Chapter

Embodied Presence: Realistic Avatars and Trust Verification in Metaverse Environments

Jaime Gallego Vila

Abstract

The metaverse is posited as the next evolution of human–computer interaction, promising unprecedented social presence through realistic avatars. Yet, realism alone is insufficient: As photorealism increases, so does the feasibility of identity spoofing (such as deepfakes) and real-time impersonation. Realism concerns how closely an avatar aligns with a user's appearance and behavior; verification concerns how reliably we can establish – and *continuously* maintain – the link between the avatar and the human who claims it. This chapter confronts that tension by surveying the psychological foundations of presence, state-of-the-art AI methods for photorealistic, behaviorally congruent avatars, and the resulting trust crisis in multiuser environments. We then introduce the AI-driven Continuous Trust Verification Framework: a layered architecture that integrates active/passive biometrics, behavioral signatures, decentralized cryptographic proofs, and interoperability mechanisms as the necessary components of a robust identity fabric. Finally, we examine the transdisciplinary implications and argue that embodied presence and verified trust are inseparable: Without continuous verification, realism erodes plausibility; with it, realism can safely underpin governance, commerce, and public services in virtual worlds.

Keywords: metaverse, virtual worlds, embodied presence, realistic avatars, trust verification, artificial intelligence, digital identity, deepfake, biometrics, decentralized identity, uncanny valley

1. Introduction

1.1 The metaverse as the next paradigm of human–computer interaction

The metaverse represents more than an incremental evolution of the Internet; it is conceived as a paradigm shift in how humans interact with information, computation, and, fundamentally, with each other. It transcends 2D screens to offer a spatial, persistent, and synchronous Internet where individuals coexist and interact through digital representations, or avatars [1]. In this new plane of digital existence, human activities – from collaborative work and commerce to socialization and entertainment – are expected to achieve far greater fidelity and immediacy, blurring the lines between the physical and virtual worlds.

1.2 The double helix of immersion: Plausibility, presence, and the embodied avatar

The success of this paradigm hinges on its ability to generate a profound sense of "presence." Influential researcher Mel Slater breaks this experience down into two critical components: place illusion (PI), which is the sensation of truly "being there" in the virtual environment, and plausibility illusion (Psi), the sensation that the events occurring within that environment are actually happening [2]. While PI can be achieved with basic immersive technology, Psi is more elusive and fundamental; it requires the virtual world and its avatars to behave in a believable manner. Without plausibility, interactions feel artificial, and the sense of presence shatters.

It is at this juncture that we must draw a fundamental distinction between two types of realistic avatars:

1. **Visually similar avatars:** These are 3D representations that achieve a high degree of photorealism, replicating a user's appearance from photographs or scans. However, their animation might be generic or limited. These avatars answer the question "Who am I?" but in a static way. They contribute to PI but can harm Psi if their behavior is incongruent.

2. **Dynamically embodied avatars:** These realistic avatars move beyond mere resemblance. They are equipped with a humanoid rig that is animated in real time using the user's motion data, captured via the virtual reality (VR) headset and controllers. This type of avatar does not just *look* like the user; it *moves* like the user, replicating gestures, posture, and head movements. This direct sensorimotor mapping is the true engine of *embodied presence* and a massive catalyst for Psi. When the avatar acts as a natural extension of one's own body, the mind accepts the virtual events as genuinely "real."

Therefore, presence and realism form a double helix: Visual likeness (the avatar's appearance) and behavioral congruence (the embodied animation) intertwine to build a convincing and plausible immersive experience.

1.3 The obstacle of perfection: The uncanny valley

The pursuit of realism is not a linear progression toward greater user acceptance. This journey is fraught with a significant perceptual pitfall known as the "uncanny valley," a concept first introduced by Masahiro Mori in his seminal 1970 essay, "Bukimi no Tani" (the uncanny valley) [3]. The hypothesis posits that as a robot or digital avatar becomes more humanlike, our affinity for it increases, but only up to a certain point. When the representation is almost perfectly human but fails in subtle, almost imperceptible details – such as eye movement that is slightly too slow, unnatural skin texture, or imperfect lip-syncing – our affinity plummets dramatically, replaced by a sense of eeriness, strangeness, or even revulsion.

As illustrated by the uncanny valley hypothesis (see **Figure 1**), an avatar that falls into the valley directly destroys the Psi described by Slater [2]. It breaks immersion

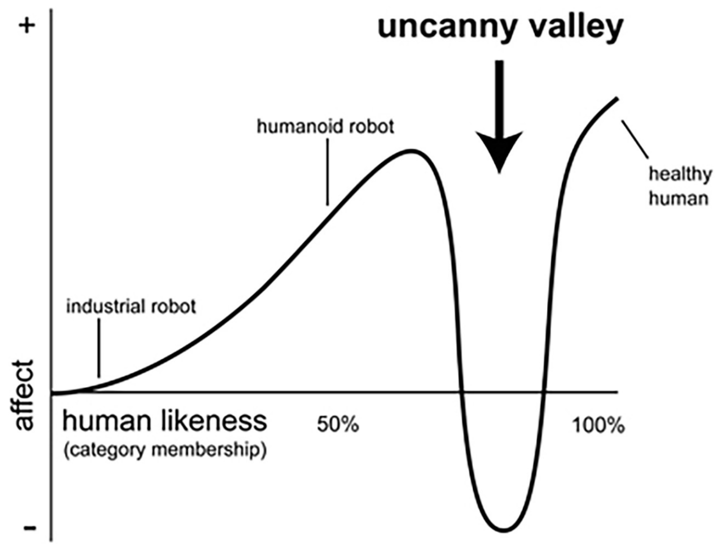

Figure 1.
The uncanny valley hypothesis, as proposed by Mori. As a representation approaches perfect human likeness, a point is reached where subtle imperfections cause a sharp drop in affinity, creating a sense of eeriness. The graph illustrates this dip. (Image recreated by the author based on Mori, 1970).

