Metaverse Standards Forum
Humanity Attestation in Metaverse Environments

Last Update: May 13, 2025

Status: Approved for Public Distribution

Version: 1.0

Reviewer	Due Date	Status	Contact
Digital Asset Management Working Group	December 17, 2024	Complete	digital_asset_management@lists.metaverse-standards.org
MSF Domains (Peer Review)	March 05, 2025	Complete	oversight@lists.metaverse-standards.org
Use Case Taskforce	May 13, 2025	Complete	use_case_task_force@lists.metaverse-standards.org

The purpose of this template is to provide a structured framework for collecting and documenting use cases within the Metaverse Standards Forum (MSF). Use cases are essential for understanding real-world scenarios where metaverse technologies are applied and where interoperability challenges may arise. This template guides MSF members in providing a concise yet comprehensive description of a use case, including its title, identifier, and summary. It also encourages contributors to list the benefits of the use case, identify actors or entities involved, and describe the use case scenario in detail, emphasizing interactions, challenges, and requirements. Additionally, it prompts the inclusion of relevant technical information, such as implementations, success metrics, and challenges faced. This template aims to facilitate the gathering of valuable use-case data to inform standards development and foster collaboration within the MSF community.

MSF members and MSF Domain Groups are invited to submit use cases.

NOTE: Organizations such SDOs who want to submit and add a use case would need a sponsor that is an MSF member. This process is established in order to have a contact person in MSF that can handle discussions and resolve open issues within regular meetings.

Eligible submitters:

- MSF Domain Groups
- MSF Members (Principal and Participant)
- External Organizations with Liaison Agreements (with the support of a MSF member that acts as sponsor)
- Standard Development Organizations (with the support of a MSF member that acts as sponsor)

Minimum Requirements for MSF Member Submissions not part of a Domain Group:

- Minimum required number of proposers: 3
- Minimum required number of supporters: 5

NOTE: Use cases submitted by SDOs and Liaison Organizations would also need to fulfill the same requirements (and would need a sponsor) unless they are submitted by a Domain Group.

MSF: *Metaverse Standards Forum*
POG: *Pre-qualified Organizations and Groups*
SPP: *Standards Related Publications and Projects*
DWG: *Domain Working Groups*
WG: *Working Group*
SDO: *Standards Development Organization*

Use Case Title

Unified Reputation Management for Metaverse Entities

Use Case Identifier

MSF2024-POH-001

- Version 1.0
- Year of Release: 2025

Summary of Use Case

Description: This use case focuses on the development and implementation of mechanisms that help users distinguish between avatars controlled by humans and those operated by artificial intelligence (AI) in metaverse environments.

Benefits:

- Enhances trust among users
- Improves safety within virtual environments
- Supports transparency across digital platforms
- Fosters genuine user interactions

Contributors and Supporters

- Digital Asset Management Working Group
- MSF Domains (Peer Review)
- Use Case Taskforce

Keywords

Digital Identity, Authentication, AI, Human Verification, Avatar, Trust, Safety, Metaverse Security

Actors/Entities

- **Human Users:** An Individual engaging with the Metaverse through a Human-Controlled Avatar. This includes Humans augmented by AI.
- **AI Entity:** A Non-Human, AI-driven agent participating within the same environments. This includes AI Entities directed by a Human with different levels of Human Engagement and control of the AI Entity.
- **Human Verification System:** A Technology or System capable of accurately distinguishing between Human and AI-controlled entities.
- **Data Controller:** The Entity that Collects, Processes, Saves, Shares and Deletes Identity and usage data in the System.
- **Platform:** a virtual world in the Metaverse.
- **Platform Operator:** The Entity responsible for Integrating and Managing the Platform.

Detailed Description of Use Case/Scenario

Preconditions:

- **Active Participation:** Human Users and AI Entities must be active within a Metaverse Platform that supports diverse interactions (e.g., conversations, transactions).
- **Verification System:** The Platform has integrated a Human Verification System.

