

PRIVATE PROOF OF HUMAN: CRITICAL INFRASTRUCTURE FOR HUMANITY IN A WORLD WITH ADVANCED AI

Tools for Humanity

March 25, 2026

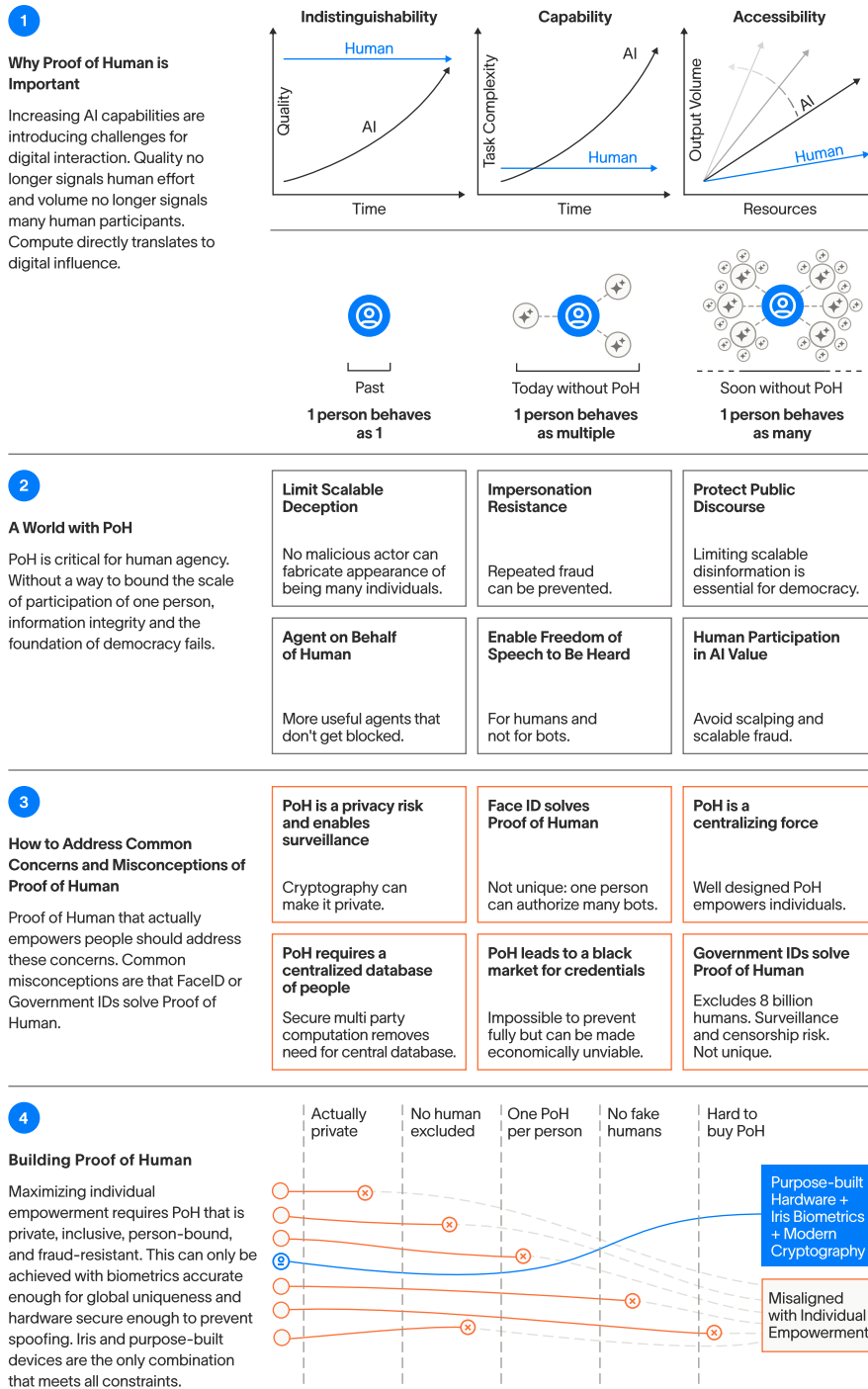
Key Takeaways

- Building Proof of Human is unexpectedly challenging.
- Government IDs aren't effective and create surveillance risks.
- FaceID doesn't confirm uniqueness.
- Without uniqueness, AI agents can outsource Proof of Human to “human farms” (think call centers) that fabricate human presence for AIs 24/7.
- Face-based uniqueness can be spoofed and isn't accurate, ultimately excluding billions from getting a Proof of Human.
- As a result, an anti-surveillance and effective solution that actually empowers people requires new technology.

