

AI-Driven Cognitive Manipulation Risks in Metaverse Environments: Challenges, Implications and Security Solutions

Khushi Dinesh Krishnani¹, Sunny Nahar²

¹Student, Department of Master of Computer Applications, Vivekanand Education Society's Institute of Technology (VESIT), Mumbai, India

²Professor, Department of Master of Computer Applications, Vivekanand Education Society's Institute of Technology (VESIT), Mumbai, India

Abstract - The Metaverse is advancing gradually as a virtual reality universe with AR and AI avatars where users are involved by interacting with it. These technologies, in addition to the known cyber issues, pose new security/privacy challenges and provide opportunities for user engagement and personalisation. An emerging challenge is the cognitive manipulation (via artificial intelligence) that can analyse the behaviours, emotional responses, interaction patterns and personal preferences of the users and target them to the manipulation of their perception, decision making and virtual experiences. Cognitive manipulation is an attack on human judgement, not like traditional attacks that are aimed at data theft or unauthorized access. AI uses user data to create more effective and personalized virtual experiences. The study investigated the effects of AI on users' privacy, freedom and trust in the Metaverse. It also refers to things that can influence users such as personal recommendations, content based on emotions as well as different virtual activities. The research also explores issues that can occur when AI influences virtual environments. A Cognitive Security Framework is proposed to mitigate these risks. This approach allows users to learn how AI works, gives them more control over their data and helps to secure their information. Insights underscore the importance of correct use of AI and reinforced governance frameworks to guarantee the safe and trustworthy use of the Metaverse. It will also help those interested in the new risks that virtual spaces and AI may pose, and offer some guidelines for the human-centric and ethical Metaverses of the future.

Key Words: Metaverse, Artificial Intelligence, Cognitive Manipulation, Psychological Privacy, Behavioural Profiling, Digital Trust, Virtual Reality, Cognitive Security

1. INTRODUCTION

The Metaverse is an emerging virtual space that enables users to connect, communicate, collaborate and engage in virtual experiences using technologies like Virtual Reality (VR), Augmented Reality (AR), and Artificial Intelligence (AI). The Metaverse, by combining these technologies, offers immersive and interactive experiences that are beyond the scope of traditional internet platforms. AI is a key component of the Metaverse development. Applications of AI can involve virtual assistants, realistic digital avatars,

personalized recommendations, auto-generated, and virtual environments. These features provide greater content user engagement and help create more realistic virtual experiences. The existing literature is generally related to threats of data leakage, ID theft, Internet attacks, and electronic fraud. Less attention, on the other hand, has been paid to cognitive manipulation by AI, whereby AI systems can, over time, alter the perception, preferences, emotions and decisions made by the users by analyzing and adapting their behaviour. The Metaverse is what makes AI systems different from traditional online platforms, as they can interact with users for longer periods of time. This increases the chance of an indirect influence on their behaviour due to the possibilities of making target-oriented suggestions, emotionally flexible content and tailored virtual experiences. This can have an impact on user autonomy, psychological privacy and digital trust. It is essential, therefore, to understand the risks related to metaverse scenarios that can include cognitive manipulation with AI. This paper investigates the root causes, implications, security concerns, and proposes solutions for the creation of secure, transparent and user-centric Metaverse platforms.

2. PROBLEM STATEMENT

AI is playing a crucial role in the Metaverse by creating personalized suggestions from algorithms, intelligent virtual assistants, lifelike virtual avatars, and personalized virtual experiences. In addition to providing better user engagement and interaction, these technologies also give an opportunity for excessive behavioural influence and cognitive manipulation. Today's Metaverse systems are always gathering information about users, including their history of interactions, emotional reactions, likes and dislikes, attention spans, and behaviour. The AI uses this data to deliver personalized content and virtual experiences to users. However, too much personalisation can affect how users think, feel and make decisions without them knowing it. Earlier studies mostly focused on privacy problems, identity theft and cyber threats. Less attention was given to how AI can slowly influence user behaviour in virtual environments. This may affect user privacy, freedom of choice, and trust. Therefore, it is important to understand and reduce these risks such that virtual environments can be made safe and useful for users.