Main Flow:

1. **User Login:** a Human User logs into the Platform.
2. **Optional Step:** Human Verification System verifies the Human User is a Real Person.
3. **AI Entity Activation:** AI-controlled Avatars now are active within the Platform (either introduced by the Platform or outside the Platform as a user login).
4. **Interaction Initiation:** an Avatar initiates an interaction with another Avatar (one of the two controlled by a Human User) and the Verification System analyses the avatars, as well as the entities controlling the Avatars. Note that an interaction can be as simple as one Avatar viewing another in the Platform or as complex as a chat system or API.

5. **Result Display:** the Verification System displays data to the Human User, indicating the nature of the other Entity (Human or AI, as well as the “grey area” in between) and (optionally) the Verification System’s confidence of the indication.
6. **Continued Interaction:** based on the Verification Result, the Human User decides whether to continue the interaction.
7. **Persistence:** Verification System is able to continually evaluate Users.

Alternative Flow

- **AI System Unavailability:** in cases where the Verification System is undergoing maintenance or updates, Human Users are preemptively notified of the system's unavailability, and (optionally) a general caution is advised for interactions.

Postconditions

- **Successful Verification:** the Human User is informed about the nature of the avatar, leading to informed decision-making regarding the interaction.
- **System Update:** feedback from the Verification Process contributes to ongoing improvements in the Verification System, enhancing its accuracy and reliability.

Implementations and Demonstrations or Technical Feasibility

Specific implementations or demonstrations explicitly focused on differentiating between Human-Controlled and AI-Controlled Avatars within Metaverse Environments are not widely documented. However, various techniques and technologies employed in related fields can serve as a foundation for such systems, indicating the technical feasibility of this use case.

Existing Implementations

- **Captcha and Behavioral Analysis:** traditional web environments utilize Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) systems to distinguish between Human Users and automated Bots. Similar foundational technology, enhanced with behavioral analysis (e.g., movement patterns, interaction timing, location information, IP tracking), could be adapted for real-time Avatar Verification in Metaverse Platforms.
- **Voice Verification Systems:** technologies that analyze voice for authentication purposes, distinguishing Human Speech from Synthetic Voices, could be adapted to verify whether an Avatar is Human-Controlled based on voice interactions within the Metaverse.
- **Biometric Authentication:** while direct Biometric Verification might not be applicable for avatars, the principles behind Biometric Authentication (unique, non-replicable identifiers) can inspire the development of unique “Digital Signatures" for Human Users, differentiating them from AI Entities.
- **KYC/AML Verification Services:** normally Verifies the Identification of a User but could be repurposed to Identify Humans from AI.
- **Circle of Trust Systems:** Systems that rely on Human Attestations such as Proof of Humanity.
- **Video Verification:** Video Recordings of the user logged into a platform to determine if the user is a Human or AI.

Technical Feasibility

The integration of such systems into Metaverse Platforms is technically feasible but requires careful consideration of privacy, cybersecurity, and digital ethics.

Ensuring the System's Reliability while maintaining User Anonymity and Consent is paramount. Success would be measured by the System's Accuracy in Identification, User Trust in the system's verdicts, and the minimization of false positives or negatives.

Challenges:

- **Accuracy and Reliability:** ensuring the system accurately identifies avatars without false positives or negatives. Distinguishing between sophisticated AI and human behavior can be complex, requiring advanced algorithms and continuous updates.
- **Control Switching:** recognizing that an Avatar may initially be Controlled by a Human User, but the Human User may turn over control of the Avatar to an AI Entity. This action has implications for liability and could be difficult to prevent.
- **Additional Accounts:** a User might try to use different accounts to circumvent the Verification System.
- **Privacy and Consent:** balancing the need for Verification with the Privacy Rights of Users. Any System implemented must Respect User Privacy, requiring minimal personal data and ensuring that data collection, processing, and deletion comply with global privacy standards.
- **Adaptability and Scalability:** the System must be adaptable to various Metaverse Platforms and scalable to accommodate growing numbers of users and evolving AI technologies.
- **User Experience:** maintaining a seamless and non-intrusive user experience. Verification mechanisms should not disrupt User Engagement or require cumbersome processes that could detract from the Immersive Experience.
- **Cybersecurity:** protecting the Verification System from potential attacks aimed at bypassing or fooling the System. This includes ensuring the Security of Data Collected for Verification Purposes.
- **Ethical Considerations:** addressing ethical concerns related to AI identification, such as Potential Biases in Verification Algorithms and the Right of AI Entities to participate in certain interactions without Explicit Identification.
- **Interoperability:** ensuring the System can operate across different Metaverse Platforms and technologies, promoting a standardized approach that supports broad adoption.

