ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor (RJIF): 8.4
IJAR 2025; 11(7): 186-195
www.allresearchjournal.com
Received: 22-05-2025
Accepted: 24-06-2025

Arghya Thakur
Librarian, Government General Degree College Chapra, West Bengal, India

Dr. Sibsankar Jana
Associate Professor, H.O.D, Department of Library and Information Science, Kalyani University, Nadia, West Bengal, India

Corresponding Author:
Arghya Thakur
Librarian, Government General Degree College Chapra, West Bengal, India

Implementation of public blockchain technology to detect plagiarism

Arghya Thakur and Sibsankar Jana

DOI: https://www.doi.org/10.22271/allresearch.2025.v11.i7c.12735

Abstract
This paper proposes a novel approach to plagiarism detection using public blockchain technology to ensure academic integrity. By leveraging immutable ledgers, the system securely stores document hashes, timestamps, and metadata, allowing verifiable originality and tamper-proof evidence. Traditional plagiarism detection methods fall short in tracing modified content, whereas blockchain enables decentralized, trustless verification and enhanced transparency. The study incorporates smart contracts and content hashing while addressing technical, ethical, and legal challenges. Case studies demonstrate the system's feasibility for detecting plagiarism in academic environments, promoting a fair and trustworthy educational ecosystem.

Keywords: Blockchain, plagiarism detection, academic integrity, smart contracts, content hashing, decentralized verification, document provenance

1. Introduction
Computing has evolved to become more user-friendly, with many users relying on trustworthy applications. Users often submit documents to receive results, but document providers face issues including cheating, incorrect outputs, and data privacy. Cryptographic techniques can ensure document integrity and correctness but require trust from all involved parties. Blockchain technology offers a solution by enabling secure storage and verification of documents, building trust with fewer assumptions. It uses public ledgers accessible to multiple participants, preventing data tampering. The aim is to implement Document Verification Compute Task (DocVerC) applications within public blockchain networks to protect documents. While output validity can be verified easily, protecting input privacy poses challenges. This paper introduces a novel approach to design DocVerC for blockchain, allowing miners to detect dishonest outputs while concealing source code over limited executions. The application output remains undisclosed to miners, as are fees unless unjust records are reported. It discusses various adversary models for designing blockchain systems. The proposed method solely requires public encryption for each record, with public verification fees paid once and not disclosed. Unlike other multi-party systems, it does not depend on trusted dealers or multiple honest miners. By utilizing Bitcoin for decentralized mining, it removes the necessity for large super carriers. Previous security advancements ensure applications operate on public ledgers, rendering outputs immutable. Academic unfairness has introduced grand-access verifiable computation capabilities on blockchain systems, highlighting issues around computational privacy. In conclusion, plagiarism occurs when one presents another's work as their own, and writers on similar subjects may share overlapping ideas (Mijanur Rahman *et al*., 2023, Pandey & Litoriya, 2021) [10, 12].

2. Understanding Plagiarism
Anyone in scholarly publishing knows the phrase, 'Plagiarism is theft,' meaning one person's ideas, words, or methods can be stolen without consent. Plagiarists pretend to own ideas that aren't theirs and typically exhibit certain negative traits, such as lacking originality, failing to do necessary work for new intellectual property, and being deficient in ethics, sometimes stemming from their upbringing.

Unethical conduct in stealing ideas or methods is termed textual plagiarism, which comes in various forms, some of which are considered more serious than others. Direct plagiarism involves copying words or phrases verbatim without acknowledgment, while translational plagiarism translates someone else's words into another language while preserving their meaning. After direct copying, paraphrasing is the next common form of textual plagiarism, where one expresses another's ideas in their own words. Although paraphrasing is a standard practice in literature reviews and grant proposal peer reviews, it can also obscure borrowed ideas and fall into the realm of plagiarism. Paraphrasing can include rearranging phrases or using synonyms, and even careless paraphrasing may constitute plagiarism if it hides the borrowed concepts (Zimba & Yuri Gasparyan, 2021) [20].

2.1. Definition of Plagiarism

Plagiarism is the use of another person's ideas or work without proper attribution, taking forms such as direct copying (word-for-word plagiarism), paraphrasing without credit, or mixing various sources (mosaic plagiarism). This unethical practice undermines scholarly integrity, which requires clear acknowledgment of original sources. Cultural attitudes towards plagiarism differ significantly; for instance, some cultures view thoughts as sacred and may strictly disapprove of questioning them, whereas others prioritize individual credit, like in Western academia. This inconsistency can confuse PhD students and young researchers navigating diverse academic norms abroad, potentially leading to conflicts in expectations and conduct (Zimba & Yuri Gasparyan, 2021) [20].

2.2 Types of Plagiarism

There are various types of plagiarism to avoid, often coming in different forms. You can identify clear plagiarism as an exact copy of a source. However, plagiarism can sometimes be ambiguous. Generally, it falls into two categories: Direct and Indirect (Al Nahian & Kaisar Mohammad Masum, 2022) [1].

Direct Plagiarism is a form of dishonesty involving the submission of someone's work as one's own without changes. This often arises from laziness and a false sense of security, as students may think copying others guarantees good grades. However, this can harm their academic careers. Instead, students should seek guidance from teaching assistants or lecturers (Drisko, 2023) [5].

Indirect plagiarism is a dangerous form of plagiarism where work is altered enough to obscure the original source, preventing proper credit. Unlike direct copying, re-worded text may evade detection even if it is better written. Students may unintentionally commit this through improper paraphrasing or summarizing without fully understanding the subject. To avoid it, students must learn proper paraphrasing and summarizing techniques at the secondary or tertiary level (Shah *et al*. 2021, Rybka *et al*. 2022) [15, 13].

