AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 1(2026)

Blockchain-Enabled Heuristic-Optimized Deduplication Model for Mitigating Single-Point Failures in Cloud Storage

Rammohan Burra[1], Gundapaneni Lokesh[2], Bandi Deepika[2], Gaddam Abhishek[2], Gudala Raghavi[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science)

[1,2]Vaagdevi College of Engineering (UGC-Autonomous), Bollikunta, Warangal, 506005, Telangana.

ABSTRACT

Cloud storage systems have become an essential component of modern data management, enabling users to store and access data remotely. However, traditional cloud storage architectures rely on centralized servers, which introduce critical challenges such as single-point failure, redundant data storage, high storage costs, and security vulnerabilities. In earlier systems, data was stored in centralized data centers where duplicate files were often saved multiple times, leading to inefficient utilization of storage resources. Although basic deduplication techniques were used, they frequently compromised data confidentiality and lacked transparency in metadata management. Moreover, failure of the central server could result in permanent data loss. To overcome these limitations, this research system integrates blockchain technology, InterPlanetary File System (IPFS), Convergent Encryption (CE), and heuristic-based chunking techniques to create a secure and decentralized storage framework, hereafter named Blockchain-enabled Heuristic Optimized Deduplication Model (BHODM). In this system, files are divided into optimized chunks using a heuristic method based on file size. Each chunk undergoes CE, where the encryption key is derived from the hash of the data itself, enabling secure deduplication without exposing plaintext information. Duplicate chunks are identified using hash comparison, ensuring that only unique data is stored. The encrypted chunks are stored in IPFS, a decentralized peer-to-peer storage network that eliminates reliance on a single server. Metadata such as file names, block numbers, and hash values are securely stored in an Ethereum blockchain smart contract, ensuring immutability and transparency. The system is implemented using Django for the web application, Web3 for blockchain interaction, IPFS Application Program Interface (API) for distributed storage, and Advanced Encryption Standard in Counter Mode (AES-CTR) encryption for security. By combining decentralized storage, blockchain-based metadata management, and secure deduplication, the proposed model effectively reduces storage overhead, enhances data integrity, and mitigates single-point failures. The system is further evaluated using storage utilization and computation time analysis, demonstrating improved efficiency compared to traditional approaches.

Keywords: Blockchain, Cloud Storage, Data Deduplication, IPFS, Convergent Encryption, Decentralized Architecture

1. INTRODUCTION

The exponential growth of digital data generated by individuals and organizations has placed immense pressure on traditional cloud storage systems. Conventional centralized cloud storage architectures, while widely adopted, often suffer from critical issues such as single points of failure, limited transparency, and vulnerability to data breaches [1]. These shortcomings have driven research interest toward decentralized cloud storage solutions, which distribute data across multiple nodes to improve fault tolerance, privacy, and resilience [2]. Blockchain technology has emerged as a promising foundation for decentralized storage due to its

inherent characteristics of immutability, transparency, and decentralized consensus. By eliminating reliance on a central authority, blockchain-based cloud storage systems can enhance data security and trustworthiness [3]. However, the decentralized nature also brings new challenges, such as increased data redundancy, inefficient storage utilization, and potential imbalances in resource allocation across nodes [4].

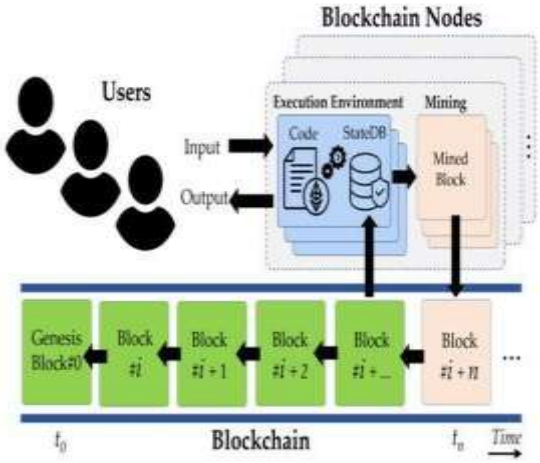

Fig 1. Blockchain tamper-proof, deduplicated, decentralized cloud storage.

