Human Verification: Decentralized Social Coordination

Trevor Thompson, Ben Walther, Sean Waters, Seth Gholson

Ethos Network

Version 0.2 — February 2026

https://ethos.network

Abstract

The ability to create unlimited pseudonymous identities is both a feature and a vulnerability of digital systems. As artificial intelligence capabilities advance, distinguishing humans from sophisticated bots becomes increasingly difficult—and increasingly important. Existing proof-of-personhood solutions have approached this problem primarily through biometrics, requiring users to surrender intimate personal data to centralized entities. We propose an alternative approach that leverages human coordination around credibility assessment. Rather than asking "can a machine verify this is a human?" we ask "can trusted humans verify this is a human?" By utilizing Ethos's existing credibility infrastructure and validator ecosystem, we address the primary challenge of social-graph-based verification—bootstrapping a trusted network—while preserving privacy and avoiding dependence on specialized hardware or biometric data collection. This paper describes the mechanism design, economic incentives, accountability structures, and known limitations of Ethos Human Verification.

1. Introduction

The ability to create unlimited pseudonymous identities is both a feature and a vulnerability of digital systems. A single actor can spin up thousands of wallets, social accounts, and digital personas at near-zero cost. As AI capabilities advance, distinguishing humans from sophisticated bots becomes increasingly difficult—and increasingly important.

Existing proof-of-personhood solutions have approached this problem primarily through biometrics: iris scans, facial recognition, and other physical markers. While technically sophisticated, these approaches require surrendering intimate personal data to centralized entities and trusting that data will be stored, processed, and governed responsibly. This represents a significant privacy tradeoff and creates honeypots of sensitive information.

We believe there is a better way.

2. The Ethos Approach

Ethos Human Verification leverages what the network already does well: **human coordination around credibility assessment**. Rather than asking "can a machine verify this is a human?" we ask "can trusted humans verify this is a human?"

This approach has precedent. As Buterin noted in his analysis of proof-of-personhood systems [1], social-graph-based verification offers distinct advantages: no specialized hardware, no biometric data collection, and the potential for greater privacy preservation. The primary challenge he identified—bootstrapping a trusted network from scratch—is one Ethos has already solved through its existing credibility infrastructure and validator ecosystem.

2.1 Design Philosophy

Current approaches to human verification—whether biometric, institutional, or behavioral—share a common flaw: they require substantial trust in a centralized issuer. Users must trust that biometric data is handled responsibly, that verification algorithms are unbiased, and that the issuing entity will persist and remain trustworthy.

We believe this trust should be distributed. Just as Wikipedia demonstrated that human coordination could produce reliable information without central editorial control, we believe human coordination can produce reliable identity verification without central biometric databases.

This is not a new idea. It is how trust has worked throughout human history—through vouching, reputation, and social networks. We are simply bringing these mechanisms onchain with appropriate incentive structures.

3. Mechanism Design

3.1 Requesting Verification

Any Ethos user may request human verification. This signals their intent to be verified and makes them visible to validators who can perform the verification.

3.2 Validator Verification

Owners of Validator NFTs (a fixed supply of 256) may verify a user's humanity through one of three methods: **in-person meeting**, where the validator has met the user face-to-face; **video call**, where the validator has conducted a live video conversation; or **voice call**, where the validator has spoken with the user by phone.

These methods are ordered by signal strength. In-person verification provides the highest confidence, while voice-only provides the lowest. Validators exercise their own judgment in determining sufficient evidence of humanity.

3.3 The Human Bond

Upon verification, a bond is created using the Ethos Bond smart contract. This bond links the validator's reputation to the verified user's humanity status, persists for three years before requiring renewal, is recorded entirely onchain, and can be slashed if the verification is later proven fraudulent.

3.4 Achieving Verified Status

A user achieves "Human Verified" status upon receiving verifications from **three unique validators**. This threshold provides redundancy against individual validator error or compromise while remaining accessible to legitimate users.

Upon verification, the user receives a Human Verified badge visible across all Ethos integrations, a +100 credibility score bonus, and access to features and benefits restricted to verified humans.