Summary

The accelerating capabilities of AI agents create an existential challenge for the integrity of digital interaction and the stability of our society. AI-generated content and actions are becoming indistinguishable from human interaction, and automated messages become increasingly personalized and scalable. As a result, the ability to reliably distinguish humans from AI becomes critical. A world without private Proof of Human (PoH) risks mass disinformation, election manipulation, scalable fraud and privacy-invasive tracking, all of which seriously threaten the stability of democracy and human agency. At the same time, PoH protects freedom of speech by elevating human voices above bots and it empowers people through agents that are not blocked as bots but recognized to act on a human's behalf. PoH also prevents scalable disinformation and enables public input at unprecedented scale. This blogpost outlines the necessity for a globally inclusive, high-integrity, and privacy-preserving PoH mechanism that also makes it hard to delegate PoH credentials to malicious actors. Contrary to common perception, (well-implemented) PoH does not increase surveillance but protects against it because it preempts the need for privacy-invasive monitoring. It is a fundamental enabler for maintaining and increasing human agency, safeguarding public discourse, and robust benefits distribution (should they be needed) in a world with advanced AI. We detail a technical architecture based on purpose-built hardware and modern cryptography for anonymous verification of PoH. We advocate for scaling this infrastructure rapidly to prevent avoidable threats to democracy and avoid less effective, censorship-enabling, privacy-invasive measures that are incentive-misaligned with individual freedom in the long-term. World ID is an attempt at realizing this future.

Private Proof of Human:
Critical Infrastructure for Humanity in a World with Advanced AI



1. Why Proof of Human is Important

1.1. Capability Shift

AI leads to two trends that amplify each other:

Loss of reliable signals: Digital systems increasingly lack reliable ways to distinguish between human-participation and bots because AI systems make it possible to:

- Generate large volumes of text that mimic personal opinion, reflection, or experience.
- Present convincing representations of people across images, video, and voice.
- Take human-like actions on the internet, including navigation of web browsers.

Amplification without people: Increasing viability of deploying AI at scale makes it possible for a small number of actors to act as if they were many independent humans, at an increasingly large scale:

- The cost of producing high-quality synthetic output continues to fall.
- Model capabilities continue to improve and open source models keep pace with a time delay.
- Increasing task lengths become viable and increase time horizons over which models can be deployed

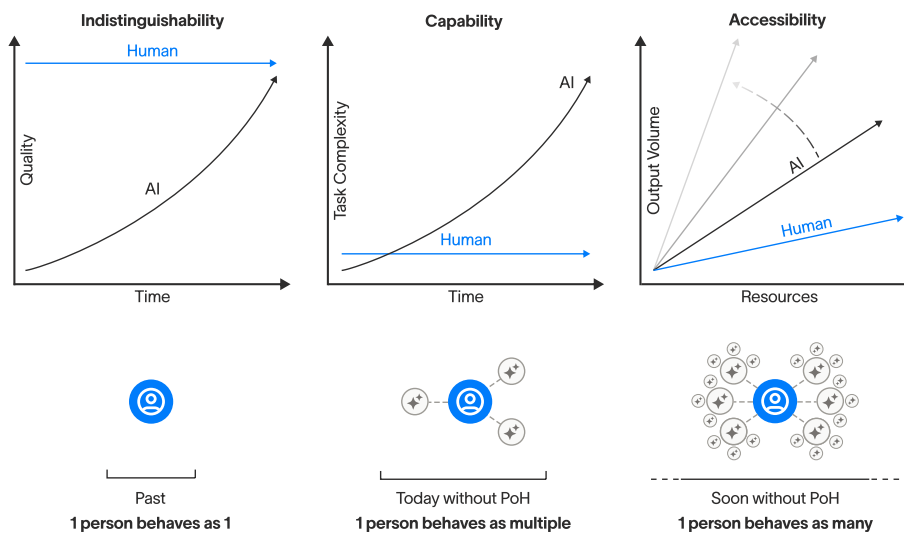


Figure 2. As AI output quality approaches human levels over time, traditional signals used to infer genuine human participation become less reliable. Additionally, AI is becoming increasingly better at completing tasks of longer length and higher complexity at lower cost. As a result, AI enables large-scale human-like output without proportional human input. Without PoH, over time it will be easier for one real human to behave as if they were many.