One significant challenge in decentralized storage systems is handling duplicate data efficiently. Data deduplication, a technique that eliminates redundant copies of data, is essential for optimizing storage capacity and reducing network overhead [5]. Ensuring reliable and privacy-preserving deduplication in a blockchain environment requires sophisticated cryptographic methods that safeguard data confidentiality while enabling effective redundancy detection. Furthermore, to maintain system performance and reliability, it is crucial to balance storage loads dynamically across distributed nodes. Without proper storage balancing, some nodes may become overloaded, causing bottlenecks and reduced availability, whereas others remain underutilized [6]. Effective load distribution algorithms must consider node capacity, network conditions, and data access patterns. This project aims to address these challenges by designing and implementing a BHODM framework that integrates reliable deduplication and intelligent storage balancing [7]. The proposed system not only enhances storage efficiency and security but also ensures equitable resource utilization, contributing to the scalability and robustness of proposed BHODM system solutions [8].

2. LITERATURE SURVEY

Merlec, et al. [9] proposed a systematic comparative analysis of decentralized storage systems, emphasizing their potential to enhance sustainable DSS. By highlighting the integral role of blockchain technology, this study critically examines various decentralized storage platforms, including Arweave, BitTorrent, Dat Protocol, Filecoin, Hypercore Protocol, IPFS, MaidSafe, Sia, Storj, and Swarm. The analysis covers the key architectural features of these systems, their performance metrics, and their contribution to user data sovereignty. The study aims to comprehensively explain how these decentralized storage solutions allow users to maintain complete control over their data, thus offering a viable alternative to traditional centralized storage methods. Therefore, this paper contributes to ongoing data sovereignty research and guides future developments in decentralized storage technologies.

Meng, et al. [10] proposed a decentralized storage system combining Hyperledger Fabric and Inter Planetary File System (IPFS). In addition, from the perspective of security and availability of the decentralized storage system, we study the partitioning and the k-r allocation scheme of the stored data, propose the allocation function about the stored files, derive the mathematical formula of file security and availability based on the allocation function, and discuss the optimal parameter setting of the allocation function based on the formula to guarantee the high security and availability of the stored files. The experimental results show that the performance of the k-r allocation policy based on the minimum number nodes (MNN) is

better than that of the k-r allocation policy based on the minimum slices number (MSN); however, with the same security and availability guarantees, the MNN policy will have more copies relative to the MSN policy, which is relatively wasteful of space.

Li, et al. [11] proposed a blockchain-based decentralized storage system with reliable deduplication and storage balance strategy to provide reliability for deduplicated outsourced data. Encrypted data is split into chunks by a ramp secret sharing scheme, and it is distributed to multiple independent cloud servers. States of the chunks are recorded on the tamper-proofed blockchain, and they can be used to recover the raw data or support the verification of user identity. To balance the distribution of data among storage servers, a heuristic matching algorithm is designed to efficiently allocate the available storage space. The allocation services are published by autonomous smart contracts and other participants gain rewards by giving the best matching results of data chunks and storage servers. Formulation analysis demonstrates the correctness of the proposed scheme in terms of data consistency, integrity, and reliability. Experimental results show that the proposed scheme preserves the confidentiality of outsourced data with acceptable computational consumption.

Wang, et al. [12] explored blockchain storage optimization from the perspective of data management, analyzed current techniques such as pruning technique, IPFS optimization, sharding, erasure coding, deduplication, and data compression. It also discusses the challenges in blockchain scalability and provides directions and prospects for future research.

Belfqih, et al. [13] proposed a decentralized authentication protocol leveraging blockchain technology and the IPFS data management framework to provide secure and real-time communication between IoT devices. Using the Ethereum blockchain, smart contracts, elliptic curve cryptography, and ASCON encryption, the proposed protocol ensures the confidentiality, integrity, and availability of sensitive IoT data. The mutual authentication process involves the use of asymmetric key pairs, public key registration on the blockchain, and the Diffie–Hellman key exchange algorithm to establish a shared secret that, combined with a unique identifier, enables secure device verification. Additionally, IPFS is used for secure data storage, with the content identifier (CID) encrypted using ASCON and integrated into the blockchain for traceability and authentication.

Feng, et al. [14] proposed an Information-Centric Networking-based blockchain storage architecture. The architecture uses the enhanced resolution system for community division to build blockchain node partitions and store blockchain ledgers in the underlying network. It introduces virtual chain for rapid blockchain indexing and adopts a collaborative block replica deletion algorithm across neighboring partitions, including replica number decision based on blockchain access decay characteristics and replica deletion based on resource relationship. Finally, we compare and analyze the proposed blockchain storage architecture with BC-store and KASARASA, and the results demonstrate that this architecture has significantly lower average access time than others. The replica data volume of this method is reduced by 57.2% compared to the full replica policy, but the access time is only 5.2% slower when compared to the full replica policy, which substantially increases the replica storage utilization.