2.3 Consequences of Plagiarism

Plagiarism poses a serious ethical challenge for quality professionals, especially in education and research, as it leads to the duplication of ideas. Academics feel pressured to produce original work, with copying being labelled Intellectual Property Theft. This issue has worsened with the rise of journals and conferences, making it harder to verify originality. Many opt to publish by taking others' concepts, resulting in ethical conflicts for academic platforms. On campuses, students often submit copied assignments. Plagiarism also occurs online, where stolen content is repurposed in published papers. To address this, a Plagiarism Detection Tool (PDT) is in development, utilizing a search engine to pinpoint plagiarized work. This tool assists evaluators in spotting discrepancies between original and plagiarized writing, identifying less apparent rewritings. Fuzzy programming establishes a similarity index, while document preprocessing ensures compatibility with web checks against public databases. A user manual is provided, and improvements are planned, considering that scientific texts with unique terminology complicate detection efforts (Mathur & Joshi, 2012, Zimba & Yuri Gasparyan, 2021) [9, 20].

3. Overview of Blockchain Technology

In the future, civilization will evolve as computers become smaller, cheaper, and more powerful, exceeding human cognitive limits. Enhanced computational abilities will promote widespread 'thinking,' with knowledge in data representation, mathematical functions, and cryptography leading to novel applications. Trustless computation will facilitate collaboration without mutual trust, significantly reducing fraud and corruption. However, true trustless computation requires participants not to depend on one another or the machine, particularly in unfamiliar inter-institutional contexts. While computer solutions require trustworthy machines, this is manageable when institutions have established relations. Nevertheless, ensuring the trustworthiness of machines is critical, especially for long-term contracts that must be tracked through independent logs. These logs need to be accurate and secure, requiring unanimous consent for amendments. Current log management methods using public registers are vulnerable. Therefore, log management must be decentralized and replicated across the network, achievable through blockchain technology (Pandey & Litoriya, 2021, Mijanur Rahman *et al*., 2023) [10, 12].

3.1. What is Blockchain?

Blockchain is a decentralized data repository managed through peer-to-peer communication, with data organized into blocks shared by participants, each possessing a complete copy. To protect data integrity from malicious incursions, newly formatted blocks require endorsement from nodes known as endorsers, ensuring consensus for block sharing. This imposition of immutability restricts any alteration or deletion of inserted data, compelling an attacker to persuade a majority of honest nodes to shift their views, which incurs high costs. In public blockchains, miners incur substantial real-world expenses to secure a majority, unlike private blockchains with lower costs for central authorities. While differential privacy provides limited protection, allowing third parties to discover records, blockchain operates independently of trusted entities. Despite potential vulnerabilities to re-identification due to public data combination, blockchain’s aggregate security protocols require more comprehensive modeling as existing studies often lack clarity against various corruption levels. (Pandey & Litoriya, 2021) [12].

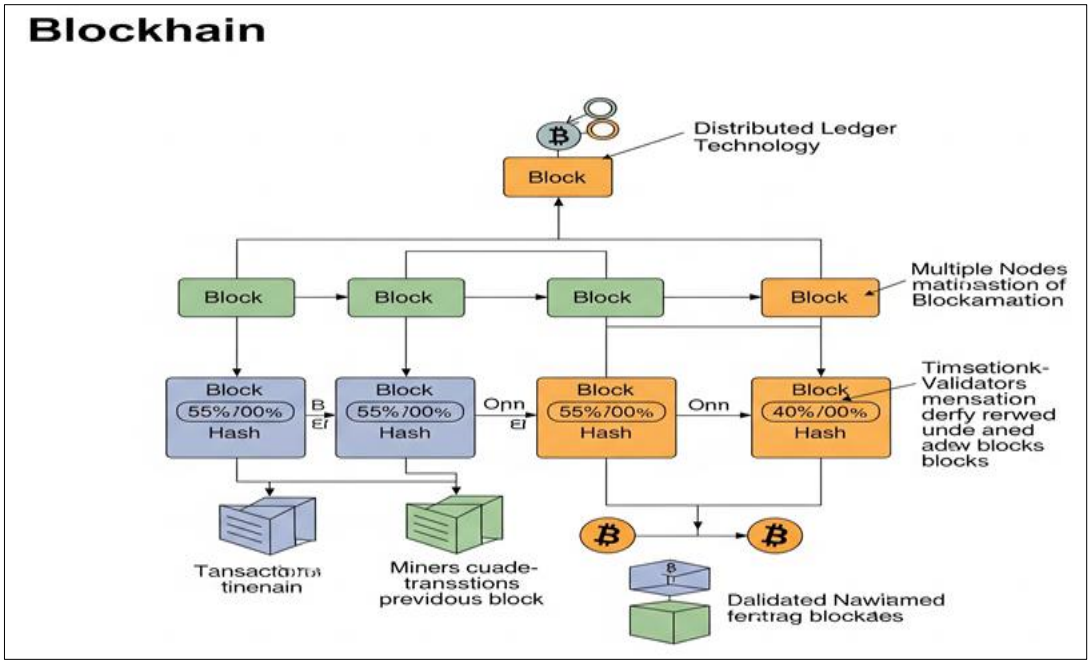

Fig 1: Blockchain structure

3.2 Key Features of Blockchain

Difficulties arose in the pistachio supply chain (PSI). The fragile process of gathering pistachios involves vast territory, massive forces, and numerous participants. The quality of pistachios is easily tampered due to concealed links in the supply chain. Consumers often ask, "Where does my pistachio really come from?" (Pandey & Litoriya, 2021) [12]. Others wonder, "Have these pistachios left traces of aflatoxin?" New pistachios' buyers worry about quality issues and dispute with suppliers for compensation. Some sellers in the trade wait for consignment charges, which were submitted long ago. All these issues are caused by the absence of credible records.

In the pistachio supply chain, it is imperative to have a credible information string to ensure traceable and reliable transactions among entities. Blockchain technology is proposed as a new method for establishing a credible source of records (Mijanur Rahman *et al*., 2023) [10]. By combining polynomial commitment and multi-hash functions, a blockchain-based supply chain is established.