3. LITERATURE REVIEW

The Metaverse platforms have been progressing different domains like education, healthcare, entertainment, and business with the latest advancements in Artificial Intelligence and immersive technologies. The applications of AI for the development of Intelligent Virtual Environments, Realistic Avatars, Personalized User-Experiences and Adaptive Digital Ecosystems have been investigated.

Many studies highlight the importance of AI technologies such as Machine Learning, Natural Language Processing, Computer Vision and Generative AI to improve the interaction with the users in virtual environments. These technologies enable users to communicate with each other, create content and get useful suggestions. But there can also be concerns about privacy, security and safe use of AI.

While there have been studies on the advantages and difficulties of these AI-driven Metaverse systems, few have focused on the potential dangers to the cognitive system. Most of the research is on technical and security threats and the effects of AI systems on user behaviour, perception and decision-making is an emerging field.

Table -1: Summary of Existing Literature

Author	Year	Key Contribution	Limitation
Mystakidis	2022	Discussed Metaverse concepts and applications	Cognitive risks not addressed
Huynh-The et al.	2023	Investigated the role of AI technologies within Metaverse platforms	Limited focus on behavioural influence risks
Dwivedi et al.	2023	Analysed opportunities and challenges of Metaverse adoption	Psychological privacy not discussed
Zhou et al.	2025	Reviewed Generative AI applications in the Metaverse.	Cognitive manipulation risks not fully addressed.

The literature review reveals that although there has been considerable research on Metaverse technologies, security and personalisation mechanisms, there is a need for further research into the new threat of cognitive manipulation from using AI. The gap identified is the basis on which the current research is developed.

4. RESEARCH GAP

From the literature review, it is found that there is significant research done on Artificial Intelligence, virtual reality technologies, personalisation mechanisms, and

cybersecurity issues in the Metaverse. The majority of studies set parameters in privacy protection, identity management, data security, user engagement and immersive virtual experiences.

But there has been little focus on the new danger posed by cognitive manipulation using AI. Modern AI capabilities can track users' actions, reactions, interactions, and preferences to tailor virtual experiences to the user. Personalisation can make users happy and interested in virtual environments. But it can also influence how people think, decide and behave.

In many of the earlier studies, the long-term implications of AI on user privacy, freedom and trust were not considered to the same extent. As the use of AI in the Metaverse platforms is increasing, it is essential to be aware of these risks and ensure that the virtual environments are safe for the users. Hence, this paper aims to recognize and investigate the risks of cognitive manipulation in the AI world and suggest some security measures to build a safer Metaverse environment.

5. AI-DRIVEN COGNITIVE MANIPULATION IN METAVERSE ENVIRONMENTS

AI has served as an enabling technology for capabilities and interactions in Metaverse platforms including the provision of personalized experiences, intelligent virtual assistants, realistic avatars, and adaptive virtual environments. These technologies are constantly learning the user's behaviour, preferences and interaction, which leads to a better interaction with the user.

Cognitive manipulation powered by artificial intelligence is the use of personalisation and behavioural analysis to shape, guide or influence a person's thoughts, perceptions, feelings or choices. AI systems in the Metaverse will be able to gather large amounts of data from users, including their browsing behaviours, interactions, emotional reactions, levels of engagement and communication styles.

Then the data is used to create customized virtual experiences for individual tastes. Personalisation improves user satisfaction, but over time, the overreliance on AI-generated recommendations can subtly influence user decisions, unbeknownst to them. The algorithm might then decide to show the user the same product, point of view, service or virtual experience again and again.

Behaviour, decision-making and perception can be impacted by prolonged exposure. The Metaverse is, by nature, more immersive and therefore the cognitive manipulation is likely to be more impactful. Users spend a lot more time in virtual environments than they do on traditional digital platforms, and these environments allow them to interact with AI systems that are capable of dynamically adapting content in real time, on an ongoing basis.

Unlike conventional digital platforms, users spend long durations in virtual environments, where they can interact with AI systems that can continually tailor the content in real-time. This fosters the possibilities for subtle behavioural influence which can impact the autonomy and psychological privacy of the user.