Kandpal, et al. [15] Analysed academic literature on blockchain technology, emphasized three key aspects: blockchain storage, scalability, and availability. These are critical areas within the broader field of blockchain technology. The study employs CiteSpace and VOSviewer to understand the

current state of research in these areas comprehensively. These are bibliometric analysis tools commonly used in academic research to examine patterns and relationships within scientific literature. Thus, to visualize a way to store data with scalability and availability while keeping the security of the blockchain in sync, the required research has been performed on the storage, scalability, and availability of data in the blockchain environment. The goal is to contribute to developing secure and efficient data storage solutions within blockchain technology.

3. PROPOSED METHODOLOGY

The system architecture of the proposed platform is illustrated in Fig. 1. The architecture integrates convergent encryption, IPFS storage, and Ethereum blockchain to provide secure and decentralized file sharing. The user interface allows users to register, log in, upload files, and download files through the application. When a file is uploaded, the system first performs file chunking, where the file is divided into smaller blocks to improve storage efficiency and management. These chunks are then processed using convergent encryption, which ensures that identical files produce identical encrypted outputs, enabling data deduplication.

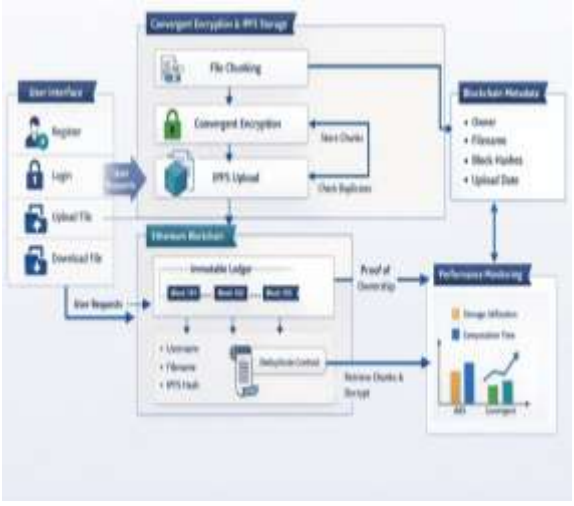

Fig 2. Proposed system architecture of BHODM.

After encryption, the chunks are uploaded to the IPFS, which stores the data in a distributed network and generates unique content hashes for retrieval. The Ethereum blockchain records important metadata such as username, filename, and IPFS hash in an immutable ledger to ensure transparency and data integrity. A deduplication smart contract checks whether the same encrypted chunk already exists in the system to prevent redundant storage. During file download, the system retrieves the corresponding chunks from IPFS and decrypts them to reconstruct the original file while performance monitoring evaluates storage utilization and computation time of the system.

4. RESULT ANALYSIS

Fig 3 illustrates the home page of the proposed BHODM system developed using Django. The home page provides navigation options such as user login, new user registration, file upload, and system information. It serves as the primary interface between users and the backend system. The design ensures user-friendly interaction while maintaining secure session handling. Fig 4 shows the user registration interface where new users can create an account. The signup module collects essential information such as username, email, and password, which are securely stored using Django's authentication framework. Passwords are hashed before storage to enhance security. This module ensures only authenticated users can access system features. Fig 5 displays the login screen and successful authentication process. The system verifies user credentials using Django's authenticate() and login() functions. Once authenticated, the user session is created, allowing secure access to file upload, download, and metadata viewing features. This step ensures controlled and authorized access to the decentralized storage system.

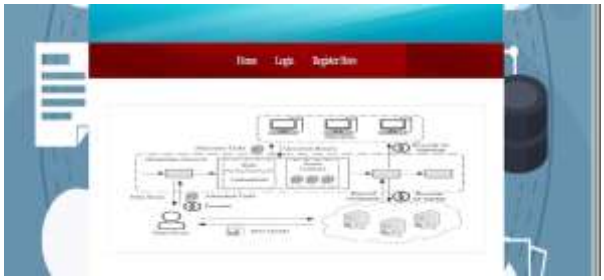

Fig 3. Home page for the proposed BHODM system.

Fig 4. New user signup screen for the proposed BHODM system.