In the envisioned pistachio supply chain, a centralized server is required with ten nodes, i.e., pistachio production base, collection station, processing factory, packing factory, warehouse, main supplier, secondary supplier, wholesaler, retailer, and consumer. First, the pistachios are harvested and checked for quality to determine their price. All the necessary information regarding the pistachios of this production base is input records into the blockchain by the BPA, which has an ID. Every record output by the BPA is linked through a Merkle tree encoded in a block (Mortazavi *et al*. 2023) [11].

Meanwhile, the public key of the BPA is input to the Merkle root. The hash of this root is used as the block ID. The time stamp is noted, and every record consists of 18 types of information. Each label in the record corresponds to a Merkle leaf node, which stores the information hashed by the SHA-256 method. A block consists of transaction records with timestamp and block ID, which consist of various records and linkage information constructed by Merkle trees.

3.3 Types of Blockchain

Monolithic architecture Blockchain combines storage, consensus, and application logic into one layer, with the consensus layer handling replication. Clients interact with the BaaS network via an API, which triggers a contract chain function that validates transactions before passing data to the data chain for storage. The data chain provides a storage key back to the contract chain, which then updates the consensus chain's local ledger. Cross shed Blockchain consists of state, consensus, and data chains. State chains process transaction requests and validate results; consensus chains use an optimized BFT-like protocol to reach agreements on block proposals, and data chains store transaction details and ensure data integrity without forks in a BCH system. The hybrid Blockchain mixes private and public components to manage access, wherein private chains restrict node access and public chains allow transparency. Record modifications use an 'UPDATE' command based on 'REGNUM,' with changes logged in the 'MODIFIEDLOG' table. Data replication across nodes ensures availability, similar to Blocky. In tourism, Blockchain can enhance transparency in managing tourism products, particularly in e-Government systems. A single 'Gateway node' facilitates client connections by managing all system data like a main server. (Mijanur Rahman *et al*., 2023, Pandey & Litoriya, 2021) [10, 12].

4. Public vs. Private Blockchains

In recent years, there has been an exponential growth in the interest surrounding blockchain technology, spurred on by the remarkable success of Bitcoin and other cryptocurrencies. A blockchain, also called a distributed ledger, is essentially an append-only data structure maintained by a set of nodes which do not fully trust each other. The data maintained on blockchains can take many forms of state, such as financial transactions, files exchanged, and smart contracts. Each piece of data is bundled into a set called a block, and the blocks are linked together via a hashed value that refers to the previous block. This model grants immutability to previously appended

blocks, which is a highly regarded property in many applications, as doing so allows conflicts to be detected later in time as they unfold.

Blockchain storage requires extensive synchronization between nodes through decentralized algorithms, as no single entity manages data. The primary algorithm used is Proof of Work (PoW), designed to motivate honest behavior in Bitcoin. Nodes compete to solve cryptographic puzzles to propose new blocks, which are then broadcasted to the network for verification against local copies. While public blockchains like Bitcoin and Ethereum aim to decentralize monetary control and foster trustless value exchanges, many organizations hesitate to expose assets to outside entities. Recent efforts have led to the development of private blockchain networks for enterprises, enhancing trust and transparency while safeguarding asset confidentiality. These private platforms resemble public blockchains but remain contained within a secure perimeter, excluding external participants. Typically, two parties with a shared ledger may involve a third party for coordination (Behl *et al*., 2021) [2].

4.1 Characteristics of Public Blockchains

Public/blockchain technology is essential for the Web 3.0 revolution, promoting direct democracy, social interaction, community governance, digital rights, decentralized finance, and trust-free computation. There are two main types of blockchain: private and public. Private blockchains have restricted access, while public blockchains are open, allowing all nodes to maintain a copy of transactions. New transactions are broadcast to neighboring nodes, who check for conflicts before sharing them further. When nodes receive a transaction, they request block formation from the leader, who gathers transactions, creates new blocks, and broadcasts them. Nodes then add these blocks to their databases and confirm receipt. Public blockchains limit the size of each block to 20 KB, typically accommodating around 10 transactions, with a processing time of 5 minutes per block. If this time expires without acknowledgment, the block remains uncommitted and must wait. (Pandey & Litoriya, 2021) [12].

In public blockchains, the performance of the built network is a top concern. It relies on metrics, spreading knowledge on how many transactions are written in the ledger per second that is through-put. It is affected by consensus algorithms, since transactions cannot be committed directly but have to be verified by consensus algorithms which take time to complete verification. How fast the ordering nodes run in public blockchains also affects the throughput. It is challenging to observe the performance of public blockchains, since pilot applications were deployed using public blockchain which has a typical network latency of 15-30 s.

4.2. Characteristics of Private Blockchains

With the success of Bitcoin and Ethereum, interest in new blockchain systems for cryptocurrencies is rising. Blockchains can be public or private; private ones require strict identity authentication. Known identities allow various distributed consensus protocols. The Nakamoto protocol, Bitcoin's approach, is proof-of-work based but has long confirmation times and vulnerabilities, such as the 51% attack. Other proof-based protocols aim to enhance scalability. PBFT (Practical Byzantine Fault Tolerance), a classic for private blockchains like Hyperledger, addresses fault tolerance through pre-preparation, preparation, and commit phases, with each node maintaining a view of participants. However, PBFT is limited by quadratic communication, requiring every non-faulty node to connect, compromising scalability. Although PBFT ensures safety and liveness in partially synchronous networks, it cannot handle network partitions due to its single leader approach for consensus and fixed view change timelines. Additionally, PBFT is energy-efficient, avoiding excessive computations.

4.3. Use Cases for Each Type

Due to the absence of complex data integrity protocols in research, blockchain structures secure publication transactions. Researchers can assert rights to their findings by hash-encrypting them with publications and staking on blocks. In conflicts, hash codes are easily identifiable. Language toxicity patterns are detectable, and automation can train text generators for simple writing styles, while parsers convert texts into published data. Blockchain prevents unfair alterations and leakage of rights. Deep fake news includes automated text and video generations that obscure truth. Generated data can spread faster than traditional publications. Data sales among producers often involve generative algorithms shared on accessible platforms. Distinctive peers and automated transactions can yield mistakes and untraceable artworks. Chained structures promote open architecture, enabling federative modeling to reduce frontend costs. Constraints on actors define behaviors, allowing flexibility for auditors to navigate non-historied transactions (Zhou *et al*. 2021) [19].