4. Validator Requirements

Not all Ethos users may verify others. To act as a human verifier, one must own a Validator NFT (one of the 256 fixed-supply governance tokens), be Human Verified themselves, maintain a credibility score above 1600 demonstrating sustained ethical behavior, and have available Verification slots distributed by protocol governance.

4.1 Why Validators?

Validators represent the most trusted participants in the Ethos network. They have been present since the protocol's earliest days, have demonstrated commitment through financial and social investment, and have the most to lose from protocol degradation.

By restricting human verification to this group, we ensure that verifiers have significant reputation at stake, that the verification supply is naturally rate-limited, that bad actors cannot easily infiltrate the verifier pool, and that accountability is traceable to known, high-reputation entities.

4.2 Verification Allocation

Validators do not have unlimited verification capacity. Each validator receives a finite number of "Verification slots" that are distributed by protocol governance (initially the Ethos team, transitioning to community governance over time). Once a validator exhausts their slots, they must wait for additional allocation.

This mechanism prevents rapid sybil creation even by compromised validators, creates scarcity that encourages careful verification decisions, and allows the protocol to scale verification capacity deliberately.

4.3 Future Verifier Expansion

While Human Verification is currently restricted to Ethos Validators, this constraint is not permanent. As the protocol matures and demand for verification exceeds validator capacity, we may extend verification privileges to other high-reputation users who meet stringent criteria—for example, users above a certain credibility threshold who have maintained good standing over an extended period.

This expansion is not guaranteed; it depends on observed behavior, attack patterns, and the protocol's ability to maintain verification quality at scale. The validator-only model provides a controlled environment to refine the mechanism before potentially opening it to a broader set of verifiers. Any expansion would be implemented gradually with appropriate safeguards and subject to governance approval.

5. Economics

Human verification operates within Ethos's XP economy, creating aligned incentives for both users seeking verification and validators providing it.

5.1 Verification Cost

Requesting human verification requires **10,000 Ethos XP**, collected at the time of request. This cost serves multiple purposes: it provides sybil resistance by ensuring attackers cannot create verified identities without first accumulating meaningful XP through legitimate protocol participation; it signals commitment by demonstrating that users requesting verification have invested in the Ethos ecosystem; and it creates sustainable incentives by funding validator compensation without requiring external subsidy.

5.2 Escrow Model

Upon requesting verification, the 10,000 XP is collected: 1,000 XP is immediately burned and 9,000 XP is held in escrow. Validators are compensated incrementally as verifications occur. As each of the three validators verifies the user, 3,000 XP is released to each validator immediately upon their verification.

Event	XP Movement
User requests verification	9,000 XP escrowed; 1,000 XP burned
First validator verifies	3,000 XP released
Second validator verifies	3,000 XP released
Third validator verifies	3,000 XP released

Verification requests remain open indefinitely until completed. There is no expiration or timeout. This ensures users in regions with limited validator access are not penalized for slower verification timelines.

5.3 Validator Compensation

Validators receive 3,000 XP immediately upon each successful verification, regardless of whether the user ultimately achieves full Human Verified status. This means validators are compensated for their time even if a user abandons the process, users who receive partial verification (one or two validators) have remaining XP held in escrow until completion, and there is no refund mechanism for incomplete verifications.

5.4 Economic Safeguards

The XP incentive is deliberately modest relative to slashing penalties. A validator who verifies fraudulently risks 800 credibility score loss (substantially more impactful than 3,000 XP gain), permanent loss of verification privileges (forfeiting all future verification income), and reputational damage within the Ethos community.

The asymmetry between reward (3,000 XP) and penalty (800 credibility plus permanent verification ban) is intentional. Verification should be worthwhile for honest validators but economically irrational for those considering fraud.

5.5 Future Development

The use of XP as the economic mechanism is a temporary measure for the initial implementation. XP provides a low-friction way to bootstrap incentives while the protocol matures, but it is not the intended long-term model. As Human Verification gains adoption and the protocol stabilizes, we anticipate transitioning to actual capital—ETH or stablecoins—for both verification fees and validator compensation. This would strengthen economic incentives, increase the cost of attacks, and create sustainable revenue for validators who provide high-quality verification services.