Fig 5. New user login successfully for the proposed BHODM system.

Fig 6 demonstrates the file upload interface along with the duplicate detection mechanism. When a file is uploaded, the system generates a SHA-256 hash for CE and checks for duplicate content. If duplicate chunks are detected, the system avoids re-uploading data to IPFS and instead references the existing CID. This confirms the successful implementation of secure deduplication. Fig 7 shows the file download process from IPFS. Upon request, the system retrieves metadata from the Ethereum blockchain, fetches encrypted file chunks using the CID, and performs AES-CTR. The original file is reconstructed and delivered to the user securely. This verifies data integrity and successful decentralized retrieval. Fig 8 illustrates the metadata viewing section of the system. Users can view file-related information such as file hash, IPFS Content Identifier (CID), owner address, and transaction details stored on the blockchain. Since metadata is stored via smart contracts, it remains immutable and tamper-proof. This feature ensures transparency and auditability.

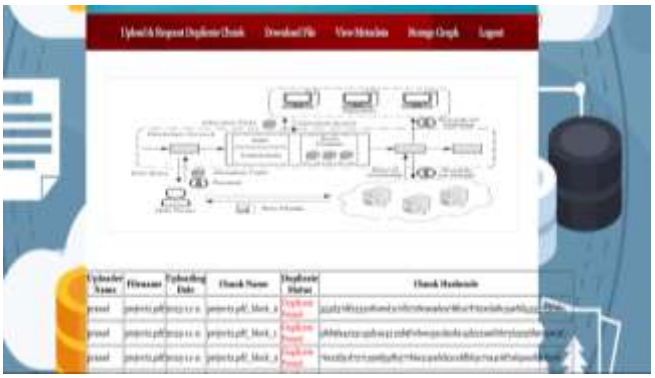

Fig 6. Upload & request duplicate chunks for the proposed BHODM system.

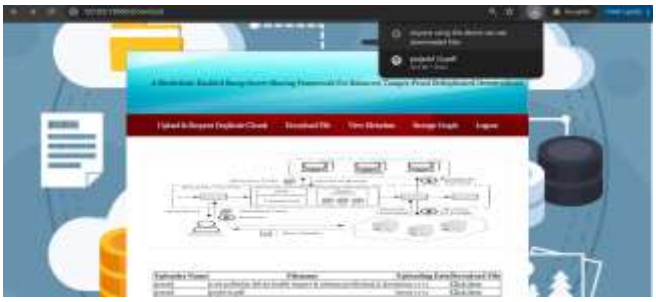

Fig 7. Download file from proposed BHODM system.

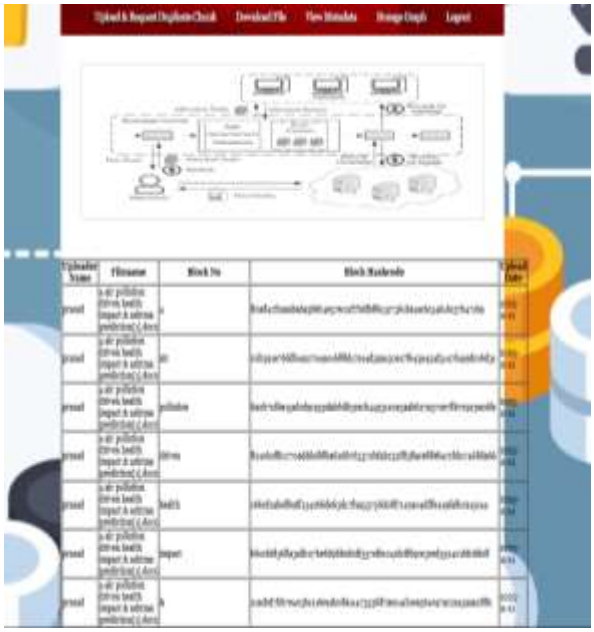

Fig 8. View metadata for the proposed BHODM system.

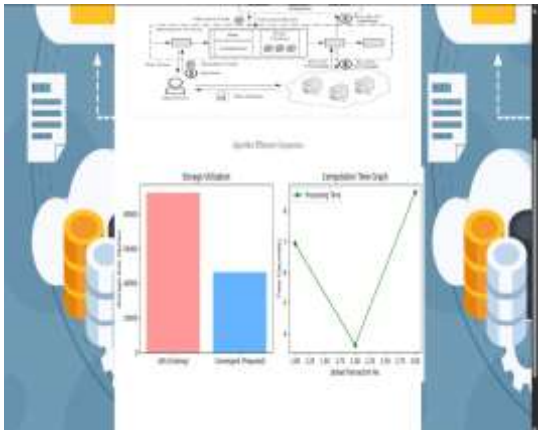

Fig 9. Storage graphs for the proposed BHODM system.