by constantly signaling that something is "wrong," even if the user cannot consciously pinpoint the flaw. Overcoming this valley is therefore not merely a matter of better 3D modeling but of achieving behavioral realism through the precise, AI-driven animation of the subtle cues that constitute authentic human interaction. This challenge is not merely theoretical; it is a practical obstacle in fields like 3D avatar reconstruction (see **Figure 2**).

This effect is equally prominent in robotics, where famous examples such as the SAYA robot (**Figure 3**) are often cited for their uncanny appearance, despite being

Figure 2.
An example of the uncanny valley in real-time avatar generation. (a) the real person; (b) a reconstruction that captures their likeness but whose subtle imperfections in skin texture and expression can place it in the uncanny valley, creating a sense of unease.

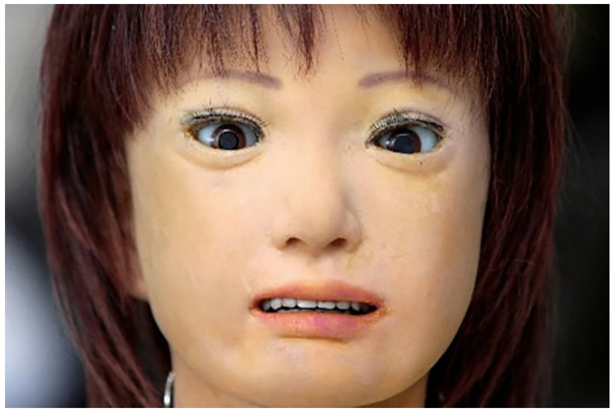

Figure 3.
The animatronic head of the robot SAYA, developed at the Tokyo University of Science by Prof. H. Kobayashi. Designed to express basic emotions, its mechanical nature and subtly unnatural expressions often place it deep within the uncanny valley, highlighting the profound challenge of simulating human behavior.

designed for social interaction [4]. This very challenge highlights the core tension of this chapter: The technologies essential to making an avatar feel genuinely human are the same ones that could be leveraged to create a dangerously deceptive fake.

1.4 The crisis of trust in an age of artificial realism

Paradoxically, the very technology that promises to create realistic, embodied avatars is the source of an existential threat to the metaverse: the systematic erosion of trust. While the current Internet era already suffers from a crisis of confidence fueled by disinformation and phishing, the metaverse magnifies these risks exponentially. It weaponizes the tools of realism – the very AI-driven animations of micro-expressions and gestures needed to cross the uncanny valley – to enable the creation of real-time deepfake avatars. These malicious representations are capable of impersonating identities in a manner that is functionally indistinguishable to unaided human perception [5].

This threat constitutes a direct assault on the Psi that is foundational to presence. Imagining high-stakes interactions – a medical consultation, a corporate negotiation, a legal testimony – in an environment where the interlocutor's identity cannot be guaranteed reveals a fundamental vulnerability. Any hint of inauthenticity shatters plausibility, replacing immersion with suspicion and anxiety. This could not only stymie the mass adoption of the metaverse for critical uses but also confine it to an "era of suspect reality," where the potential for deep human connection is perpetually undermined by the fear of deception. Furthermore, these challenges are magnified in a fragmented digital ecosystem. For any solution to be effective, it must be interoperable, ensuring that identity and trust can be maintained across different virtual worlds and platforms, rather than being trapped in proprietary silos.

1.5 Central thesis and chapter structure

This chapter posits that authentic embodied presence and verified trust are not independent goals but are symbiotic and inseparable. A photorealistic avatar without a verifiable seal of authenticity fails to generate a sustainable Psi, thus creating not presence but anxiety and suspicion. Therefore, the construction of a functional metaverse requires that the technological sophistication of avatars be matched by an equally sophisticated trust architecture. The main contributions of this chapter are as follows:

- A structured survey and synthesis of the foundational psychological theories (embodied presence, plausibility) and the state-of-the-art in AI-driven avatar reconstruction, creating a comprehensive taxonomy of methods ranging from parametric models to neural volumetric representations.

- The proposal of a novel conceptual framework, the AI-driven Continuous Trust Verification Framework (AI-CTVF), serves as a structured solution to the identified trust crisis.

- A critical analysis of the transdisciplinary implications and open research challenges, providing a forward-looking agenda for building safe and trustworthy virtual worlds.

To address these points, we introduce an AI-based conceptual framework for continuous identity verification. The chapter is structured as follows: Section 2 explores the theoretical foundations of embodied presence. Section 3 analyzes the AI technologies that enable the creation of realistic avatars. Section 4 delves into the trust paradigm and its threats. Section 5 details our original proposal: a multilayered framework for trust verification. Section 6 analyzes the transdisciplinary implications. Finally, Section 7 presents case studies and future directions, before offering conclusions in Section 8.