We've passed the Turing Test. Intelligence, digital social interaction and soon video streams are no longer reliable signals of humanness: AI systems can sustain coherent interaction, personalize messaging, and pursue objectives over long horizons. As a result, this enables amplification through multiplicity: A small number of bad actors can use AI to create a large number of seemingly distinct participants.

1.2. Failure Modes Without PoH

If current trends in AI continue, synthetic content will become ubiquitous. As a result, coordinated networks of agents can manufacture apparent support for or opposition to legislation or territorial

claims; seed and amplify conspiracy theories; and mobilize protests or movements by organizing participants and coordinating messaging.¹ In the extreme case, as bot activity on the internet increases, these dynamics can become a serious threat to democracy and contribute to election interference, abrupt shifts in political legitimacy, and regime destabilization, as has already occurred when online coordination translated directly into real-world governance outcomes.² Importantly, this does not require false or novel content; even amplification of existing content can meaningfully shift how it is perceived.³ The important property is not whether content is human or AI-generated, but rather the impression of how many people agree with it. As cost for AI capabilities is declining and automation enables creating the false impression of human consensus, influence and impersonation scale with compute rather than people.

As these techniques mature and become more cost effective, the set of actors capable of exploiting them expands beyond traditional state influence operations, for example:

- State and quasi-state actors using information warfare to influence or destabilize political systems domestically or abroad.
- Terrorist and extremist groups leveraging automated content and coordination to recruit, radicalize, and organize.
- Financially motivated actors including large-scale scam networks, advertising and review fraud operations.

As synthetic participation scales, digital systems no longer enable authentic human interaction. Omnipresence of human-looking AI agents makes it harder to surface human voices, and perceived consensus no longer represents actual people. At the same time, as technological and economic change accelerates, democratic systems need tighter feedback loops outside of elections to understand how people are affected and respond effectively. These same feedback channels are increasingly disrupted by large-scale automated participation, causing them to fail precisely when they are most needed. Similar failures affect companies and institutions that depend on large-scale human input to guide decisions and evaluate real-world impact. Free expression remains formally intact, but becomes ineffective as authentic participation is drowned out by bots and individual agency declines. In parallel, fraud and identity abuse expand with impersonation, phishing, benefits fraud, and identity theft becoming easier to automate and harder to detect, supported by synthetic media and forged facial and document-based credentials. Public support programs, including emergency and social-benefit systems, become vulnerable to large-scale exploitation, as demonstrated during recent crises in which fraudulent claims reached unprecedented levels as seen with COVID-19 relief funds.

The resulting erosion of reliable signals may contribute to reality apathy—where individuals increasingly discount information altogether because distinguishing truth from fabrication becomes impractical. Therefore, content authenticity is no longer sufficient; the critical factor becomes participation: who is acting and whether those actors correspond to real humans.

As these failures accumulate, the response will likely be to collect more personal information in an effort to distinguish real users from automated activity. This would likely lead to broader activity tracking, tighter identity verification requirements, and more intrusive presence checks. These

¹See this on how AI agents were deployed at scale to covertly and persuasively influence real users in online discussions on Reddit.

²See the 2025 Nepal case here

³“Bots almost exclusively retweeted original posts by Twitter users who are human, the scientists noted. In turn, many humans retweeted the bots messages that aligned with their political leanings, which then led to additional retweets and replies.”

measures, however, fail to solve the underlying problem and instead introduce new risks. Storing larger amounts of sensitive personal information makes data breaches more damaging and creates new opportunities for impersonation and identity abuse.

1.3. Why Other Defenses Do Not Scale

While existing approaches like content provenance, watermarking, automated detection, and frontier-model mediated monitoring are very valuable in specific contexts, none of them address the dominant failure modes described above. In particular, they do not meaningfully constrain large-scale, seemingly human activity manufactured by bot networks.