With the advancement of AI technologies, the grasp and management of cognitive manipulation risks will play a pivotal role in the development of safe, ethical, and trustworthy Metaverse environments.

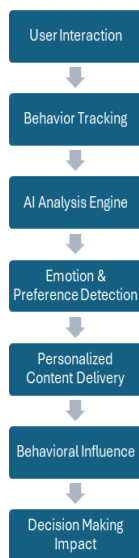


Fig-1: Cognitive Manipulation Process

6. CAUSES OF COGNITIVE MANIPULATION

Cognitive manipulation with Artificial Intelligence in the Metaverse is the process of intelligent systems continuously observing users' behaviour and building a profile based on this data to manipulate their perceptions, emotions, and decision-making. There are a number of technological reasons for this. These factors allow for very individualized virtual experiences, yet can likewise put the individual at a hazard for behavioural influence and psychological manipulation.

6.1 Behavioural Profiling

Behavioural profiling is one of the main reasons for cognitive manipulation. The user behaviour of AI systems includes tracking of browsing, interactions, communication, virtual purchases, and movement in digital environments. The information is fed to make a detailed user profile of the user that can predict the user's future interests and actions. Profiling strengthens personalisation, but could also lead to the display of repeated content, thus shaping users preferences and decision-making in a cumulative way.

6.2 Emotion Recognition Systems

With the advancements of AI technology, emotions can be detected by analysing facial expressions, voice characteristics, eye movements, and interactions with the system. Emotion recognition plays a pivotal role in providing Adaptive and personalized experiences in Metaverse environments.

But emotional information can also be utilized to share articles that resonate with certain emotions and responses. This can heighten the risk of manipulation of thoughts and emotions.

6.3 Personalized Recommendation Algorithms

Recommendation algorithms are aimed at suggesting content that is of interest and preference to the user. These systems also dynamically adapt to user behaviour and adjust the presentation of content to ensure the highest engagement. While recommendations provide a better user experience, over-personalisation could reduce users' access to a variety of information and lead to over-familiarity with certain perspectives, goods, or services.

6.4 Generative AI Influence

Generative AI technologies can make realistic virtual characters, conversations, images and experiences – in real-time. These features help to provide a more immersive experience in the Metaverse and increase engagement. But as more and more incredibly realistic AI-generated content appears, users will find it harder to tell whether information is real or fake. This raises further questions around trust, transparency and mental security.

Table -2: Factors Contributing to Cognitive Manipulation

Factor	Purpose	Potential Risk
Behavioural Profiling	User Analysis	Behavioural Influence
Emotion Recognition	Emotion Detection	Emotional Manipulation
Recommendation Algorithms	Personalisation	Decision Bias
Generative AI	Content Creation	Cognitive Influence

7. SECURITY AND PRIVACY IMPLICATIONS

AI can bring many benefits to Metaverse applications, but it can also introduce some security and privacy issues. Cognitive manipulation can influence user's thinking, feeling and behaviour in virtual environments. As AI develops, we should understand these risks and make virtual environments safe and trustworthy for users.

7.1 Psychological Privacy Risks

The psychological privacy is a human right to have his/her thoughts, feelings, ideas and mental conditions kept private from the influence of others. The Metaverse systems will be data-driven, gathering and tailoring behavioural and emotional data to provide unique virtual experiences with AI. Too much of such information gathered and used can reveal psychology and potential for mental manipulation. This is about personal freedoms, emotional safety and the ethical application of Artificial Intelligence technologies.

7.2 Reduced User Autonomy

User autonomy is someone's ability to make choices about life; to make choices uninfluenced by others. With time, recommendations, virtual spaces adapted to the users and AI interactions start to shape users' decisions. This can cause users' impressions of the algorithmic suggestions to be mistaken to be their own decisions. This may lead to the loss of freedom by the users and their dependence on intelligent systems.

7.3 Manipulative Advertising and Persuasive Content

AI advertising systems can track user activities and emotional responses to deliver highly targeted content. The marketing of personalised ads can be more effective, but at the same time too much of the good thing can be more manipulative marketing methods and techniques that target individual susceptibilities. These can influence the user's buying decision, opinions and behaviour without them being aware of it.