2. Foundations of embodied presence

Having established the critical role of realism in the introduction, this section delves into the foundational psychological mechanisms that enable embodied presence. We move beyond basic definitions to explore the formal models and cognitive processes that explain how a user's mind comes to accept a digital representation as an extension of the self.

2.1 Presence as a dual illusion

As introduced earlier, the experience of presence arises from the interaction of two orthogonal illusions first articulated by Slater [2] and later probed empirically by Skarbez [6]: PI – the feeling of "being there" in the virtual scenario – and Psi – the belief that the events unfolding are actually happening. We can therefore regard presence as a function of both factors.

$$Presence = f(\mathrm{PI}, \mathrm{Psi}), \tag{1}$$

PI can be understood as a largely thresholded phenomenon – either the conditions are sufficient for "being there," or they are not – whereas Psi is continuously modulated by social believability and trust. Crucially, Psi holds a *power of veto*: If users doubt that interlocutors are genuine or that events are authentic, presence collapses regardless of visual fidelity. This is not merely experiential; it undermines *governance* in high-stakes domains (telemedicine, finance, public services), where plausibility is a prerequisite for accountability and due process. This theoretical foundation explains why a single suspected *deepfake* avatar can shatter the integrity of a high-stakes virtual interaction: It directly attacks and nullifies the Psi component, making genuine presence impossible.

2.2 Mechanisms of embodiment: Body ownership and agency

The feeling of an avatar being an "authentic extension" of the self, known as embodied presence or embodiment, relies on two specific neurological illusions: the illusion of body ownership and the illusion of agency, concepts whose application to virtual environments has been both theorized [7, 8] and extensively validated through empirical studies [9].

- **The illusion of body ownership:** It refers to the feeling that an avatar's body *is* one's own body. This is achieved through the congruence of sensory information. For instance, when a user looks down and sees a virtual body that visually matches their own posture and movement (visuomotor correlation), or feels a haptic vibration when their avatar touches an object (visuotactile correlation), the brain begins to adopt the digital form.

- **The illusion of agency:** It is the feeling of being in control of the avatar's actions. This arises from the seamless mapping of a user's intentions to the avatar's behavior. Low latency and precise tracking of movements from VR hardware are critical for maintaining this illusion. Any lag or error breaks the sense of control.

Achieving both illusions is the technical goal of "dynamically embodied avatars." It is this deep, sensorimotor link – the feeling that "this body is mine" and "I am controlling it" – that makes interactions in the metaverse feel impactful and real.

2.3 The avatar's influence on the self: The proteus effect

The relationship between user and avatar is not unidirectional. Remarkably, the characteristics of our digital representation can alter our own behavior and attitudes, a phenomenon known as the proteus effect [10]. Studies have shown that users assigned taller avatars negotiate more confidently, while those with more attractive avatars exhibit greater sociability and self-disclosure.

This effect has critical implications for the trust-verification paradigm. It means that the danger of a deepfake avatar is not merely one of visual impersonation. A malicious actor assuming the avatar of a figure of authority (e.g., a CEO, a police officer) may subconsciously begin to act with more confidence and assertiveness, making their impersonation more convincing and manipulative. The proteus effect thus acts as a psychological amplifier for deception, underscoring that an avatar's appearance is not a neutral costume but a powerful tool that shapes social dynamics, for better or for worse.

DOI: http://dx.doi.org/10.5772/intechopen.1012607

3. The technological frontier: A survey of AI-driven avatar reconstruction

The creation of a realistic, animatable 3D avatar from simple visual input is a foundational goal for the metaverse. This section provides a survey of the state-of-the-art AI technologies that form the modern reconstruction pipeline, highlighting how the field has evolved to overcome key limitations, moving from structured models to flexible neural representations.

3.1 The foundation: Parametric human models

Modern reconstruction pipelines rarely start from a blank slate. Instead, they leverage statistical, parametric models of the human body that encapsulate a vast amount of prior knowledge. The most influential of these is the skinned multiperson linear model (SMPL) [11] and its extension, SMPL-X, which incorporates expressive hands and face [12]. These models can generate a wide variety of human bodies by controlling a low-dimensional set of shape and pose parameters. This approach is powerful because it constrains the problem: AI models only need to infer the correct parameters that best fit a person in an image, a task often achieved via deep neural networks [13, 14]. The primary limitation, however, is that these models have a fixed mesh topology and cannot represent person-specific details such as hairstyles or loose clothing.

3.2 Single-view refinement of parametric models

To address the limitations of the base SMPL model, a significant branch of research focuses on refining its shape using a single photograph.

- **Silhouette and map-based deformation:** Methods such as photo wake-up [15] project the initial SMPL model onto the image plane and then “warp” its vertices to match the person’s 2D silhouette. This allows the model to capture the shape of clothing and hair. The refined 2D information is then back-projected to create a 3D mesh.

- **Texture-space detail inference:** Other approaches, such as Tex2Shape [16], operate in the UV space of the SMPL model. They use an input image to infer detailed normal or displacement maps, which are then applied to the base mesh to add high-frequency details such as wrinkles and fabric folds.

- **Complete end-to-end pipelines:** Building on these foundational techniques, other research focuses on creating complete, automated pipelines designed to produce a fully animatable and usable avatar. For example, Beacco et al. [17] present a workflow that integrates SMPL fitting, silhouette warping, and, crucially, a separate high-fidelity head reconstruction module. The resulting head mesh is then stitched onto the reconstructed body, creating a final avatar that prioritizes not just geometric accuracy but also the user’s recognizability and immediate usability in VR applications. **Figure 4** shows a graphical representation of the method presented in [17].