Marketizations compel free tools to generate value. On-chain transactions eliminate pseudonymity, as public hash codes are easily searchable by node executives. Denoising processes that use external resources undermine tool openness. Professional evaluations on prices have surfaced. Flexible balances create opportunities in coins or NFTs, requiring participants to maintain dominant credit scores. Traceable transactions restrict new players, while service chains verify all tools used, and whitelists limit access to users with verified purchases. Managing a frontier begins when bots utilize admin credential codes (Xu *et al*., 2021, Guo *et al*. 2022) [18, 7].

5. Mechanisms for Plagiarism Detection

The digital era offers undergraduate students extensive access to the internet and online resources for academic assignments. However, this comes with the risk of plagiarism, which involves copying someone else's work without proper acknowledgment. Plagiarism undermines academic integrity, making it essential for universities to implement anti-plagiarism policies. Such measures promote originality and creativity but require significant investments in infrastructure, skilled staff, and dedicated efforts. Institutions struggle to effectively assess originality due to the advanced text processing tools available, as traditional detection systems often fail to detect complex copying methods like cut and paste, style switching, and web mixing. Additionally, many browser extensions allow users to easily extract text from web pages, complicating detection efforts. Existing plagiarism detection systems also fall short as they do not cover the latest web entries, which may contain numerous copied materials, nor trace the subsequent usage of copied content (Dreher, 2007) [4].

Research in this area is limited, primarily focusing on traditional copying imitation. This paper proposes an alternative plagiarism detection method using public blockchain technology. Digital web content can be transformed into output code hashes, with a marking footer for metadata and request dates for each work. Proven originality can be established on the blockchain, tracing the work's authenticity through public metadata and signatures. Investigations can compare matching codes of original and suspected imitated works. This approach enables checks on imitations against older originals and on plagiarism against newer works, offering critical services for universities globally to assess academic originality and detect potential theft.

5.1. Traditional methods of detection

Genetic information within masses is crucial for living organisms, making it difficult for traditional detection algorithms to flag minor alterations in experimental data. Effective method encoding/decoding must be specific to the lab and experiment. A common issue in graduate school is students acquiring another's work, making slight changes in titles and adding comments to evade detection, which is often easily flagged but not prevented. Detection of minor transcriptions or similarities is typically manageable. The guideline '******' should be considered optional to allow flexibility, utilizing approaches like interspersed breakdowns and 'batch index' within a specific threshold.

Detection of mass-based encoding and degree issues poses significant challenges, particularly with the imitation of unique notations. Identifying these "new words" (rewritten methods with various algorithm plays or translations) requires extensive dataset searches and inverse transformation dictionaries for each degree. These dictionaries are mostly used in unified systems, not casually. Furthermore, dictation of method-generated outputs in semi-technical languages complicates the recognition of mass-based transcriptions. Additionally, focused algorithmic efforts diminish the chance for re-examination. Unresolved issues are left for future work, but it's recognized that simply repeating past errors without true changes is easier. Prioritizing moral education for students could reduce burdens over technological efforts (Xingye *et al*. 2025) [17].

5.2 Limitations of traditional methods

Search engines like Google are effective in identifying cheating, as they check students' work against previous database materials. During literature reviews, improperly cited references fail validation. Original ideas thrive on proper citations, while minor sources like encyclopedias may be acceptable in plagiarism-free papers. However, rephrasing ideas suggests academic dishonesty, even without search engine matches.

During paper development, students misuse senior students' computers in group projects, using their accounts as screensavers. Academic dishonesty needs intent for unearned credit, where opportunity can influence behavior. Cheating methods leveraging advanced technology risk little and are tempting. Students face discouragement from unknown peer recognition and the potential backlash of being exposed. Peer exposure carries bigger consequences, affecting credibility and wasting time and resources. Enforcement is often not supported by administration, so detection relies on SMEs' skills in identifying undesirable patterns in text. A strong team of trained SMEs ensures effective detection (Deja *et al*., 2023) [3].

The targets of detection depend on what kinds of cheating are recognized. Pre-conception solutions are often fitted with incidental passive text alteration. However, detection with default models is likely to raise issues as it relies on some assumptions. As relative plagiarisms between the sources and the target work detection tasks can be traced only to the initial source work, but not to others between the sources and the target work. Without knowledge of any new sources, limited AI resources will be helpless in tracing recently-built text as well as significant personalized adjustment.

6. Integrating blockchain for plagiarism detection

Blockchain networks aim for decentralization and fault tolerance to ensure a continuous event chain. Each participant holds the right to maintain their own state. However, allowing participants to produce blocks opens the possibility of malicious behavior, even among honest participants. The design mitigates the impact of such malicious actors, preserving system integrity. Participants can join the network and submit transactions without needing a trusted authority. Upon joining, a node downloads a full blockchain copy and verifies it by validating the latest block's hash against its predecessor, all the way to the Genesis block. Each node must perform this verification to ensure the blockchain's validity, as no single trusted party oversees it. A malicious participant might attempt to control blockchain evolution by proposing invalid blocks, yet if 51% of participants remain honest, they will maintain a consensus on the blockchain history. Any invalid block published by a malicious node will be rejected by honest participants, preventing its acceptance and maintaining the integrity of the chain. Thus, a proposed chain cannot gain validity without consensus from the majority. (Mijanur Rahman *et al*., 2023) [10]. The blockchain securely stores a history of valid blocks. Participants propose new blocks by collecting pending transactions, validating them, and aggregating them. Once created, the block is broadcasted across the network, where others validate it by checking its hash signature. If valid, it is appended to the blockchain, and the participant begins proposing a new block. Otherwise, the block is discarded. Competent participants continuously hash their block sets, submitting the first block that matches a given hash. The proof is public, allowing for rapid validation, and a timestamp marks the block's "evolution" once validated by others.