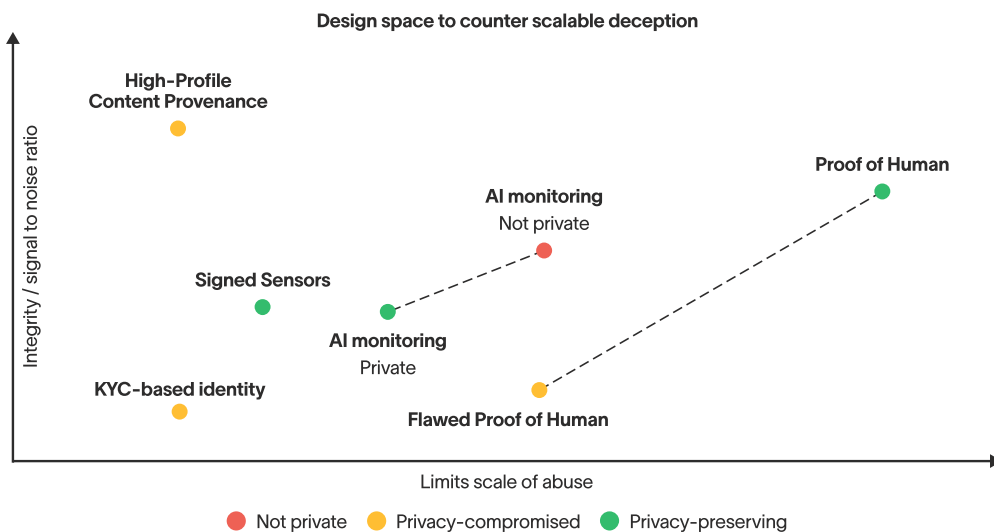


Figure 3. Various approaches to countering scalable deception and increasing authenticity come with different tradeoffs (green: privacy-preserving; orange: limited privacy; red: not private). Counter-intuitively, a well-implemented PoH system can not only address scalable deception but also indirectly increase authenticity.

Content provenance systems—such as signed camera streams and standards like C2PA—provide cryptographic metadata about the origin and modification history of digital artifacts. However, they do not attest to semantic accuracy and rely on trust assumptions that are difficult to maintain in adversarial environments.⁴ For example, a smartphone camera may generate a fully signed video stream while filming a high-resolution display that is displaying a deepfake. The resulting video carries strong provenance guarantees about the camera, the device, and the integrity of the capture pipeline, yet the depicted event is entirely fabricated and leads to false confidence.⁵ Therefore, in addition to “sensor provenance”, the content also needs to be signed by the person endorsing the content as authentic. For high-profile settings, including institutions, journalists, and public figures, it is relatively straightforward to issue every high profile entity a key and therefore highly effective in establishing accountability and verifiability for their artifacts.⁶ Establishing a trustworthy database of unique keys for the general public is challenging. If uniqueness cannot be guaranteed, individuals could possess multiple keys, rendering the system ineffective. Therefore, the efficacy of content provenance in preventing scaled misinformation by means of manufacturing large scale

⁴Pan, Bofeng, Natalia Stakhanova, and Suprio Ray. “Data provenance in security and privacy.” *ACM Computing Surveys* 55.14s (2023): 1-35.

⁵This is also why purpose-built secure hardware is needed for PoH. Please see Section 4.

⁶See Web of Trust and this for more information.

participation is almost exclusively dependent on PoH.

Additional hardware-based signals, such as depth sensing, LiDAR, or near-infrared imaging on consumer devices, can further raise the cost of creating images or videos that appear authentic by validating aspects of physical capture. These approaches are directionally valuable but remain limited in scope. They apply only to images and video (not text or audio), depend on compute integrity that is difficult to guarantee on general-purpose devices, and are vulnerable to replay and sensor-spoofing attacks which could result in false confidence.⁷ Additionally, even authentic photos and videos can be taken out of context to support disinformation.

Watermarking and model interface monitoring approaches face a different limitation. In the short term, frontier model providers can monitor the user’s interaction with gated models or embed signals into generated content that enable attribution. These techniques are useful for policy enforcement on specific platforms. However, they rely on strong black-box assumptions: that the generator is known, cooperative, and operating under constraints that adversaries cannot alter. In practice, large-scale disinformation does not require frontier models. Open models have consistently been catching up with frontier capabilities in a matter of months and are already capable of producing convincing text, images, audio, and video at scale. As generative models approach the true distribution of real-world data, reliable detection will likely become infeasible in adversarial real-world environments.