Figure 4.
Complete pipeline of the avatar reconstruction method presented in [17]. The method runs several deep learning techniques over a frontal RGB image of a single person such as region segmentation, pose detection, normal and depth maps, head reconstruction, and combines them to automatically retrieve a fully usable and recognizable 3D animated avatar.

Despite these advances, the fundamental limitation of all single-view methods is that they produce "2.5D" avatars, where any occluded part is either missing or inferred from visible ones.

3.3 Beyond fixed topology: Implicit function-based reconstruction

To overcome the constraints of fixed-topology models, a significant breakthrough came from methods using implicit neural representations. Instead of defining a surface with a discrete mesh, these approaches learn a continuous function that models the 3D shape. A seminal work in this area is PIFu [18], which learns a function that can infer whether a 3D point is inside or outside a surface based on its corresponding 2D image features, allowing for the reconstruction of arbitrary topologies. However, these methods still suffer from monocular ambiguity. To improve robustness, hybrid approaches such as PaMIR [19] condition the implicit function on an SMPL prior, combining the best of both worlds.

3.4 From single-view ambiguity to volumetric shape: Temporal and multiview fusion

To achieve a truly volumetric, 360-degree model, information from multiple viewpoints is essential.

- **Multiview photogrammetry:** The classic approach uses synchronized camera arrays to capture a subject from all angles, generating highly detailed meshes via structure-from-motion and multiview stereo.

- **Video-based refinement:** A more accessible approach uses a single monocular video. As a person moves, they naturally reveal different aspects of their body. Representative systems -such as the one proposed in [20] and related methods- initialize a model from one frame and iteratively refine it, gradually "filling in" the missing geometry and texture to create a fully volumetric model.

3.5 The rendering revolution: Neural volumetric representations

The most recent paradigm shift concerns how the final avatar is represented and rendered, moving toward neural techniques that offer unprecedented realism.

- **Neural radiance fields (NeRFs):** A NeRF [21] is an implicit representation that uses a neural network to encode a scene's color and density. Methods such as HumanNeRF can create stunningly photorealistic, deformable human avatars, though they are often slow to render.

- **3D Gaussian splatting:** To solve the speed problem, 3D Gaussian splatting [22] has emerged as a revolutionary alternative. This explicit method represents a scene as millions of tiny 3D Gaussians, achieving NeRF-level quality at real-time frame rates, making it a leading technology for the metaverse.

The rapid progress toward perfect realism makes the central thesis of this chapter – the link between realism and trust verification – more urgent than ever.

3.6 From technical breakthroughs to persistent multiuser systems

Realistic reconstruction is necessary but not sufficient for live, shared worlds. Persistent ecosystems require (i) identity continuity across sessions and worlds, (ii) asset portability and entitlements that survive context switches, (iii) networked animation with low-latency state sync, and (iv) defenses against impersonation that operate end-to-end. In practice, this couples avatar tech to identity and permissions: The same pipelines that reconstruct a face or gait must bind to verifiable credentials and runtime trust signals so that "Who controls this avatar?" is answered *continuously* as scenes, servers, and devices change.

4. The trust paradigm: The challenge of verification in anonymous environments

Having established the technological pipeline for creating realistic avatars, we now confront the profound social and security crises this very technology engenders. This section deconstructs the paradigm of trust, analyzing how physical-world mechanisms break down in the metaverse and defining a threat model for embodied identity.

4.1 The digital identity problem: From physical anchors to abstract presence

In the physical world, trust is built upon verifiable identity "primitives" that are difficult to forge: government-issued ID cards, biometric signatures, and, most fundamentally, the immutable reality of physical copresence. These mechanisms underpin nearly all high-value interactions [23]. The metaverse dissolves these anchors. Digital identity becomes malleable, an abstract data stream separate from the physical self. The challenge is not merely authenticating a user at login but ensuring the continuous, unbroken link between the human operator and the stream of photons and data that constitutes their avatar's presence.

4.2 A threat model for embodied identity

The primary threat is not simple anonymity but active, malicious impersonation enabled by the same AI tools used for reconstruction. Such risks were recently examined in an impersonation experiment published in [24]. We can define a threat model with several key attack vectors:

- **Identity spoofing and deepfake impersonation:** This involves creating a photorealistic avatar of a known, trusted individual (a CEO, a politician, or a family member) to commit fraud, spread disinformation, or gain unauthorized access. *Impact:* This is a direct attack on social and economic trust structures. In the context of presence, it is designed to exploit the victim's preexisting mental model of the person being impersonated.

- **Real-time avatar puppeteering:** A more sophisticated attack involves an actor modulating their own voice and using their movements to animate a victim's stolen or replicated avatar in real time. *Impact:* This is a direct assault on the Psi. An interaction with a "puppeted" avatar, even if subtly flawed, injects an element of the uncanny and unreal, shattering the victim's sense of presence and replacing it with anxiety.

- **Proteus effect weaponization:** Malicious actors can exploit the proteus effect [10] by choosing avatars designed to psychologically manipulate others. An avatar appearing as a doctor could elicit unwarranted trust and disclosure of private medical information; an avatar designed to appear vulnerable could be used for large-scale social engineering scams. *Impact:* This attack vector weaponizes social psychology, using the avatar's appearance as a tool to alter social dynamics and manipulate behavior, making the deception more effective.