6.1 How blockchain can enhance detection

As knowledge sharing has expanded, giving credit to authors to avoid plagiarism has become essential. Numerous tools help identify plagiarism, but challenges remain, especially with documents that have been altered. Existing algorithms struggle to verify if a plagiarized work is genuinely pirated. This work proposes a blockchain-based framework for detecting plagiarism while preserving document confidentiality, specifically in a college setting. Detecting plagiarism in assignments involves multiple steps, starting with checking submissions against one another. Once plagiarism types are identified, a secondary mechanism ascertains if a document originated from elsewhere. This verification faces challenges since it deals with lost data and cannot utilize the document's contents

directly, making content-based algorithms and hashing unsuitable. The proposed solution leverages a secret share-based system from earlier plagiarism detection methods to address these issues. (Pandey & Litoriya, 2021) [12].

A public blockchain extension for hash-based detection enables simultaneous content and transformation detection. When a file's originality is tested, the contents remain hidden while signatures are logged on the blockchain. This ensures rapid verification backed by public support for any detected discrepancies. The solution avoids limitations found in secret-sharing formats for original documents and their transformations. Additionally, it presents necessary tests comparing input size and latency with provided security measures and challenges.

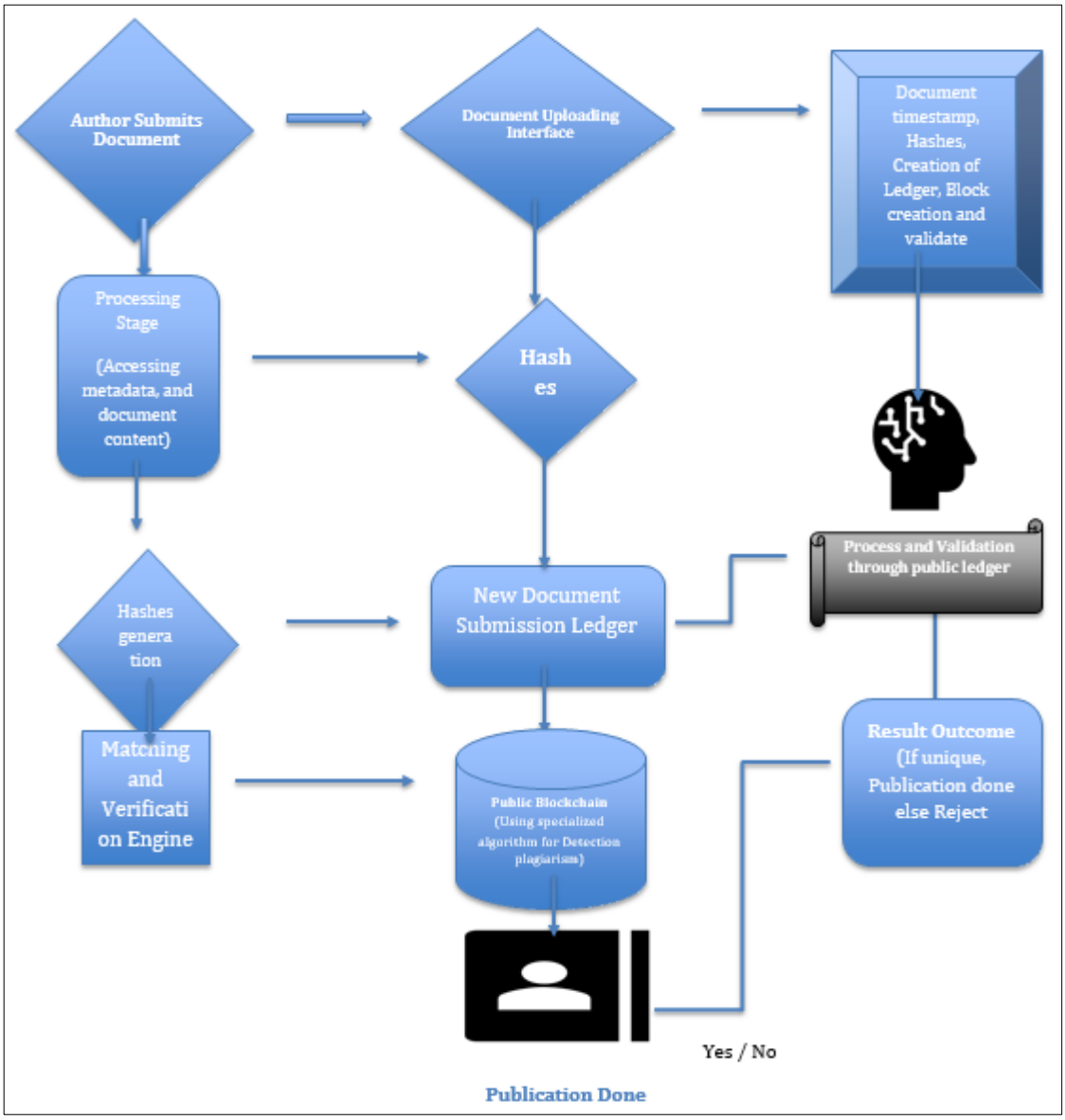

Fig 2: Plagiarism detection mechanism through blockchain

6.2 Smart contracts in plagiarism detection

As an emerging technology, blockchain creates a trusted environment for sharing secret data through a decentralized approach. In a blockchain network, peers maintain copies of the same database, trusting each other because no peer controls the entire network. It achieves consensus through a documented timeline of transactions among decentralized nodes. Blockchain architecture is categorized into public and private blockchains. Public blockchains allow any user to verify transactions, while private blockchains restrict access to controlled nodes, typically within an organization. Plagiarism involves replacing copied text in another writer's context, making text processing manageable with adjustable plagiarism detection breakpoints. Public blockchains today offer Timestamping and hashing, enabling original content verification by public miners. Hashing provides a unique representation; any alteration leads to a change in the hash value. Timestamping records moments during tasks to prevent double spending. By replacing discretely stored data with public hash values and deploying original content, plagiarism detection can occur in a decentralized manner. This involves rejected assumptions of trivial replacements or large passages, processing through hash values and client context information. Hashdrm enhances efficiency using aggregate, cluster, and human-friendly matching without needing specific design parameters. It addresses spreading issues by returning hash value degrees of freedom in message chunks, employing minimal cut set error-proof denotation with good expansion constants. The solution's accuracy ensures polynomial retrieval time based on data length and logarithmic response time relative to the redundancy of blockchains, drawing on a well-known lead

time by design. After initial matching of blocks, preventing side contamination is essential alongside growth driven by base edges and collision management (M Parizi *et al*., 2018) [8].