As the protocol matures and we observe real-world usage patterns, we also anticipate adjustments to verification pricing, validator compensation structures, and burn ratios. Parameters will be governed through standard Ethos governance processes as the protocol decentralizes.

6. Accountability: Slashing

Human verification bonds are slashable. If a validator verifies an entity later proven to be non-human (a bot, a duplicate identity, or a coordinated sybil), the bond may be slashed through the standard Ethos slashing mechanism.

6.1 Proving Non-Humanity

A natural question arises: how does one prove that a verified user is *not* human? The answer lies in Ethos's existing social consensus mechanism for slashing, which has been battle-tested across other protocol functions.

Any Ethos user with sufficient standing may initiate a slash proposal against a Human Verification bond by presenting verifiable evidence that the verified entity is non-human. This evidence might include proof of duplicate identities (the same person verified under multiple profiles), evidence of bot behavior (automated posting patterns, impossible activity timestamps, or coordination with known bot networks), admissions or public statements revealing non-human status, or technical evidence linking the profile to known sybil operations.

The proposer must stake their own credibility as a bond, creating skin in the game that discourages frivolous or malicious slash attempts.

6.2 Social Consensus Resolution

Once a slash is proposed, the resolution follows Ethos's established social consensus process. Reputable users—weighted by their credibility scores—review the evidence and vote on the outcome. This is not a popularity contest; voters are expected to evaluate the evidence objectively, and the credibility-weighted voting ensures that users with proven track records have greater influence on outcomes.

If the slash is approved, the verifying validator suffers the penalties described below. If the slash is rejected, the proposer loses their staked credibility, discouraging bad-faith accusations. This mechanism has proven effective for other slashing contexts within Ethos and provides a decentralized, transparent process for adjudicating disputes about humanity status.

6.3 Slashing Penalties

Human Verification Slashing carries enhanced penalties: the verifying validator loses **800 credibility score** per fraudulent verification, the validator **permanently loses the ability to verify humans**, and standard slashing procedures apply to the bond itself.

These severe penalties reflect the seriousness of the trust being extended. Verifying a non-human pollutes the verified pool and damages the credential's value for all legitimate participants.

7. Vision

7.1 Universal Human Verification

Human Verification is designed to verify *any* human—not just existing crypto users. We are committed to making the verification process accessible to non-crypto-native individuals and to integrating human verification wherever identity matters.

We envision a future where platforms can confidently surface content from verified humans, where airdrops and rewards flow to real people rather than bot armies, where democratic governance mechanisms have meaningful sybil resistance, and where the "Human Verified" credential is recognized and valued across the internet.

7.2 Why This Matters

We believe that companies will want to target specific users who are human, elevate them in their platforms, and ensure that they are rightfully rewarded. As AI-generated content and autonomous agents proliferate, the ability to credibly signal humanity becomes increasingly valuable.

8. Known Limitations and Attack Vectors

We present this mechanism with intellectual honesty about its current limitations and potential attacks.

8.1 Collusion Risk

A group of validators could coordinate to verify sybil identities. The three-validator requirement increases coordination costs, slashing penalties create mutual destruction incentives, validators risk their own permanent verification rights, and pattern analysis can detect suspicious verification clusters. However, with 256 validators and over 2.7 million possible combinations of three, detecting collusion requires sophisticated analysis.

8.2 Validator Compromise

A validator could be compromised through account theft, coercion, or corruption. Limited verification slots cap damage from any single compromise, slashing provides retroactive accountability, the validator pool's size provides redundancy, and multi-validator requirements mean single compromises are insufficient to create verified sybils.

8.3 Social Engineering

Bad actors could deceive validators through sophisticated impersonation. We encourage in-person or video verification over voice-only, rely on validator discretion in choosing verification standards, develop community norms around verification rigor, and impose reputational consequences for validators who are frequently deceived.

8.4 Absence of Negative Signaling

In the current design, validators can only take positive action (verify) or no action (ignore a request). There is no mechanism for validators to flag a request as suspicious, warn other validators, or remove problematic requests from the verification queue.