- **Cross-platform vulnerabilities.** A further risk is *identity spillover*: Compromises in one ecosystem (e.g., a cloned avatar or leaked credential) can be replayed elsewhere if platforms accept screenshots, recordings, or unverifiable tokens as proof. Without portable cryptographic proofs and liveness checks, attackers can bootstrap trust in World B using artifacts harvested from World A. Mitigation requires cross-platform verification primitives and revocation mechanisms that travel with the user, not with a single vendor.

4.3 The perception dilemma: The failure of human intuition

Herein lies the central paradox: the technologies that allow us to cross the uncanny valley and achieve true embodied presence are the same ones that give adversaries the perfect toolkit to subvert it. This creates a technological arms race, but it is an asymmetric one. Human perception, honed for the physical world, is ill-equipped to detect sophisticated digital forgeries [5]. Relying on users to spot subtle artifacts in a real-time deepfake is a strategy doomed to fail, as it is often computationally easier to generate a convincing fake than it is to detect one definitively.

4.4 The architectural imperative: Toward a verifiable digital identity fabric

For the metaverse to mature beyond entertainment, it requires a new digital identity infrastructure – a "fabric" woven into its very foundation. The

"Digital Passport" analogy is useful, but its implementation must be dynamic, privacy-preserving, and user-centric. This points toward architectures based on self-sovereign identity (SSI). The key components of such a fabric are as follows:

- **Decentralized identifiers (DIDs):** Unique, persistent identifiers that are owned and controlled by the user – not a corporation or government. They are the anchor of identity in a decentralized world [25].

- **Verifiable credentials (VCs):** Tamper-proof, cryptographically signed claims about a user (e.g., "is over 18," "is a verified employee of Company X") that can be presented without revealing unnecessary data.

- **Zero-knowledge proofs (ZKPs):** A powerful cryptographic technique that allows a user to prove a statement is true without revealing the information underlying the proof. For example, a user could prove they are on a list of verified employees without revealing which employee they are, perfectly balancing verification with privacy.

A paradigmatic and widely adopted application of these principles in the Web3 ecosystem is the use of nonfungible tokens (NFTs) as a form of VC for digital assets, including avatars. An NFT, typically based on standards such as ERC-721 on the Ethereum blockchain, acts as a unique, publicly verifiable, and user-owned certificate of ownership. This enables a powerful verification mechanism known as "token-gating," where access to a digital resource (like using a specific avatar in a game) is granted only if the user can prove they own the corresponding NFT (note: proves ownership/usage rights of avatar assets, but not human personhood).

A practical implementation of this workflow is demonstrated in [26], which details a system where the user's Ethereum wallet address is used to query the blockchain in real time. The system verifies if the user is the registered owner of the ERC-721 token associated with a 3D avatar, whose metadata and model are stored decentrally on IPFS. This approach serves as a tangible example of a "Digital Passport" for avatars, grounding the abstract concepts of SSI in a functional, implemented solution. It is important to note, however, that these cryptographic solutions are not a panacea; they introduce their own challenges related to key custody, wallet security, phishing vulnerabilities, and the complex issue of credential revocation.

These architectural primitives are the necessary building blocks for the continuous, AI-driven verification framework we propose in the next section.

5. AI-powered trust-verification architectures

To resolve the stated dilemma, we propose the AI-CTVF. This framework does not rely on a single point of verification but on a "defense-in-depth" model with multiple continuously operating layers to establish a dynamic trust level for the user's identity.

5.1 A multilayered model for continuous identity verification

1. **Layer 1: Active and passive biometric verification**

This layer establishes the initial identity and monitors it.

- *Active verification (onboarding):* The initial enrollment process can utilize robust biometric methods, such as 3D facial scanning with *liveness detection* (to prevent photo or video spoofing) or voice authentication.

- *Passive verification (continuous):* During the session, AI noninvasively monitors biometric data. For instance, it can continuously analyze the user's voice biometrics for anomalous changes or verify subtle, unconscious patterns such as blink rate or head micro-movements, which are difficult to forge in real time.

2. **Layer 2: AI-powered behavioral analysis (behavioral signature)**

This is perhaps the most innovative and difficult-to-spoof layer. AI can learn a unique "behavioral signature" for each user, composed of idiosyncratic patterns.

- *Behavioral biometrics:* This includes gesture patterns, speech cadence, eye movement patterns when scanning a scene, and even the way a user interacts with virtual objects.

- *Coherence analysis:* An AI model can detect inconsistencies between the expected behavior of the legitimate user and the observed behavior, such as unusual hesitation or a movement pattern that does not match historical data. An impostor can replicate a face, but replicating a decade of learned behavioral habits is computationally infeasible in real time.

3. **Layer 3: Cryptographic and decentralized verification**

This layer protects user privacy and ensures identity portability.

- *DIDs:* The user's identity would not be controlled by a single corporation but by the user themselves via a blockchain-based system, giving them control over their data [25].

- *ZKPs:* This powerful cryptographic tool allows a user to prove a fact without revealing the underlying information. For example, a user could prove their avatar is verified as "human, over 18" without revealing their name, face, or birth date, thus preserving anonymity when desired.