As AI systems become more capable, they can be used to counteract disinformation as well. Today’s models can already perform deep research, verify factual claims, surface inconsistencies, and act as personal moderators for the information people consume (e.g. grok on X). In principle, much of what individuals encounter online could be contextualized or verified before it is consumed. AI mediated monitoring of public and private conversations could analyze conversations for manipulative patterns, coordinated influence, or biased framing, and warn users or dampen certain attacks. This approach seems likely to be effective in some contexts, but it requires large volumes of user interaction to be observed, uploaded, or mediated by the model provider. Privacy preferences around such monitoring vary widely, and this approach concentrates significant power—by choosing what (not) to flag and how to flag it, including politically motivated censorship—in the hands of a small number of actors. As a result this will likely be coupled with regulatory and policy requirements governing acceptable speech and behavior, pushing model-level moderation toward roles traditionally associated with centralized content control.⁸ Separately, the effectiveness of such approaches meaningfully depends on the integrity of the signals they rely on. If feedback, moderation signals, or apparent consensus are dominated by other malicious agents, even the best AI models may amplify distortion rather than correct it.

⁷In practice, this can be bypassed via blocking the laser emitter and spoofing the receiver with real world simulators like this one that is synchronized with the emitter. These limitations are exacerbated on general-purpose consumer devices, where compute integrity and execution environments are difficult to fully secure. World solves this with the Orb (introduced later) which is a custom hardware device that combines multiple secure elements, trusted execution where possible, and complementary multispectral sensors and illuminators across the electromagnetic spectrum, alongside continuously evolving presentation-attack detection algorithms.

⁸In multiple EU countries, individuals have faced fines or prosecution for online speech under national hate-speech and insult laws.