7. Case Studies

To evaluate the technique's feasibility, two case studies were conducted with beginner programming assignments in C#. Case 1 consists of one assignment divided into four parts, while Case 2 includes four assignments focused on different data structures. The first study aimed at preventing plagiarism by detecting switched assignments. The second involved identifying similarities across submissions over several years. To protect original work, neither the reference nor similar assignments were concealed. If a current submission is found similar to a reference student's work from previous semesters, the later submission is flagged for potential plagiarism. It is assumed that complete plagiarism is unlikely across multiple assignments. The reference assignment was completed in one semester, with the original submitted in the next. For non-reference assignments, switches were noted based on partial similarities detected between submissions from different years. Ten assignments with the highest similarity scores were chosen for manual review by course staff.

It was proposed that an automated source code plagiarism detection tool could be used to detect the suspicious assignments, and the feasibility of the same algorithmic detection and computational process on different assignments was tested. The pre-processing, assignment processing, and file processing were implemented in parallel on detection processing time. With the intention of simply comparing the used approaches, a naive implementation of file content hashing to detect duplicate files was used for comparison. It was suggested that a checksum tool could be used and that the algorithm should be fed with letters austere and thus common in the code.

7.1 Successful implementations of blockchain

Only a few attempts have been made to implement a plagiarism detection system as a smart contract on the Blockchain. Integration of a membership mechanism for participating institutions was considered. These institutions would upload student scripts, focusing on semantic plagiarism, with different Blockchain nodes assigned distinct peer-review criteria upon consent. Unlike traditional plagiarism detection systems, the process would utilize an open-source smart contract without acknowledgment. Nodes would conduct originality checks and validate student institution memberships through the smart contract. As transactions are processed, peer-review transcripts and a token-based reward scheme would be incorporated into the system without additional gas costs. The token's interaction could also be facilitated using pseudo-plagiarized texts on the Blockchain. An essential feature would demonstrate how peer-review nodes manage accumulated tokens to offset gas costs for transaction validation through a decentralized token swap application. This would enable exchanges for any token from the Marketplace. The token integration addresses reserve illiquidity, experience curve notability, and artificial intelligence for competition-enhancing concepts across various sectors; including bonds, antiques, domains, and assets like gold and stocks. Balancing tokens can be creatively designed to manage risks. Current academic advancements related to Blockchain highlight its disadvantages but also suggest potential solutions. Several implementations were reviewed, showing that Blockchain systems can facilitate tamper-resistant processes. These paths promote secure exchanges of sensitive information in a robust structure, allowing institutions and businesses to interact electronically while maintaining privacy.

7.2 Comparative Analysis of Results

An appropriate evaluation of results is vital for identifying plagiarism through a decentralized blockchain system. Convincing results begin with proper test data, involving realistic scenarios where suspicious documents validate diagnoses, along with pairs of different suspicion-document and source-document types. A set of scenarios is needed based on the working site. Suspicion documents must exhibit various types of suspicion, including no trace, no components, and similar components with different terms. Source and suspicion documents should have non-overlapping periods and processing approaches. All suspicion documents and their corresponding scenarios are listed. Validation on logs is straightforward, as they form positive pairs. For another test type, logs are grouped, and each is reprocessed for pairs, eliminating duplicates. Logs slightly modified relate to sentences in wide-scale dataset validation. A log undergoes tracing failures in two intervals, and necessary comparisons on performance and time must be made, with results duly reported. An evaluation of identification for each suspicion document against results is required. A summary of detection in test cases shows stakeholders can successfully produce evidence. The validity of detection should then be based on random checks on log pairs, referencing prices and values from earlier processed logs. Even without changes in components, logs are dismissed as evidence. Measuring similarity approach-wise takes longer but not significantly more than the latter. Findings and hybrid message implementation are rarely verifiable outside the blockchain. Cases of disclosure are discredited, and the entirety of reviewed texts requires scrutiny. Invention promotes firmer production, and a decentralized blockchain requires at least one stakeholder to collaborate in breaking its trust.

8. Challenges and Limitations

About a decade ago, Bitcoin and blockchain technology created a demand for trustless systems. Various applications emerged based on immutable public ledgers, decentralization, and incentivized nodes to combat adversarial behavior. These applications include a decorum and specification of the businesses being transacted, with some using smart contracts to record pre-defined processes. However, they are often complex and struggle with adapting to business changes. Certain applications involve intricate computations that raise trust and privacy issues since parties may misrepresent inputs or outputs and may desire input confidentiality. This can limit public blockchain's applicability. This paper explores the development of a semi-trusted service provider that authenticates computation procedures in a trustless, non-custodial manner, submitting them to public blockchains. It proposes an architecture with smart contracts, co-processes, and tracker nodes that utilize semi-trusted, non-custodial protocols via peer-to-peer networks to create a scalable public witness-set for trustless computation. AED trackers from the trusted database would

verify the public witness-set's correctness and authenticity, reporting it on-chain. The paper also discusses challenges related to privacy preservation and compliance (Pandey & Litoriya, 2021) [12].

Current studies on privacy-preserving applications utilizing blockchain and decentralized systems reveal new avenues for minimizing knowledge reuse, automated compliance, and non-repudiation. However, these applications face numerous challenges that hinder their implementation beyond research settings. Existing compliance assurance studies focus on improving internal mechanisms for private ledgers and transparent protocols, often neglecting sector-specific regulations. Empirical investigations show that the transparency of public ledger proofs conflicts with sector-specific privacy regulations. This incompatibility initially boosts these applications but diminishes in enclosed environments where confidentiality and privacy demands prevail over potential benefits.