- *Verifiable ownership via NFTs:* A direct and practical implementation of this layer involves linking the avatar asset to a NFT. This "token-gating" architecture, as demonstrated in systems built with the Ethereum blockchain [26], ensures that only the cryptographic owner of the NFT can use the associated avatar, providing a robust, on-chain verification of rights without a central authority.

4. **Layer 4: Perceptual watermarking and digital forensics**

This final layer adds a seal of authenticity directly into the avatar's data stream.

- *Perceptual watermarking:* An imperceptible signal, detectable by an algorithm but not by humans, can be embedded in the avatar's video or audio stream. The absence or alteration of this watermark would indicate that the avatar has been tampered with or is not from a verified source. However, the robustness of such watermarks against transformations such as video re-compression or adversarial attacks remains an active area of research.

- *Digital forensics:* In the event of a dispute, cryptographic audit logs—together with data from the system's earlier layers—enable forensic verification of whether the session was legitimate or spoofed.

5. **Layer 5: Interoperability and portability**

Trust must be portable across devices and worlds. This layer ensures that identities, credentials, watermarks, and audit proofs are *standards-aligned* and verifiable beyond a single vendor stack. Practically, this means: (a) binding avatar controllers to DIDs/VCs that any conformant verifier can check; (b) shipping trust signals (e.g., liveness attestations, watermark hashes) as signed, privacy-preserving metadata alongside media streams; and (c) defining revocation and key-rotation flows that work across organizations and jurisdictions.

Figure 5 summarizes the AI-CTVF layered trust-verification architecture described above. The model follows a defense-in-depth approach: The upper layers focus on continuous authentication through direct human signals, while the lower layers provide cryptographic assurances and forensic traceability.

5.2 The role of AI in real-time anomaly and deepfake detection

Artificial intelligence is the engine that drives the AI-CTVF. It acts as a real-time anomaly detection system, constantly comparing incoming data streams (biometric, behavioral) against the established trust profile for the user. If a significant deviation is detected in one or more layers, the avatar's "trust level" could be visibly downgraded (e.g., its verification badge changes from green to yellow), alerting other users to a potential risk.

5.3 Implementation notes: Scaling to consumer devices

To run continuously on consumer hardware, AI-CTVF emphasizes: (i) on-device inference for sensitive biometrics, (ii) model compression and quantization for low-latency behavioral models, (iii) adaptive sampling (graceful degradation under load), and (iv) edge offloading for heavy cryptographic proof generation with end-to-end encryption. Together, these choices keep verification responsive without sacrificing privacy or battery life.

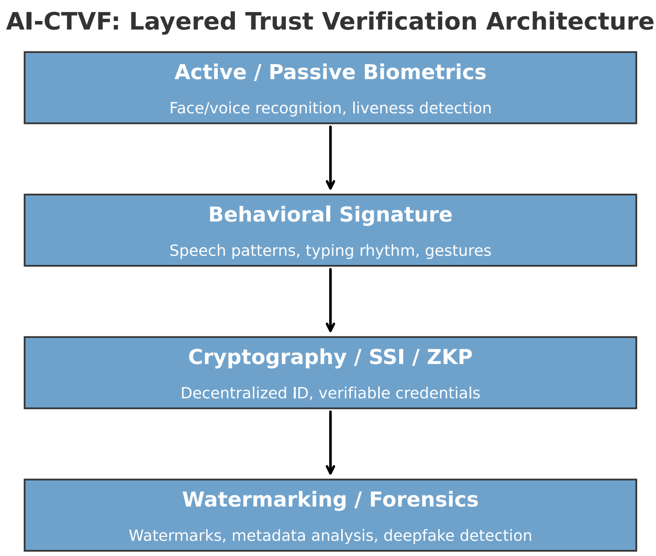

Figure 5.
AI-CTVF: layered trust-verification architecture. The framework integrates (a) active/passive biometrics *– such as face/voice recognition and liveness detection; (b)* behavioral signature *– for example, speech patterns, typing rhythm, and gestures; (c)* cryptography/SSI/ZKP *– decentralized identity, VCs, and ZKPs; and (d)* watermarking/forensics *– watermark embedding, metadata analysis, and deepfake detection. Together, these layers ensure robust, multimodal verification of trusted embodied presence in virtual environments.*

6. Transdisciplinary implications and ethical considerations

The implementation of systems like the AI-CTVF transcends engineering and has profound societal consequences.

6.1 Psychology

The existence of a visible "trust score" could fundamentally alter interpersonal perception. How do we interact with someone whose avatar is marked as "partially verified"? This could influence judgment formation, trust, and self-presentation. Furthermore, the pressure of being continuously monitored, even passively, could generate anxiety or induce less spontaneous behavior.

6.2 Sociology

These systems could create new forms of social stratification: a "verified class" with access to high-value interactions and an "anonymous class" relegated to less secure spaces. This could exacerbate existing digital divides, where those without access to the necessary technology for verification are marginalized from the metaverse economy and society.

6.3 Law and regulation

Complex legal questions arise. Who is legally responsible for the actions of a deepfake avatar: the software creator, the platform that hosted it, or the individual who operated it? A legal framework for digital identity is urgently needed to define the status of a "verified avatar" and establish liability in cases of fraud or impersonation. Jurisdiction in a global, decentralized metaverse is another monumental challenge.