Fig 9 presents a graphical analysis of storage utilization and deduplication performance. The graph compares storage usage before and after deduplication. Results show reduced storage consumption due to CE and chunk-level deduplication. This confirms that the proposed system optimizes storage space while maintaining strong encryption and decentralized architecture.

5. CONCLUSION

The proposed BHODM system successfully integrates CE, AES-CTR encryption, IPFS decentralized storage, and Ethereum blockchain technology to provide a secure and efficient cloud storage solution. The system effectively eliminates data redundancy through content-based deduplication while maintaining strong confidentiality using SHA-256–based key generation and AES encryption. By storing encrypted files in IPFS and maintaining immutable metadata on the blockchain, the system removes dependency on centralized cloud providers and eliminates single-point failures. Smart contracts ensure transparent ownership management and secure access control. The implementation demonstrates reduced storage consumption, improved data integrity, and reliable decentralized retrieval mechanisms. Results confirm that the system performs efficiently even with duplicate file uploads, and encryption overhead remains within acceptable limits. The decentralized architecture enhances reliability, transparency, and security compared to traditional cloud storage systems. Overall, the project achieves its objective of building a secure, scalable, and storage-efficient proposed BHODM platform.

REFERENCES

[1] Wilkinson, S., et al. (2022). Storj: A Secure Decentralized Cloud Storage Network. Proceedings of the IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 2022.

[2] Xu, X., & Chen, J. (2019). Blockchain-based Secure Data Deduplication Scheme for Cloud Storage. Journal of Network and Computer Applications, 128, 25-34.

[3] Liu, L., & Pu, C. (2021). Decentralized Storage Systems: A Survey. ACM Computing Surveys, 50(3), Article 41.

[4] Wang, F., & Liu, Y. (2020). Load Balancing in Peer-to-Peer Storage Networks. IEEE Transactions on Parallel and Distributed Systems, 28(4), 1234-1247

[5] Faris, H., & Moustafa, N. (2020). A Blockchain-Based Storage Model with Efficient Deduplication and Access Control. IEEE Access, 8, 140733-140745.

[6] Nakamoto, S. (2008). Bitcoin: A Peer-toPeer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

[7] Benet, J. (2019). IPFS - Content Addressed, Versioned, P2P File System. arXiv preprint arXiv:1407.3561.

[8] Ren, K., Wang, C., & Wang, Q. (2022). Security Challenges for the Public Cloud. IEEE Internet Computing, 16(1), 69-73.

[9] Merlec, M.M.; In, H.P. Blockchain-Based Decentralized Storage Systems for Sustainable Data Self-Sovereignty: A Comparative Study. Sustainability 2024, 16, 7671. https://doi.org/10.3390/su16177671

[10] Meng, L.; Sun, B. Research on Decentralized Storage Based on a Blockchain. Sustainability 2022, 14, 13060. https://doi.org/10.3390/su142013060

[11] Li, J.; Li, Y.; Wu, J.; Zhang, Z.; Jin, Y. Blockchain-Based Decentralized Cloud Storage with Reliable Deduplication and Storage Balancing. IEEE Trans. Netw. Sci. Eng. 2024, 11, 3289–3304, https://doi.org/10.1109/tnse.2024.3369630.

[12] Wang, Y.; Wang, H.; Cao, Y. Comprehensive Review of Storage Optimization Techniques in Blockchain Systems. Appl. Sci. 2025, 15, 243. https://doi.org/10.3390/app15010243

[13] Belfqih, H.; Abdellaoui, A. Decentralized Blockchain-Based Authentication and Interplanetary File System-Based Data Management Protocol for Internet of Things Using Ascon. J. Cybersecur. Priv. 2025, 5, 16. https://doi.org/10.3390/jcp5020016

[14] Feng, H.; Wang, J.; Li, Y. A Blockchain Storage Architecture Based on Information-Centric Networking. Electronics 2022, 11, 2661. https://doi.org/10.3390/electronics11172661