7.4 Digital Trust Challenges

One of the key success factors of Metaverse platforms is trust. If the possibility of 'cognitive manipulation' exists, the trust users have in the suggestions, interactions and experiences they have online could diminish. Users' trust in an AI system could waver if they feel it's being used to manipulate them, whether by trick or design. It is therefore critical to have transparency and accountability in the creation of secure and reliable Metaverse ecosystems.

7.5 Long-Term Behavioural Impact

The exposure to AI technologies of influencing behaviours may have an impact on a user's behaviour over a longer period of time. Personal preferences, practices and perspectives can have long-term consequences as a result of interactions with customised digital media. This can over time restrict access to a wide range of viewpoints and help to foster behavioural dependency. These are potential (likely) long-term effects that are essential in the responsible use of AI in the immersive digital space.

Table-3: Security and Privacy Implications of Cognitive Manipulation.

Security Concern	Impact on Users	Severity Level
Psychological Privacy Loss	Exposure of emotional and behavioural data	High
Reduced User Autonomy	Influenced decision-making	High
Manipulative Advertising	Behavioural persuasion	Medium
Digital Trust Issues	Reduced confidence in AI systems	High
Long-Term Behavioural Impact	Dependency and perception changes	High

8. PROPOSED COGNITIVE SECURITY FRAMEWORK

The primary goal of this framework is to ensure users' safety, privacy, and trust while allowing them to enjoy personalized experiences. The framework consists of four core components: a Transparency Layer, a User Consent Management, a Behavioural Privacy Protection and AI Accountability Mechanisms. These Components make AI systems safe and easy to use.

The Transparency Layer explains how the AI collects and uses users' data. User Consent Management lets users control how their personal information is used. Behavioural Privacy Protection assists in reducing too much tracking of user activities and protects user information from being used without permission.

AI Accountability Mechanisms assist in the evaluation of AI systems' operation and in confirming adherence to safety standards and guidelines." These measures help make the

Metaverse safer and support responsible AI use. They also help to mitigate the risk of cognitive manipulation.

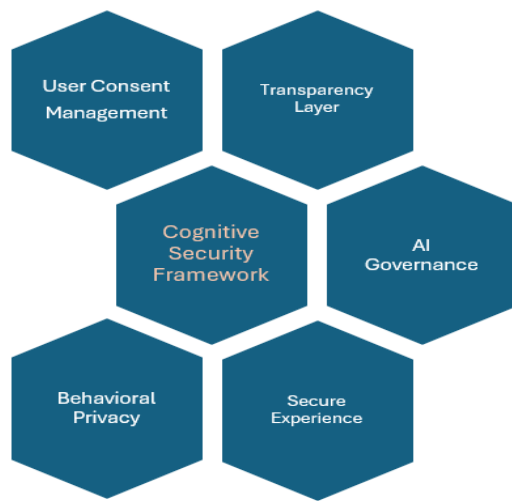


Fig- 2: Proposed Cognitive Security Framework for Metaverse Environments.

9. FUTURE SCOPE

Thanks to the current technological advances, the metaverse is poised to be even more powerful in the future as AI and immersive technology continue to evolve and mature. And every day, virtual environments are getting smarter and more engaging, which means there is even more need for cognitive security and ethical AI. Therefore, further research needs to be done in order to develop machinery that can detect, monitor or diminish the process of cognitive manipulation.

With the emergence of new technologies, e.g. XAI, privacy-preserving machine learning and AI governance frameworks, they will also contribute to further improving transparency and user trust in virtual environments. These technologies could help users better manage their personal, emotional and behavioural information, and minimize the potential for unwanted and unauthorized influence. Future platforms may include monitoring capabilities to detect anomalous user behaviour and potential manipulation attempts, as well as the use of personalized content.

There is also a need for research on common rules, for instance, the ethical guidelines and the international governance regimes. Responsible use of AI is important in immersive digital environments. This will contribute to the design of safe, dependable and user-friendly metaverse environments, with user protection at the core.