6.4 Ethics

The tension between security and privacy is central. To what extent is it ethical to require biometric and behavioral monitoring to participate in digital society? It is crucial to ensure that verification algorithms do not contain biases that discriminate against certain demographic groups. For example, biometric systems are known to exhibit biases based on skin tone or age, and behavioral analysis could unfairly penalize users with atypical movement patterns, a phenomenon known as "biometric drift." The ownership and control of biometric and behavioral data – arguably the most intimate information we possess – must remain firmly in the hands of the user, as proposed by decentralized identity architectures. The right to anonymity, vital for freedom of expression and the protection of dissidents, must be carefully balanced with the need for security.

6.5 Governance models for trust infrastructure

Three governance archetypes emerge, each with trade-offs:

- **Regulatory (public-sector-led):** Clear accountability and due process; risks include slower iteration and over-centralization.
- **Consortium-led (industry standards):** Faster alignment with practical needs; risks include capture and uneven user protections across vendors.
- **Decentralized (SSI/DAO-aligned):** User-controlled keys and credentials; risks include key management burdens, fragmented assurance levels, and complex liabilities.

A polycentric approach – public baselines for rights and redress, consortium standards for interoperability, and SSI for user agency – can balance safety, innovation, and autonomy.

7. Case studies, applications, and future directions

7.1 Case study 1 (high stakes): Telemedicine in the metaverse

A surgeon, whose avatar is verified by the AI-CTVF, remotely guides a local medical team through a complex operation. Continuous verification assures the hospital and the patient that the guidance is from the legitimate expert. The cryptographic layer (ZKP) allows the surgeon to prove their medical license without revealing any other personal data.

7.2 Case study 2 (commerce): International contract execution

Two executives from different continents meet in a virtual boardroom to sign a merger agreement. Their photorealistic avatars display a “fully verified” status. The contract signing is executed via a cryptographic key tied to their verified identities, creating a legally binding and nonrepudiable record.

7.3 Case study 3 (social): Safe and moderated communities

A social platform allows users to choose to interact only with others who meet a certain verification level (e.g., “verified as human”). This could drastically reduce harassment, bots, and spam, creating safer spaces, especially for vulnerable users, without completely eliminating the option for anonymity in other areas of the platform.

7.4 Case study 4 (civic): Verified avatars in e-governance

Citizens access municipal services in a virtual town hall. Their avatars present a ZKP that they are residents and over 18, without revealing their names or addresses. A council hearing permits only “human-verified” avatars to speak, mitigating bot-driven disruption while preserving anonymity for whistleblowers. Audit trails – watermarked media plus cryptographic session logs – create accountable, appealable records compatible with public-sector retention policies.

7.5 Future directions: Government-issued avatars as digital identity certificates

The logical culmination of the trust-verification paradigm is the evolution of avatars from mere representations to legally recognized digital identity instruments, analogous to physical passports or digital certificates. In this future state, we envision a system where trusted entities, potentially government-certified or acting as sovereign identity providers, are responsible for creating and issuing high-fidelity, verified avatars.

The process would mirror the issuance of a passport:

1. A citizen would undergo a robust in-person identity verification process at an authorized office.

2. During this process, a photorealistic, biometrically accurate 3D scan of the individual will be captured, creating a “base” avatar.

3. This verified avatar would then be cryptographically signed and minted as a nontransferable soulbound token or a specialized NFT, and delivered to the citizen’s secure digital wallet.

This “Sovereign-Issued Avatar” would become the root of trust for all high-stakes interactions in the metaverse. It would be universally verifiable across different virtual worlds, allowing citizens to sign binding contracts, access government e-services, conduct secure financial transactions, or give legal testimony with the same level of legal certainty as if they were physically present or using a national digital certificate. This framework would solve the identity problem at its core, creating the foundational trust layer necessary for the metaverse to become a genuine extension of our civic and economic lives.

7.6 Open research avenues and pending challenges

The path to implementing the AI-CTVF is fraught with challenges:

- **Computational overhead:** Continuous analysis of multiple real-time data streams requires significant computational power.
- **Interoperability:** To avoid "walled garden" virtual worlds, open standards for identity and verification are needed that work across different platforms.
- **User acceptance:** Users must perceive the benefits of these systems as outweighing their concerns about privacy and surveillance.
- **System robustness:** The security of the verification system itself is paramount, as it would become a high-value target for attackers.

8. Conclusion

8.1 Recapitulation: The inseparable symbiosis of realistic presence and verified trust

This chapter has argued that the pursuit of embodied presence through realistic avatars in the metaverse is an incomplete and dangerous endeavor if detached from the simultaneous construction of a trust infrastructure. Realism without verification leads not to immersion but to an era of "suspect reality," where every interaction is fraught with uncertainty. Authentic presence can only flourish on the fertile ground of guaranteed safety and identity.

8.2 AI as the "system" of the metaverse

Artificial intelligence reveals itself as a double-edged sword. It is the tool that generates both the problem (*deepfakes*, impersonation) and the solution. In proposing the AI-CTVF, we have repositioned AI not only as an architect of virtual worlds but as their essential, adaptive, continuously learning system that identifies and neutralizes threats to identity, allowing the social organism of the metaverse to function healthily.

8.3 Final reflections on building safe and trustworthy infinite worlds

The construction of the metaverse is not merely a technical challenge but a profound social, ethical, and philosophical one. The choices we make now about the architecture of identity and trust will define the nature of these "infinite worlds" for decades to come. By integrating verification into the core design of the metaverse, rather than adding it as a later patch, we have the opportunity to build the next chapter of the Internet's history, that is, from its inception, safer, more accountable, and ultimately more human.