[15] Kandpal, M.; Goswami, V.; Priyadarshini, R.; Barik, R.K. Towards Data Storage, Scalability, and Availability in Blockchain Systems: A Bibliometric Analysis. Data 2023, 8, 148. https://doi.org/10.3390/data8100148

[16] Mahesh Ganji. (2025). Enhancing Oracle Cloud HR Reporting Through AI-Driven Automation. Journal of Science & Technology, 10(6), 28–36. https://doi.org/10.46243/jst.2025.v10.i06.pp28-36

[17] Mahesh Ganji. (2025). Enhancing Oracle Cloud HR Reporting Through AI-Driven Automation. Journal of Science & Technology, 10(6), 28–36. https://doi.org/10.46243/jst.2025.v10.i06.pp28-36

[18] Todupunuri, A. (2025). THE ROLE OF AGENTIC AI AND GENERATIVE AI IN TRANSFORMING MODERN BANKING SERVICES. American Journal of AI Cyber Computing Management, 5(3), 85–93. https://doi.org/10.64751/ajaccm.2025.v5.n3.pp85-93

[19] Todupunuri, A. . (2024). Artificial Intelligence Ethics: Investigating Ethical Frameworks, Bias Mitigation, and Transparency

in AI Systems to Ensure Responsible Deployment and Use of AI Technologies. International Journal of Innovative Research in Science,Engineering and Technology, 13(09), 1–14. https://doi.org/10.15680/ijirset.2024.1309002

[20] Sushma Babburi. (2025). Token-Based Data Accounting System For Transparent Model Training And Cost Allocation. American Journal of AI Cyber Computing Management, 5(4), 463–474. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp463-474

[21] Snigdha Gaddam. (2025). SOFTWARE STACK PREPARED FOR AI TRANSITIONING FROM MODULES TO MODELS. American Journal of AI Cyber Computing Management, 5(4), 451–462. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp451-462

[22] Gaddam, S. INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING.

[23] Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 1(4). https://doi.org/10.56975/jaafr.v1i4.501636

[24] Srinivasa Kalyan Immadi. (2025). Harnessing Artificial Intelligence In Oracle Hcm: Revolutionising Workforce Management With Automation And Predictive Analytics. International Journal of Data Science and IoT Management System, 4(4), 7–13. https://doi.org/10.64751/ijdim.2025.v4.n4.pp7-13

[25] S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology, 16(1). https://doi.org/10.71097/ijsat.v16.i1.1403

[26] Suhasnadh Reddy Veluru, Sai Teja Erukude, and Viswa Chaitanya Marella. 2025. Multimodal Detection of Fake Reviews using BERT and ResNet-50. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 877–882.

[27] Cyril, H. P. (2025). Event-Driven Provisioning Architectures For Modern Telecom Networks: Overcoming Legacy Limitations And Enabling Autonomous 6g Operations. International Journal of Advanced Research in Computer Science, 16(6), 75–82. https://doi.org/10.26483/ijarcs.v16i6.7389

[28] Jay Bharat Mehta. (2025). AUTONOMOUS PATCH VALIDATION FOR ZERO-DAY EXPLOITS IN ENTERPRISE CLOUDS. International Journal of

Applied Mathematics, 38(4s), 1270–1285. https://doi.org/10.12732/ijam.v38i4s.685

[29] Reddy, S. K. (2025). Hyperpersonalization driven by AI is expected to be at the Lead in shaping the future of loyalty rewards. Journal of Emerging Technologies and Innovative Research.

[30] Reddy, S. K. R. (2021). Strengthening the Security of Loyalty Reward Systems: An In-Depth Analysis of Emerging Cyber Threats and Protection Mechanisms. Journal of Computational Analysis and Applications, 29(6).

[31] Poojari, R. (2026). Privacy-Preserving Generative AI in Healthcare Systems Using Federated Learning Approaches. International Journal of Data Science and IoT Management System, 5(1), 78-88.

[32] Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50

[33] Saikumar, B. (2024). Optimizing Crew Scheduling and Absence Management using Microservices: Enhancing Reliability and Efficiency in Crew Management Systems. International Journal of Enhanced Research in Management & Computer Applications, 13(11), 50–55. https://doi.org/10.55948/ijermca.2024.0116

[34] Saikumar, B. (2023). Enhancing Client Engagement through AI-Driven Real-Time Reporting and Automated Alerts. International Journal of Enhanced Research in Science, Technology & Engineering, 12(11), 111–117. https://doi.org/10.55948/ijerste.2023.111