Verifying AI Content Authenticity Using Decentralized Technologies

TREO Talk Paper

Arthur Carvalho
Farmer School of Business
Miami University
arthur.carvalho@miamioh.edu

Suman Bhunia
College of Engineering and Computing
Miami University
bhunias@miamioh.edu

Abstract

In the rapidly evolving domain of digital content creation, particularly with generative artificial intelligence (AI), establishing and verifying the authenticity and origin of content has become increasingly critical to maintain trust, protect intellectual property rights, and prevent misinformation. This work-in-progress explores how decentralized technologies can be employed to ensure the provenance and integrity of content generated by AI. In particular, we consider the conjunction of the *InterPlanetary File System* (IPFS) with *blockchain* technology. IPFS is a decentralized and peer-to-peer storage technology that is open and free. Data stored on IPFS gets a unique fingerprint called a *cryptographic hash* and is distributed across a network of nodes, thus ensuring redundancy and high availability. Blockchain, on the other hand, is a distributed ledger technology known for its key attributes of decentralization, immutability, and transparency. Each transaction on a blockchain is recorded in a block and hash linked to the previous block. This structure ensures that data cannot be altered retroactively once entered into the blockchain.

Integrating IPFS and blockchain provides a powerful solution to store and verify the authenticity of content created by generative AI. Specifically, that content can be stored on IPFS and be associated with a unique hash value. This hash acts as a digital fingerprint of the content, as it is immutable and unique. IPFS does not natively keep track of data ownership. Thus, the unique hash generated by IPFS can be indexed and organized on a blockchain. A similar solution was adopted by Tang et al. (2020), who suggested storing blocks of massive datasets on IPFS to achieve network efficiency while indexing hashes on a blockchain network. By storing these hashes on a blockchain, the content's integrity and origin are immutably recorded, thus providing a solution for the traceability of digital content ownership, similar to how some non-fungible-token projects currently operate (Carvalho et al., 2023). Timestamps can also be added to the blockchain entry, providing a tamper-proof record of the creation time. Anyone needing to verify the authenticity of certain content can calculate its hash and determine whether that hash is stored on the blockchain. Furthermore, anyone can also retrieve the content from IPFS with that hash. If the hashes match, it confirms that the content has not been altered after its initial creation and blockchain entry.

While the combination of IPFS and blockchain represents a significant step forward in establishing a reliable and transparent environment for creating and distributing AI-generated content, that solution is not without challenges. First, while organizations developing generative AI models can be coerced through regulations to adopt the solution described here, the same might not be feasible when it comes to individuals using open-source models locally. Another challenge is that even minor modifications to a text, image, or sound can drastically alter their hash values. Consequently, while the original content produced by generative AI models is retrievable and verifiable through the proposed system, that is no longer true after slight variations of that content done ex-post due to different hash values.

References

Carvalho, A., Zavolokina, L., Bhunia, S., Chaudhary, M. 2023. "Promoting Inclusiveness and Fairness through NFTs: The Case of Student-Athletes and NILs," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

Tang, J., Jia, T., Chen, H., Wei, C. 2020. "Research on Big Data Storage Method based on IPFS and Blockchain," in *Proceedings of the 2020 International Conference on Video, Signal and Image Processing*.