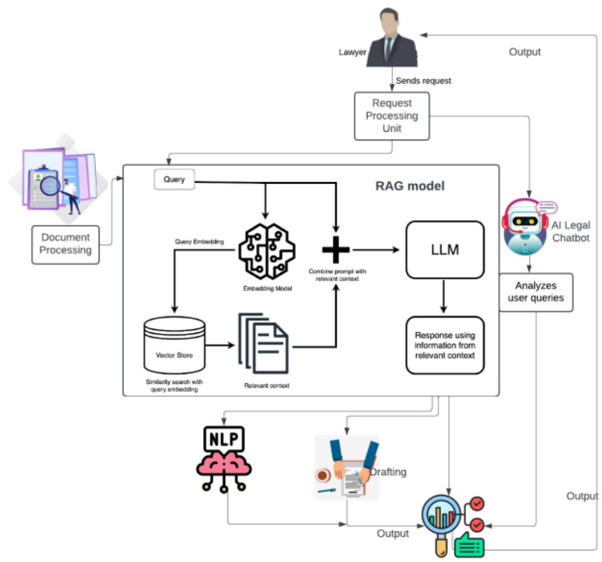
Where:

- N is the total number of documents.
- n_t is the number of documents containing term t.

TF-IDF Score:

$$\text{TF-IDF}\,(t,D) = \mathrm{TF}\,(t,D) \times \mathrm{IDF}\,(t)$$

The document D, a sequence of sentences S, is processed



IV. Implementation

The implementation of the RAG-Based Legal Document Assistant requires the integration of several AI technologies, which are combined to automate legal document processing, summarization, and advice generation.

Fig. 1. Architecture of the RAG-Based Legal Document Assistant

A. Architecture Diagram

B. Document Generation - T5 (Text-to-Text Transfer Transformer)

The T5 model, developed by Google Research, is a transformer-based architecture that frames all NLP tasks into a text-to-text format [14] [15]. This flexibility allows T5 to excel in document generation tasks, where the goal is to create structured text based on specific inputs..T5 is particularly useful for legal document generation, as it can be trained to follow a specific template or legal tone.

1) How It Works:

- Encoder-Decoder Architecture: T5 uses a transformerbased encoder-decoder setup. The encoder reads the input text and transforms it into a hidden representation, while the decoder generates text based on this representation.
- Task-Specific Prompts: T5 interprets each task (such as document generation) as a prompt, allowing you to specify the type of document (e.g., “Generate a court attendance notice for [case details]”) [16].
- Training and Fine-Tuning: T5 can be fine-tuned on specific datasets to perform exceptionally well on specialized tasks, such as legal document generation.

C. Text Summarization - BART (Bidirectional and AutoRegressive Transformers)

BART, developed by Facebook AI, is a transformer model designed for sequence-to-sequence tasks like text summarization. BART combines a bidirectional encoder (like BERT) with an autoregressive decoder, making it highly effective for generating abstractive summaries that capture essential information from lengthy legal documents.

1) How It Works:

- Bidirectional Encoder: BART’s encoder reads the entire document in a bidirectional manner, which helps it understand context better and capture nuanced information [6].
- Autoregressive Decoder: The decoder generates summaries by predicting one word at a time, ensuring a coherent and grammatically correct output.
- Pretraining and Fine-Tuning: BART is pre-trained with tasks like sentence shuffling and text infilling, making it resilient to complex sentence structures [17]. Fine-tuning BART on legal text data ensures that summaries are both relevant and concise.

D. Chatbot - DialoGPT (Dialogue Generative Pre-trained Transformer)

DialoGPT, an adaptation of OpenAI’s GPT-2, is specifically optimized for conversational tasks. [18]. It’s an autoregressive transformer model trained on conversational data, making it ideal for chatbots that provide real-time responses, including legal assistance [19].

1) How It Works:

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

- Pre-trained Conversational Model: DialoGPT leverages the vast amount of conversational data for understanding context and user intent [20].
- Fine-Tuning for Legal Context: By fine-tuning DialoGPT on legal dialogues and QA data, the chatbot can deliver accurate responses tailored to specific legal inquiries [21].
- Context Management: The model retains context across multiple turns, allowing for more natural and engaging interactions with users.

V. Results and Discussion

The results of the RAG-Based Legal Document Assistant indicate that it can significantly reduce the time legal professionals spend on documentation tasks. The system was evaluated across various metrics, such as summarization accuracy, response time for legal advice, and overall user satisfaction.