10. CONCLUSION

AI is used in many Metaverse platforms to provide personalized experiences, virtual interactions, and

immersive digital environments. These features make the user experience better, but they may also create concerns about cognitive manipulation. Their monitoring of user behaviour, their recommendations, their ability to recognize emotions, and their Generative AI features can affect user actions, emotions and choices. The study explored the possible risks and impact of cognitive manipulation in the AI environment (Metaverse), and proposed a Cognitive Security Framework to address these problems.

It highlights transparency, consent, behavioural privacy and AI governance to help protect against manipulation risks and to ensure that the benefits of personalisation are maintained. These results highlight the need for good AI usage, better privacy protections, and ethical AI governance to ensure that the future of the Metaverse is safe, secure, and user-friendly.

REFERENCES

- [1] D. Mystakidis, "Metaverse," Encyclopedia, vol. 2, no. 1, pp. 486–497, 2022.
- [2] T. H. Huynh-The, Q.-V. Pham, D. T. Nguyen, E.-N. Huh, and W.-J. Hwang, "Artificial Intelligence for the Metaverse: A Survey," Journal of Information Processing Systems, vol. 19, no. 1, pp. 1–23, 2023.
- [3] Y. Dwivedi et al., "Metaverse Beyond the Hype: Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy," International Journal of Information Management, vol. 66, 2023.
- [4] Z. Lv, "Generative Artificial Intelligence in the Metaverse Era," Cognitive Robotics, vol. 3, pp. 208–217, 2023.
- [5] M. Ball, The Metaverse: And How It Will Revolutionize Everything. Liveright Publishing, 2022.
- [6] N. Stephenson, Snow Crash. New York: Bantam Books, 1992.
- [7] H. Zhou, X. Chen, J. Li, Z. Zhang, Y. Fu, M. P. Liva, D. Greenbaum, and P. Hui, "Generative Artificial Intelligence in the Metaverse Era: A Review on Models and Applications," Research, vol. 8, Article ID 0804, 2025, doi: 10.34133/research.0804.
- [8] S. O. Olanipekun, "Computational Propaganda and Misinformation: AI Technologies as Tools of Media Manipulation," World Journal of Advanced Research and Reviews, vol. 25, no. 1, pp. 911–923, 2025, doi: 10.30574/wjarr.2025.25.1.0131.
- [9] L.-H. Lee, T. Braud, P. Zhou, L. Wang, D. Xu, Z. Lin, A. Kumar, C. Bermejo, and P. Hui, "All One Needs to Know About Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda," arXiv preprint arXiv:2110.05352, 2021.

[10] B. Falchuk, S. Loeb, and R. Neff, "The Social Metaverse: Battle for Privacy," IEEE Technology and Society Magazine, vol. 37, no. 2, pp. 52–61, 2018.

[11] T. F. Ask, S. Sütterlin, L. Müller, R. G. Lugo, D. Saari, H. Grahn, M. Canham, D. Hermansen, and B. J. Knox, "Cognitive Security: The Study and Practice of Protecting the Human Mind and Other Cognitive Assets from Cognitive Threats," OSF Preprints, 2024.

[12] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2021.

[13] K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, 2021.

[14] IEEE Standards Association, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," IEEE, 2023.

[15] T. Guo, Q. Wu, and F. Cai, "When Privacy Concerns Drive AI Adoption: A Psychological Perspective," Journal of Research in Interactive Marketing, vol. 20, no. 3, pp. 440–454, 2026, doi: 10.1108/JRIM-02-2024-0122.

[16] J. D. N. Dionisio, W. G. Burns III, and R. Gilbert, "3D Virtual Worlds and the Metaverse: Current Status and Future Possibilities," ACM Computing Surveys, vol. 45, no. 3, pp. 1–38, 2013.

[17] J. Park and Y. Kim, "A Metaverse: Taxonomy, Components, Applications, and Open Challenges," IEEE Access, vol. 10, pp. 4209–4251, 2022.

[18] L. Floridi and J. Cowls, "A Unified Framework of Five Principles for AI in Society," Harvard Data Science Review, vol. 1, no. 1, 2019.

[19] V. Dignum, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer, 2019.