This creates risk: verification requests remain open indefinitely, so a sophisticated attacker could submit numerous requests and wait for three careless, compromised, or colluding validators to verify each one. Without negative signaling, the community cannot efficiently filter obviously fraudulent requests.

Future protocol versions may introduce mechanisms for validators to flag suspicious requests, including reputation-weighted negative signals, threshold-based request removal, and slashing for users who submit provably fraudulent verification requests. This limitation is acknowledged as a known gap in the initial implementation.

8.5 Uniqueness Is Not Guaranteed

Human Verification confirms that a profile belongs to a human, but it does not guarantee that the human has only one verified profile. In principle, a single person could complete verification multiple times under different identities, obtaining multiple Human Verified badges.

This is a deliberate design choice. Ethos addresses uniqueness through complementary mechanisms rather than attempting to solve it at the verification layer. The Ethos Credibility Score provides the primary uniqueness signal: replicating a meaningful credibility score across multiple accounts requires sustained, genuine participation over time. An attacker attempting to maintain ten verified accounts with credible scores would need to build ten distinct social graphs, accumulate ten sets of vouches, and maintain ten histories of positive interactions—a substantial and ongoing cost.

Compared to the current state of crypto, where spinning up thousands of wallets is trivial and costless, Human Verification creates meaningful friction. We do not claim to eliminate sybils entirely; rather, we significantly raise the cost of operating them. This approach accepts that perfect uniqueness verification may be unachievable without invasive biometrics, and instead offers improvement via negativa—making the undesirable behavior substantially harder rather than impossible.

8.6 Bootstrapping Challenges

Initially, few validators may be geographically distributed enough to verify users in all regions. We acknowledge this as a growth constraint and are exploring mechanisms to expand validator coverage deliberately.

9. Related Work

Several projects have attempted to solve the proof-of-personhood problem with varying approaches.

Proof of Humanity [2] requires users to upload a video of themselves and obtain vouching from existing users, with disputes resolved through Kleros decentralized courts. While accessible (requiring only a smartphone), this approach publicly exposes user faces, creating significant privacy concerns.

BrightID [3] uses video call verification parties where users verify each other, combined with a trust graph (Bitu) for higher-level verification. This approach has bootstrapping challenges similar to other social-graph systems.

Worldcoin [4] uses specialized hardware (the Orb) to scan user irises, providing strong sybil resistance but requiring trust in hardware manufacturers and creating biometric data collection concerns.

Idena [5] uses synchronized captcha games to verify humanness, an innovative approach that avoids biometric collection but requires participation at specific times.

Ethos Human Verification differs from these approaches by leveraging an existing high-trust validator network, avoiding biometric data collection, and utilizing onchain reputation and economic bonds for accountability.

10. Conclusion

Ethos Human Verification represents a novel approach to human verification that prioritizes privacy, decentralization, and economic alignment over biometric data collection. By leveraging the existing trust network of Ethos validators and binding verifications to meaningful economic and reputational stakes, we create a system where honest behavior is rewarded and fraudulent behavior is costly.

The protocol is early in its development. The parameters described here—including bond duration, slashing penalties, credibility bonuses, and validator requirements—are subject to revision as we learn from real-world usage. We are committed to iterating transparently and incorporating community feedback as this mechanism matures.

We believe that through human coordination, we can solve the unique-human problem without surrendering our biometric data to centralized authorities. The technology exists. The incentives can be aligned. What remains is execution and iteration.

References

[1] V. Buterin, "What do I think about biometric proof of personhood?" *vitalik.eth.limo*, July 2023. https://vitalik.eth.limo/general/2023/07/24/biometric.html

[2] Proof of Humanity, "A system combining webs of trust with reverse Turing tests." https://proofofhumanity.id/

[3] BrightID, "A social identity network that allows people to prove that they're only using one account." https://www.brightid.org/

[4] Worldcoin Foundation, "Worldcoin Whitepaper." https://whitepaper.worldcoin.org/

[5] Idena, "Proof-of-Person blockchain." https://www.idena.io/

[6] Ethos Network, "Ethos Whitepaper." https://whitepaper.ethos.network/