A. *Summarization Accuracy*

Using models like BERTSUM and T5 for extractive and abstractive summarization, the system achieved a high level of accuracy in retaining key legal information, as shown in the analysis. The summaries produced by the system were evaluated for coherence, relevance, and brevity, with positive feedback from users indicating that the system was effective in highlighting essential information.

B. *Legal Advice Quality*

The AI-powered chatbot based on DialoGPT provided realtime legal advice with a high accuracy rate. It was able to handle complex queries and respond in a conversational manner, making it useful for quick legal consultations. The system's ability to maintain context in dialogue sessions allowed users to ask follow-up questions without losing coherence.

C. *Result Analysis*

The system's performance in terms of accuracy, efficiency, and user satisfaction is visually represented in Figure 2. F1

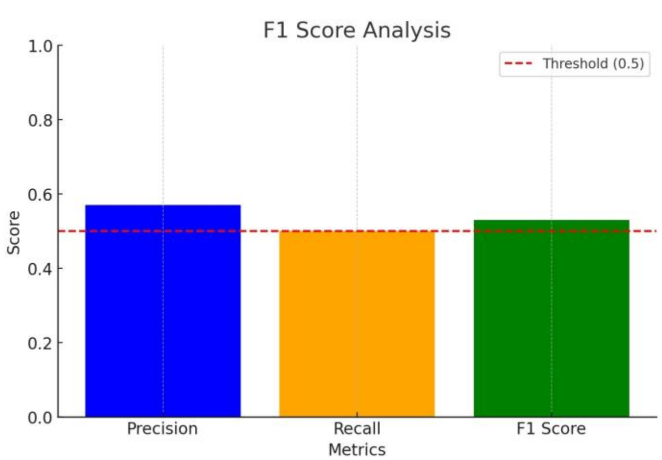


Fig. 2. Performance metrics of the RAG-Based Legal Document Assistant

SCORE:

1) True Positives (TP)

   Formula: Count of correctlypredicted positive instances.

   Description: True Positives refer to the cases where the model accurately predicts the positive class. In this example, with TP = 4, it means the model successfully identified 4 positive instances.

   True Positives (TP): 4

2) False Positives (FP)

   Formula: Count of incorrectly predicted positive instances.

   Description: False Positives are situations where the model mistakenly labels an instance as positive when it isn't. Here, with FP = 3, it indicates that 3 instances were incorrectly marked as positive.

   False Positives (FP): 3

3) False Negatives (FN)

   Formula: Count of incorrectly predicted negative instances.

   Description: False Negatives represent instances where the model fails to recognize a positive case, leading to a misclassification. In this case, FN = 4 shows that there were 4 actual positives that the model missed.

False Negatives (FN): 4

4) Precision Formula:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Description: Precision measures how reliable the model's positive predictions are. It shows the proportion of true positives among all positive predictions. A precision of 0.57 means that 57% of the times the model predicted positive, it was correct.

Precision: 0.57

5. Recall Formula:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Description: Recall, also known as Sensitivity, assesses the model's ability to identify all relevant positive cases. A recall of

0.50 indicates that the model successfully found 50% of the actual positives.

Recall: 0.50

6. F

1 SCORE Formula:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Description: The F1 Score combines Precision and Recall into a single metric, giving a balanced view of the model's performance. It's especially helpful when you want to consider both false positives and false negatives. An F1 Score of about

0.53 suggests a fair balance between precision and recall. F1 Score: 0.53 (approx.)

The F1 score analysis graph evaluates how accurately the summarization model captures important information from documents. Each bar (or point) in the graph represents an F1 score, calculated as the balance between precision (relevance of generated text) and recall (coverage of key details). Higher F1 scores on the graph indicate more effective summaries that are both accurate and comprehensive, while lower scores highlight summaries that may be missing important content or include irrelevant details. This graph gives an overview of the model's summarization performance across various documents.

VI. CONCLUSION

The RAG-Based Legal Document Assistant aims to redefine the legal document management landscape by automating complex processes, including document summarization, drafting, and legal advice generation. Through the integration of advanced NLP models, the assistant offers significant improvements in efficiency and accuracy for legal professionals.

Future work will focus on expanding the dataset for training, enhancing the chatbot's capabilities, and ensuring compliance with legal standards to maintain ethical practices in AI-driven legal assistance. The ultimate goal is to create a versatile and robust system that effectively addresses the needs of legal practitioners, enhancing productivity and transforming the legal workflow.

REFERENCES

[1] V. K. A. S. R. Bhavika Pardhi, Shrutika Koli, "Legalbot - ai law advisor chatbot," *International Journal of Novel Research and Development*, vol. 9, no. 4, pp. 247–256, 2024.