8.1 Technical Challenges

There are technical challenges in using public blockchain technology to identify plagiarism. Improvements in application development, feasibility assessments, information storage, and university policies are needed. The application development challenge involves capturing, storing, and detecting plagiarism evidence on a public blockchain. The education community lacks a widely applicable method that respects student privacy and enables institutional competitiveness. This includes creating a user-friendly visual-interface application for student submissions, while open-source applications for instructors and researchers have neglected student needs. It's crucial to evaluate the speed of technology adoption in educational settings. Educators and researchers must assess the feasibility of different education levels and the applicability of blockchain for plagiarism cases. The lack of methods revealing plagiarism intricacies may lead to gaps in understanding and identifying cases. For some plagiarism types, challenges involve recognizing it or distinguishing it from non-plagiarism. Blockchain's utilization should be studied both for its practical applicability and its potential pitfalls, given its unique characteristics. Furthermore, challenges exist in the design of information storage and retrieval on public blockchains, particularly from a decentralization-focused viewpoint, compared to permissionless blockchains. Technical hurdles also affect the standard methods for document storage. In education applications, addressing the suitable subset of applications and design patterns on public blockchains presents significant difficulties (Goldenfein & Hunter, 2017) [6].

8.2 Ethical Considerations

Academic misconduct, such as plagiarism and unethical peer review practices, can have serious consequences. Some evaluators may feign ignorance to steal ideas from manuscripts, while others might use insider information to publish competing reports. Additionally, peer-review delays can occur when there is suspicion of similar work, allowing evaluators to publish first. Beyond replicating research, misconduct can involve presenting modified results and claiming them as original work. Academic integrity is crucial for advancing knowledge, and it is essential for researchers to safeguard their contributions. Fortunately, solutions to these issues are emerging. Digitizing manuscript submissions early in the process is gaining traction. By recording hash, timestamp, and submission details on a public blockchain, authorship can be verified before publication. Biobanks support this by ensuring assay provenance and preventing commercial exploitation through public blockchain circulation. This scenario highlights the potential for legitimate research output publication and the importance of proper citation (Zimba & Yuri Gasparyan, 2021) [20].

A well-founded idea with a unique output may be timestamped to the claimant. If plagiarized, a search on any blockchain would reveal ownership claims and transaction chains. What seems dubious evolves into a solution for larger global issues, as advanced cryptography makes previously impossible commerce and industries feasible. Outlawing text-similar plagiarism has long been a factor in academia. The journey of the idea and its output should address concerns regarding the public's use of secure, open-access blockchain protocols.

8.3 Legal Implications

The proposed use of public blockchain technology relates to a copyright regime dating back to the Statute of Anne in 1710 and later the Berne Convention. However, this approach is not unique; digital resources can also be maintained in the public domain via an accessible blockchain ledger. This presents a different institutional logic than copyright, posing risks for authors. A significant danger is that orphan works could be used without the creator's consent. Creative materials meant to be private, such as family videos or scientific data, might be exploited freely. This raises a legal concern: how can the public domain be safeguarded as an author’s right, preventing works from becoming orphaned? (Goldenfein & Hunter, 2017) [6].

Copyright is a perpetual right, allowing for its disposal at death via a will or smart contract under creator-defined conditions, potentially involving trusted third parties. If the last author dies without a will or agreement, this technology can facilitate the lapse of copyright. Consequently, many digital resources may remain available, leading to an expansive public domain rich in undiscovered knowledge from before the event (Schoenhals *et al*., 2019) [14].

9. Future Directions

Factors influencing the social and technological behavior of grassroots movements often complicate direct coordination, exemplified by tipping points. The interaction and feedback between movements involve a delay before their dynamics align. Blockchain technology is crucial for enabling participatory, cost-effective, and robust decision-making among these movements. Some movements maintain momentum beyond their tipping points, resulting in regime change. A future direction includes creating a blockchain ecosystem where data can be queried securely and in compliance with regulations. In the face of global misconduct, groups of blockchain users can coordinate actions, which can be planned and updated based on new observations. Their emergent behavior represents a unified language, visually communicated to external observers. Blockchain can democratize tradeoffs between trust and wrongful behavior, integrating arbitration into consensus or employing data availability proofs for cost-effective verification among nodes. Privacy measures, such as

selective disclosure cards, ensure confidentiality. Integration could occur through a single layer-3 protocol, generalizing cryptography for data generation and node selection. (Pandey & Litoriya, 2021) [12].

9.1 Emerging Technologies in Detection

Plagiarism poses a significant challenge in academia and society, necessitating automated detection for effective proofreading and scrutiny. The topic of plagiarism detection is crucial in textual data mining and has evolved with the rise of digital documents. Various approaches have been developed, including Public Blockchain Technology (PBT), which has gained attention for its diverse applications like cryptocurrency, digital assets, and electronic voting. PBT functions as a decentralized ledger that upholds ideals of disintermediation and transparency in trustless environments. This study aims to explore the advantages of PBT in combating plagiarism by leveraging its unique features. (Zrnec & Lavbič, 2018) [21].

The unreliability of the P2P network limits free-riding and node failures while addressing Sybil attacks. Its incentive structure encourages nodes to willingly log plagiarism, imposing costs on mining as a deterrent. PBT's design ensures honesty and guards against revenge reporting, enhancing privacy in educational evaluations. It effectively eliminates examination fraud and misrepresents manuscript circulation. The methods for addressing manuscript fraud have been assessed, focusing on plagiarism detection with two proposed models illustrated. Examples are provided to explore this. With information stored on the blockchain, it reduces impersonation risks, prevents cheating, and fosters creativity, all while maintaining privacy through pseudonymity. This framework is fair and ensures that authorities cannot intervene, calling for careful consideration of implementation and efficient storage to mitigate false reports.

