Decentralized Metaverse Governance using Blockchain with Attribute-based Identity

Md Facklasur Rahaman, Mohtasin Golam, Esmot Ara Tuli, Dong-Seong Kim, and Jae-Min Lee
Networked Systems Laboratory, Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
(facklasur, golam248, esmot, dskim, and ljmpaul)@kumoh.ac.kr

***Abstract*—As the metaverse grows, a mix of real and virtual worlds grows quickly, and there are worries about offensive language and behavior online. This paper introduces a way to manage the metaverse that uses blockchain and attribute-based identity, making it decentralized. It uses smart contracts to stop hate speech, creating a friendly space for users. The model includes a credit scoring system: if users say offensive things, they get penalties, making them accountable. The model has been tested using Remix IDE, and it works well for controlling access, keeping records, and punishing rule-breakers. This model is compared with others, showing it's good at stopping hate speech in the metaverse. This approach promotes openness, inclusiveness, and fair management, changing how people interact in the metaverse.**

***Index Terms*—Blockchain, metaverse, smart contract, unusual behavior.**

I. Introduction

The Metaverse, a virtual ecosystem that combines Blockchain, Artificial Intelligence (AI), and Web 3.0, provides a huge, unbounded digital canvas for endless innovation and frictionless mobility between two separate realities [1]. Technological advancements in online platforms have raised concerns about safe and respectful interactions, as the freedom to express opinions and emotions has led to the proliferation of harmful content.

Unusual behavior, which encompasses discriminating terminologies, may pose a substantial danger to online communities and the welfare of people, especially on immersive Metaverse platforms, where hostile behavior can inflict significantly greater harm [2]. Metaverse platforms are now plagued by incidents of harassment, violence, and hate speech, resulting in users being subjected to virtual harassment and physical attacks [3]. The research in [4] introduces a hate speech detection system specifically designed for the metaverse using deep learning (DL). The system utilizes a lightweight multilayered perception (MLP) model tailored for real-time virtual interactions. A similar methodology is introduced in [5], wherein convolutional neural networks (CNN) are employed to identify and censor objectionable phrases that may manifest in the metaverse as hate speech. The research in [6] offers MetaHate, an approach for identifying and suppressing hate speech in online gaming environments using Blockchain technology and Artificial Intelligence. Blockchain technology guarantees transparency and user responsibility, while AI algorithms detect harmful linguistic patterns. A smart contract is employed for speech moderation to improve Metaverse safety and inclusiveness. However, there is a notable absence of literature regarding managing unusual behavior occurrences in the Metaverse, particularly in keeping records, tracking, and ensuring punishment transparency.

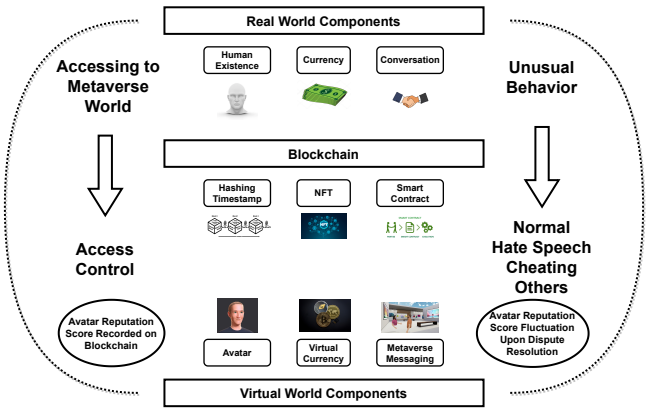

Fig. 1: Proposed approach for hate speech prevention using blockchain-based smart contract solution

In this paper, a Blockchain-enabled smart contract-based solution has been proposed to address the impact of unusual behavior like hate speech, promote user welfare, and prevent virtual harassment by keeping a record of user activity and tracking access.

II. Proposed System

The proposed approach is illustrated in Fig. 1, depicting components' sequential progression and interplay across the workflow. The main purpose of this framework is to explain how it works and how it improves safety within the metaverse using a solution based on smart contracts. The framework has two layers, assuming that AI will detect obscene speech. Subsequently, a blockchain-based solution will employ smart contracts to implement the necessary mechanisms for prevention. The natural language processing (NLP) technique examines input text to identify unusual behavior, prompting immediate reactions. Blockchain technology guarantees transparency and security by issuing alerts or prohibitions according to the level of seriousness. The proposed Metaverse approach implements a credit scoring system for every user, wherein the decentralized governing management decides the baseline credit levels. Users who disseminate unusual behavior are subject to sanctions such as credit deductions or warnings. Upon reaching a specific credit threshold, individuals encounter

```
"from": "0xd9145CCE52D386f254917e481eB44e9943F39138",
"topic": "0xa87e2c5a27f0eba604ccced4dc59f3997e7e731872951782d78bff87f34c874d",
"event": "UserRegistered",
"args": {
        "0": "0xAb8483F64d9C6d1EcF9b849Ae677dD3315835cb2",
        "1": "Paul",
        "2": "19951101",
        "3": "123456789",
        "userAddress": "0xAb8483F64d9C6d1EcF9b849Ae677dD3315835cb2",
        "name": "Paul",
        "dateOfBirth": "19951101",
        "nationalId": "123456789"
```


Fig. 2: User Registered Directly by Metaverse Authority

TABLE I: Comparative assessment of the proposed concept

Ref. Model	Method		Data Record	Access Control	Penalization
	AI	BC			
[4]	✓	×	×	×	×
[5]	✓	×	×	×	×
[6]	✓	✓	×	×	✓
Proposed Model	✓	✓	✓	✓	✓

restrictions. A smart contract solution utilizing blockchain technology effectively oversees user access across numerous accounts by monitoring unlawful actions and developing an identity access system that relies on particular attributes.