Requirements:

Technical and Functional Requirements

- **Secure Data Processing:** implement secure data handling and processing mechanisms that comply with global data protection regulations, ensuring user data privacy and consent.
- **High Scalability:** the system must be scalable to handle a large and growing number of users across various Metaverse Platforms without compromising performance.

- **Robust Security Measures:** integrate comprehensive cybersecurity measures to protect the system from unauthorized access, breaches, manipulation and attacks**.**
- **User-Friendly Interface:** provide a simple and intuitive interface for users to request and understand Verification Results without disrupting their immersive experience.
- **Minimal User Disruption:** ensure the Verification Process is quick and non-intrusive, requiring minimal or no active participation from users to avoid disrupting the User Experience.
- **Real-Time Verification:** enable real-time or near-real-time verification to support dynamic interactions within the Metaverse.
- **Feedback Mechanism:** implement a mechanism for users to provide feedback or challenge Verification Results, enhancing system accuracy and user trust over time.

Interoperability Requirements:

- **Standardized Protocols:** develop and adhere to standardized protocols for Avatar Verification, facilitating compatibility, seamlessness and interoperability across different Metaverse Platforms and technologies.
- **Cross-Platform Compatibility:** ensure the system's methods and technologies are compatible with various Metaverse Environments and Virtual Reality Platforms.

Other Key Considerations:

- **Privacy:** incorporate privacy-by-design principles, ensuring that the system respects user privacy at every stage of the Verification Process and obtains necessary consent.
- **Cybersecurity:** robust cybersecurity measures to safeguard User Avatar Data Verification details against breaches and unauthorized access. This includes performing ongoing monitoring and update of Platform protocols to adapt to emerging cybersecurity threats.
- **Identity Verification:** clear Avatar metadata and provenance review process to ensure the accuracy and reliability of Verification Results.
- **Networking and Latency:** optimize network protocols to reduce latency in Verification processes, ensuring timely responses without compromising the quality of User Experience.
- **Provenance:** include mechanisms to capture and track details about AI origin, ownership, function in the platform, types of data collected, offers, benefits, age-gating, cost for interaction (if any), and changes to these from time to time.
- **Ownership:** provide AI Entities and Human Users with the ability to maintain oversight on their Verification data usage, storage and sharing with third parties to ensure continuous compliance with the consent granted to the Data Controller, Platform and Operator.
- **Digital Ethics:** address ethical considerations, such as bias in AI algorithms and the rights of AI Entities, ensuring fair and unbiased Verification Processes.
- **Accessibility:** ensure the Verification System is accessible to users with disabilities and varying levels of technical expertise and accessibility requirements.

Relevant Domain Working Group (WGs):

- NA

Relevant Pre-qualified Organizations and Groups (POGs):

- NA

Relevant Specifications, Publications and Projects (SPPs):

- Proof of Humanity (PoH) is a Sybil-resistance protocol co-developed by Kleros to verify real human identities and help distinguish them from AI (www.docs.kleros.io/products/proof-of-humanity).
- World ID (formerly Worldcoin) is a global Digital Identity project designed to verify real human identities, ensuring authenticity and reducing the impact of AI-driven fake identities (https://worldcoin.org/learn/world-id).
- Privado ID (formerly Polygon ID) is a Self-Sovereign Identity (SSI) solution and Zero-Knowledge Proof protocol that preserves user privacy while enabling human attestation and PoH efforts (https://docs.privado.id).

Related Use Cases

- Identity Verification for Digital Asset Creators Use Case (MSF2024-IDVER-001)
- Shared KYC/AML Verification in the Metaverse Use Case (MSF2024-REPSKA-001)

Additional Comments

- This document is a living artifact and may be subject to revisions on a periodic basis to reflect the future state of Humanity Attestation in Metaverse Environments, and or based on feedback received from MSF stakeholders that warrants an update in the future.