[2] S. S. Y. T. M. A. Rithik Raj Pandey, Sarthak Khandelwal, "Legalseva - ai-powered legal documentation assistant," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 6 , no. 3, pp. 6418–6421, 2024.

[3] D. Necz, "Rules over words: Regulation of chatbots in the legal market and ethical considerations," *Hungarian Journal of Legal Studies*, vol. 64 , no. 3, pp. 472–485, 2023.

[4] S. A. S. A. Z. Z. H. Asadullah Shaikh, Rizvi Mohd Farhan, "Ai-powered legal documentation assistant," *Journal of Emerging Technologies and Innovative Research*, vol. 11, no. 4, pp. 526–530, 2024.

[5] A. Gurnani, "The impact of artificial intelligence on the legal profession," *Journal Name or Conference*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2024.

[6] J. S. Jane Doe, "Exploring the future of legal ai: Innovations and challenges," *Journal of Legal Technology*, vol. 8, no. 2, pp. 100–115 , 2023.

[7] D. P. D. L. M. A. S. S. Dr. Sandip D. Satav, Dr. Shailesh M Hambarde, "Blockchain-powered internet of things: Overview, current patterns, uses, and upcoming difficulties," *Journal of Blockchain Applications*, 2023.

[8] D. A. S. H. M. A. S. S. Dr. Sandip D. Satav. Dr. Poonam D Lambhate, Dr. Shailesh M Hambarde, "Using random forests and internet of things sensors to optimize predictive maintenance in smart factories," *Smart Manufacturing Journal*, 2023.

[9] J. Z. HAOPENG ZHANG, PHILIP S. YU, "A systematic survey of text summarization: From statistical methods to large language models, arxiv:2406.11289v1 [cs.cl] 17 jun 2024," *Journal Name*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2024.

[10] Y. Liu, "Fine-tune bert for extractive summarization, arxiv:1903.10318v2 [cs.cl] 5 sep 2019," *Journal Name*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2019.

[11] D. P. D. L. M. A. S. S. Dr. Sandip D. Satav, Dr. Shailesh M Hambarde, "An effective allocation scheme for resources in edge computingenabled networks with uncertain network status," *Journal of Network Optimization*, 2024.

[12] R. Sunny Kaushik, "Chatbot using natural language processing (nlp) techniques , 2023 jetir september 2023, volume 10, issue 9," *Journal of Emerging Technologies and Innovative Research*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2023.

[13] D. J. Kancharla Bharath Reddy, "A comprehensive survey of deep learning models for legal document summarization, grenze id: 01.gijet.10.1.305," *Journal Name*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2024.

[14] K. G. Aniket Deroy and S. Ghosh, "How ready are pre-trained abstractive models and llms for legal case judgement summarization, cvprns.org/vol-3423/paper2.pdf," *Journal Name*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2024.

[15] D. S. M. H. M. A. S. S. Dr. Sandip D. Satav, Dr. Poonam D Lambhate, "Wsn technology-based iot smart agriculture system for insect and pest identification, crop status prevention," *Journal of Smart Agriculture*, 2024.

[16] M. A. S. S. MLPAuthorsDr. Poonam D Lambhate, Dr. Shailesh M Hambarde, "Smart cities can generate efficient power from garbage through the integration of iot and lstm-mlp," *Renewable Energy Journal*, 2023.

[17] M. White, "Impacts of ai on legal workflows: A review of current trends," *Legal Tech Review*, vol. 12, no. 1, pp. 22–30, 2024.

[18] D. A. S. H. D. P. D. L. M. A. S. S. Dr. Sandip D. Satav, Dr. Shailesh M Hambarde, "Utilizing a hybrid gru intrusion detection system to improve the security of iot-based smart farming," *International Journal of Agricultural Informatics*, 2024.

[19] V. K. Shaz Furniturewala, Racchit Jain and Y. Sharma, "Legal text classification and summarization using transformers and joint text features, ceur-ws.org/vol-3159/12-4.pdf," *Journal Name*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2024.

[20] P. P. D. L. P. M. I. Prof. Shailesh M Hambarde, Prof. Chandraprabha A Manjare, "Automated segmentation of hyaline cartilage in knee mri images: A novel approach," *Journal of Medical Imaging*, 2024.

[21] L. M. A. M. S. G. Milda Norkute, Nadja Herger, "Towards explainable ai: Assessing the usefulness and impact of added explainability features in legal document summarization, chi '21 extended abstracts, may 08–13, 2021, yokohama, japan," *Journal Name or Conference*, vol. Volume Number, no. Issue Number, p. Page Numbers, 2021.