III. Implementation and Performance Analysis

A. Testing and Validation

This section clearly explains how the system is implemented, using the Remix IDE (Integrated Development Environment) as the tool for deploying a smart contract written in Solidity language. The smart contract serves two key purposes: overseeing access control for Metaverse users and preserving a database of behavioral credentials. Regarding access control, users are granted admittance into the Metaverse network by submitting additional requests that include necessary attributes. If the Metaverse authority determines it to be appropriate, users are included, and the authority also has the power to directly include users by entering the necessary qualities, which are depicted in Fig. 2 showing the addition of a user by the Metaverse authority.

Consequently, within the Metaverse network, user interactions facilitate discussions. If users see offensive activity, they can lodge complaints with the governing body of the Metaverse. After confirming evidence and deciding, the governing body settles conflicts by warning. If the warnings fail to have the desired effect, the user's credit score will be reduced. Upon entry, members start with a credit score of 100; any post-dispute conviction deducts 10 points, ensuring accountability. When disputes are settled in the Metaverse, the person in charge of the Metaverse will write down and keep the relevant information on a Blockchain. After that, a search tool lets users see the dispute history of a certain address. As you look through the past, each entry has the number of disputes, information on how they were resolved, and a timestamp that makes it easy to find each event's exact date and time. This open and unchangeable method makes the Metaverse ecosystem more accountable and open. Fig. 3 depicts the visual demonstration after solving any disputes by the authority.


```
"from": "0xd9145CCE52D386f254917e481eB44e9943F39138",
"topic": "0x0c60aad9f3b892511cde7a52cfff9d00f861cc52459057281fe0dfdd2e767536",
"event": "DisputeResolved",
"args": {
        "0": "0xAb8483F64d9C6d1EcF9b849Ae677dD3315835cb2",
        "1": "0x4B20993Bc481177ec7E8f571ceCaE8A9e22C02db",
        "2": 2,
        "3": "1703745169",
        "4": true,
        "complainant": "0xAb8483F64d9C6d1EcF9b849Ae677dD3315835cb2",
        "accused": "0x4B20993Bc481177ec7E8f571ceCaE8A9e22C02db",
        "behavior": 2,
        "timestamp": "1703745169",
        "resolved": true
```


Fig. 3: Dispute Resolved by Metaverse Authority

B. Performance Evaluation

The proposed model is being evaluated alongside similar approaches previously mentioned in the manuscript for hate speech management on the Metaverse platform. The proposed technique seeks to manage hate speech (access control, keeping records, and penalization) and proactively mitigate its impact on users. Thus, the comparison has prioritized the proposed concept rather than assessing the state-of-the-art model's effectiveness, as depicted in Table. I.

IV. Conclusion and Future Work

In conclusion, our proposed decentralized approach to Metaverse governance, integrating Blockchain and attribute-based identity, demonstrates a promising solution to address the rising concerns of offensive language and unusual behavior within virtual ecosystems. By leveraging smart contracts and a credit scoring system, we effectively curb hate speech, creating a safer and more accountable environment for Metaverse users. The comparative assessment highlights the model's efficacy in managing access control, record-keeping, and penalization, positioning it favorably against existing approaches. Further work should focus on refining NLP, an advanced AI algorithm for hate speech detection, and enhancing the credit scoring system.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 50%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 50%)

References

[1] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, and X. Shen, "A survey on metaverse: Fundamentals, security, and privacy," *IEEE Communications Surveys & Tutorials*, 2022.

[2] U. Upadhyay, A. Kumar, G. Sharma, B. B. Gupta, W. A. Alhalabi, V. Arya, and K. T. Chui, "Cyberbullying in the metaverse: A prescriptive perception on global information systems for user protection," *Journal of Global Information Management (JGIM)*, vol. 31, no. 1, pp. 1–25, 2023.

[3] S. Frenkel and K. Browning, "The metaverses dark side: Here come harassment and assaults," 2021.

[4] J. N. Njoku, A. U. Eneh, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Metahate: Text-based hate speech detection for metaverse applications using deep learning."

[5] R. M. Medina, J. N. Njoku, and D.-S. Kim, "Audio-based hate speech detection for the metaverse using cnn," , pp. 667–668, 2022.

[6] H. Sanghvi, R. Bhavsar, V. Hundlani, L. Gohil, T. Vyas, A. Nair, S. Desai, N. K. Jadav, S. Tanwar, R. Sharma *et al.*, "Metahate: Ai-based hate speech detection for secured online gaming in metaverse using blockchain," *Security and Privacy*, p. e343.