8.3.1 Who should own and operate the trust?

We argue for *polycentric governance*: (i) individuals own their cryptographic identifiers and credentials; (ii) independent, auditable verifiers (public interest or accredited) operate the attestation services; (iii) platforms consume standardized proofs but cannot lock users in; and (iv) regulators set baseline guarantees (revocation, appeals, nondiscrimination). Trust then becomes an interoperable utility – like the web's PKI – rather than a proprietary moat.

Author details

Jaime Gallego Vila
Polytechnic University of Catalonia, Vilanova i la Geltru, Spain

*Address all correspondence to: jaume.gallego@upc.edu

References

[1] Ball M. The Metaverse: And How It Will Revolutionize Everything. New York: Liveright Publishing Corporation; 2022

[2] Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences. 2009;**364** (1535):3549–3557

[3] Mori M. The uncanny valley. Energy. 1970;**7**(4):33–35

[4] Hashimoto T, Hitramatsu S, Tsuji T, Kobayashi H. "Development of the face robot SAYA for rich facial expressions," In: *2006 SICE-ICASE International Joint Conference*. IEEE; 2006, pp. 5423–5428

[5] Chesney R, Citron D. Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review. 2019;**107**:1753–1819

[6] Skarbez R, "Plausibility Illusion in Virtual Environments," Ph.D. dissertation, University of North Carolina at Chapel Hill, Chapel Hill, 2016. [Ph.D. dissertation]. Available: https://cdr.lib.unc.edu/concern/dissertations/v118rf14z [Accessed: 2025-April-23]

[7] Kilteni K, Gonzalez R, Slater M. The sense of embodiment in virtual reality. In: Presence: Teleoperators and Virtual Environments. **21**. Cambridge, USA: MIT Press; 2012. p. 373–387

[8] Biocca F. The cyborg's dilemma: Progressive embodiment in virtual environments. Journal of Computer-Mediated Communication. 1997;**3**(2). DOI: 10.1111/j.1083-6101.1997.tb00070.x

[9] Sanchez-Vives M, Spanlang B, Frisoli A, Bergamasco M, Slater M. Virtual hand illusion induced by visuomotor correlations. The New Science of Consciousness: States of Mind. 2010;**5**(2):e10381

[10] Yee N, Bailenson JN. The proteus effect: The effect of transformed self-representation on behavior. Human Communication Research. 2007;**33** (3):271–290

[11] Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ. SMPL: A skinned multi-person linear model. ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia). 2015;**34**(6):248:1–248

[12] Pavlakos G, Choutas V, Ghorbani N, Bolkart T, Osman AAA, Tzionas D, Black MJ. "Expressive body capture: 3d hands, face, and body from a single image." In: *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2019

[13] Bogo F, Kanazawa A, Lassner C, Gehler P, Romero J, Black MJ. "Keep it SMPL: Automatic estimation of 3d human pose and shape from a single image." In: *European Conference on Computer Vision*. Springer; 2016, pp. 561–578

[14] Kanazawa A, Black MJ, Jacobs DW, Malik J. "End-to-end recovery of human shape and pose." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7122–7131

[15] Weng C-Y, Curless B, Kemelmacher-Shlizerman I. “Photo wake-up: 3d character animation from a single photo.” In: *IEEE Proc. Conf. on Computer Vision and Pattern Recognition*. 2019, pp. 5908–5917

[16] Alldieck T, Pons-Moll G, Theobalt C, Magnor M. “Tex2shape: Detailed full human body geometry from a single image.” In: *IEEE International Conference on Computer Vision (ICCV)*. IEEE; 2019

[17] Beacco A, Gallego J, Slater M. “Automatic 3d avatar generation from a single RGB frontal image.” In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE; 2022, pp. 764–765

[18] Saito S, Huang Z, Natsume R, Morishima S, Kanazawa A, Li H. “PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization.” In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2019

[19] Zheng Z, Yu T, Liu Y, Dai Q. “Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction.” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. 2021

[20] Alldieck T, Magnor M, Xu W, Theobalt C, Pons-Moll G. “Video based reconstruction of 3d people models.” In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8387–8397

[21] Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM. 2021;**65**(1):99–106

[22] Kerbl B, Kopanas G, Leimkühler T, Drettakis G. 3d Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics. 2023;**42** (4):139–1

[23] Preukschat A, Reed D. Self-sovereign Identity. New York: Manning Publications; 2021

[24] Oliva R, Wiesing M, Gallego J, Inami M, Interrante V, Lecuyer A, McDonnell R, Nouviale F, Pan X, Steinicke F, Slater M, “Where extended reality and AI may take us: Ethical issues of impersonation and AI fakes in social virtual reality.” 2025. Available: 10.5281/zenodo.15413114

[25] W3C. “Decentralized identifiers (dids) v1.0.” W3C Recommendation. 2022. [Online]. Available: https://www.w3.org/TR/did-core/ [Accessed: 2025-March-3]

[26] Sanchez LI. “Estudi Sobre la Verificació D’identitat D’avatars 3d En Entorns de Realitat Virtual Utilitzant Nft I Blockchain.” Master’s thesis, Universitat Politècnica de Catalunya (UPC), Escola Superior d’Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa, 2025, master’s Thesis