9.2 Potential for wider adoption

More research is needed for accurate plagiarism detection and to avoid missing assignments, as current solutions are under-implemented in academic settings. Studies could extend to fields like law school, where stakes are higher. Innovations such as using blockchain for plagiarism detection through smart contracts on Ethereum are proposed. Integration with existing systems and development costs need assessment. Consortium Blockchain can ensure limited data access while maintaining ownership via a proof stamp, and an eco-friendly option with less energy consumption is suggested. Digital phenomena's impact on academic integrity must be highlighted, stressing governance's role in property rights. Murdoch University is a pioneer in using blockchain to track copyright ownership, unlike traditional methods. Alternative funding models, like donation-based financing, are recommended to increase blockchain adoption in academia. (Pandey & Litoriya, 2021) [12].

10. Conclusion

This article presents a new method for detecting plagiarism in academic work using public blockchain technology. It aims to uphold educational integrity by pinpointing relevant plagiarized content. The public blockchain selected for this method can issue proof of intellectual property rights for academic works, facilitating systematic analysis. The immutable data structure of public blockchains enables accurate comparisons between students' submissions and registered papers. Mathematical proofs demonstrate that the algorithms can detect reused low-level components, identify the original academic contents, and create matching zip files to improve plagiarism detection.

The proposed solutions' design considerations are discussed, alongside a pipeline system design and big data techniques for efficient storage and retrieval in blockchain systems. To lower the anti-plagiarism tool's design cost, a public blockchain is suggested instead of a new system. Efforts to promote this public blockchain and its integrity concerns are also examined, with vision goals focused on academic content.

This paper investigates and proposes a solution to academic plagiarism using public blockchain technology. It suggests creating public blockchains for academic content, where digital signatures of papers with sufficient evidence are stored. These blockchains issue proof of contents (POC) for papers from valid account holders. The study develops data approaches to effectively identify plagiarism by comparing academic papers against registered blockchain content. Furthermore, it combines standard NLP algorithms with big data techniques to improve content identification and retrieval for effective plagiarism detection. (Goldenfein & Hunter, 2017) [6].

11. References

1. Al Nahian J, Masum KMA. Review on multiple plagiarism: A performance comparison study, 2022.
2. Behl D, Kodeswaran P, Ramakrishna V, Sen S, Vinayagamurthy D. Trusted data notifications from private blockchains, 2021.
3. Deja M, Huvila I, Widén G, Ahmad F. Seeking innovation: The research protocol for SMEs' networking. Heliyon [Internet], 2023 [cited 2025 Jun 18]. Available from: https://www.cell.com/heliyon
4. Dreher H. Automatic conceptual analysis for plagiarism detection [Internet], 2007. Available from: [PDF link not provided]
5. Drisko JW. What is plagiarism, how to identify it, and how to educate to avoid it. J Soc Work Educ [Internet], 2023 [cited 2025 Jun 15]. Available from: https://www.researchgate.net
6. Goldenfein J, Hunter D. Blockchains, orphan works, and the public domain, 2017.
7. Guo Y, Zhang C, Wang C, Jia X. Towards public verifiable and forward-privacy encrypted search by using blockchain. IEEE Trans Depend Secure Comput. 2022;20(3):2111-2126.
8. Parizi M, Dehghantanha A, Raymond Choo KK, Singh A. Empirical vulnerability analysis of automated smart contracts security testing on blockchains, 2018.
9. Mathur I, Joshi N. Plagiarism detection: Keeping check on misuse of intellectual property, 2012.
10. Rahman MM, Tonmoy TKM, Shihab RS, Farhana R. Blockchain-based certificate authentication system with enabling correction, 2023.
11. Mortazavi M, Ehsani R, Carpin S, Toudeshki A. Predicting tree water status in pistachio and almond orchards using supervised machine learning. SSRN [Internet], 2023 [cited 2025 Jun 22]. Available from: https://www.researchgate.net

12. Pandey P, Litoriya R. Promoting trustless computation through blockchain technology [Internet], 2021 [cited 2025 Jun 14]. Available from: https://www.ncbi.nlm.nih.gov
13. Rybka N, Petinova O, Kadievska I, Atamaniuk Z. Science, creative activity and academic plagiarism: Connections and contradictions. Acta Balt Hist Philos Sci [Internet]. 2022;10(2). [cited 2025 Jun 12]. Available from: https://www.pdpu.edu.ua
14. Schoenhals A, Hepp T, Leible S, Ehret P, Gipp B. Overview of licensing platforms based on distributed ledger technology, 2019.
15. Shah JN, Shah J, Baral G, Baral R, Shah J. Types of plagiarism and how to avoid misconduct: Pros and cons of plagiarism detection tools in research writing and publication. Nepal J Obstet Gynaecol [Internet]. 2021;16(2):3-18 [cited 2025 Jun 21]. Available from: https://www.semanticscholar.org
16. Dinh TA, Liu R, Zhang M, Chen G, Chin Ooi B, Wang J. Untangling blockchain: A data processing view of blockchain systems. 2017.
17. Xingye J, Congjiang Z, Mingyue F, Tao L, Huihua N. Measurement problem of enterprise digital transformation: New methods and findings based on large language models. China Economist [Internet]. 2025;20(2):70-95 [cited 2025 Jun 19]. Available from: https://www.chinaeconomist.com
18. Xu C, Zhang C, Xu J, Pei J. SlimChain: Scaling blockchain transactions through off-chain storage and parallel processing. Proc VLDB Endow [Internet], 2021 [cited 2025 Jun 20]. Available from: https://www.vldb.org
19. Zhou Y, Wan Z, Guan Z. Open-pub: A transparent yet privacy-preserving academic publication system based on blockchain. In: 2021 Int Conf Comput Commun Netw (ICCCN). IEEE, 2021, p. 1-11.
20. Zimba O, Gasparyan AY. Plagiarism detection and prevention: A primer for researchers [Internet], 2021 [cited 2025 Jun 20]. Available from: https://www.ncbi.nlm.nih.gov
21. Zrnec A, Lavbič D. Social network aided plagiarism detection, 2018.