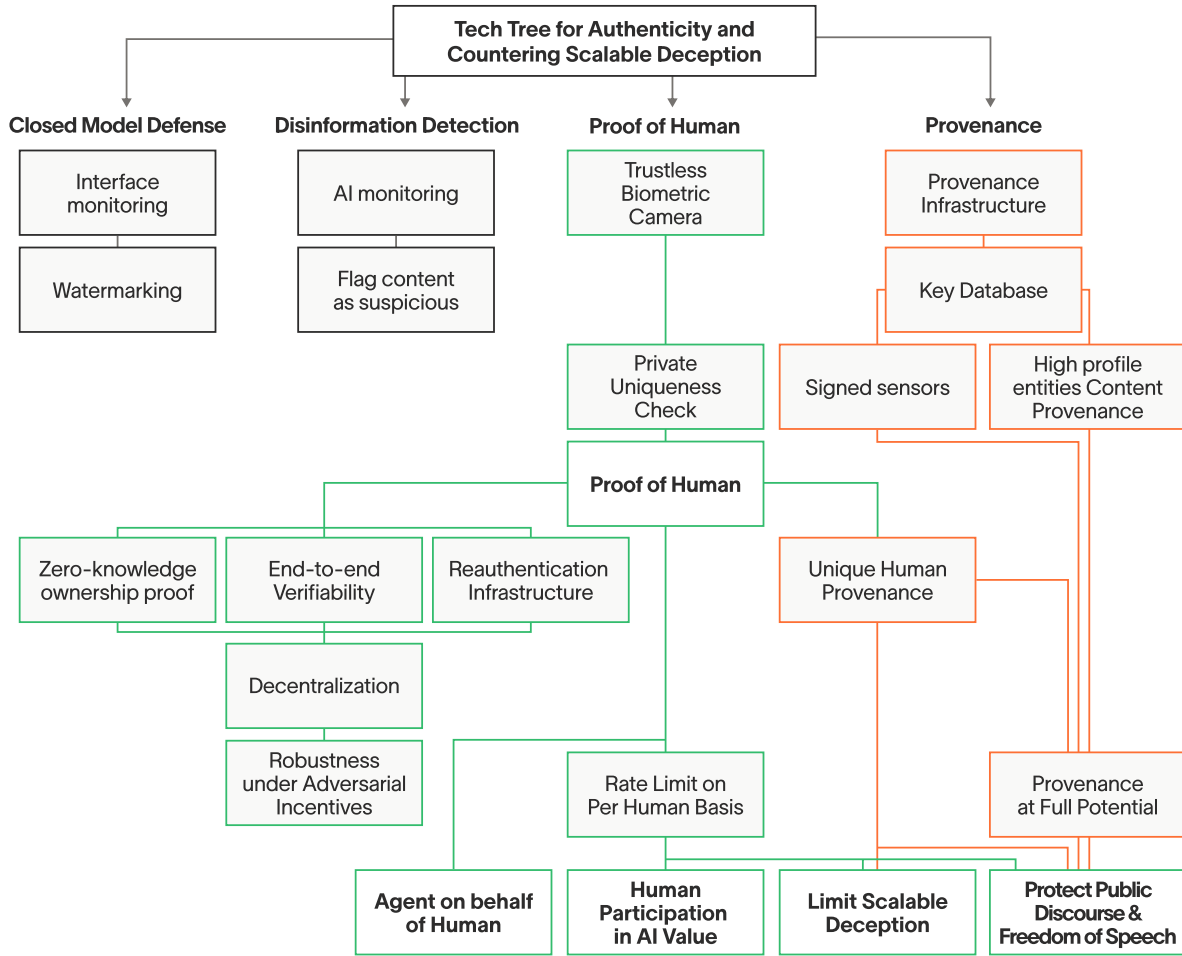


Figure 4. Tech tree for authenticity and countering scalable deception. Four complementary branches: closed model defense, disinformation detection, PoH, and provenance. PoH and provenance converge at the bottom to jointly enable limiting scalable deception and protecting public discourse, while PoH independently unlocks agent-on-behalf-of-human functionality and human participation in AI value. For more information, refer to Section 4.

Therefore, threats like disinformation, large-scale manipulation, impersonation, phishing and identity theft cannot be mitigated through content evaluation alone. Provenance, watermarking, detection, and frontier-model mediated monitoring improve detection at the margins, but they do not address who is participating or how many independent actors exist. Addressing this requires rate limiting interaction based on real humans. PoH makes it costly to manufacture the false perception of a large number of distinct human participants.

2. A World with PoH

PoH provides a way to determine whether an interaction corresponds to a real person. When combined with uniqueness, it enables rate limiting on a per-human basis, rather than per account or device. This is important since it prevents a single person from pretending they are thousands or even millions of real humans.

Importantly, this doesn't prevent multiple pseudonymous accounts or actions per person, but enables bounding the number of accounts. The key property is preventing a single actor from presenting as an unbounded number of independent participants. This approach is compatible with pseudonymity, because it does not require knowing who a person is. PoH infrastructure has many benefits:

The Benefits of Proof of Human		
Limit Scalable Deception	Impersonation Resistance	Protect Public Discourse
Agent on Behalf of Human	Enable Freedom of Speech to Be Heard	Human Participation in AI Value

Figure 5. Six key benefits of PoH. Without a mechanism to bind participation to real people, we risk disinformation, interference with democratic processes, scalable fraud, and exploitation of welfare programs.

2.1. Limit Scalable Deception

By enabling a way to constrain the number of accounts any one individual can create, PoH can directly address scalable deception at its core. Limiting the number of accounts makes artificial amplification of messages and manufactured appearance of broad consensus much more expensive and ideally uneconomical. This significantly reduces the reach and impact of disinformation campaigns. Furthermore, this limitation allows existing mechanisms—such as content provenance (outside of high profile settings), moderation, community notes, and polls—to function as they were intended.

The same participation constraint directly limits large-scale fraud and impersonation. Attacks like phishing, account selling, benefits fraud, and identity theft all rely on the ability to cheaply multiply identities. By enforcing per-human access limits, PoH makes these attacks difficult to scale. This is achieved without the need for continuous behavioral monitoring or cross-service data correlation. Consequently, systems can effectively prevent reply bots, reduce rapid re-entry after a ban, and limit the resale of high-influence accounts while still preserving pseudonymity and avoiding persistent surveillance.

2.2. Impersonation Resistance via Authentication

Many PoH implementations naturally support the creation of locally held (on the user's phone), signed selfie pictures taken during the initial verification. These selfies can be used to sign messages, images, or video streams as originating from a particular human. While this does not detect AI-generated content, it significantly raises the cost of impersonation, in much the same way that two-factor authentication reduces account takeover risk without proving intent. Essentially someone can prove that a particular video stream was authorized by someone that actually looks like the person in the video stream. This makes it very hard to impersonate someone because it is very hard to obtain an authentic signed face picture of someone (assuming the PoH issuer has good security) and then also spoofing a real-time local faceID-like authentication challenge. Applications include authenticating participants in video conferencing, signing emails, and verifying profile pictures and social interactions between individuals.

2.3. Protect Public Discourse

PoH is essential for enabling authentic public discourse in a world where humans and bots are difficult to distinguish. By preventing bad actors from scaling misinformation and manufacturing the appearance of widespread support for specific beliefs or events, PoH safeguards the integrity of

human participation in online opinion formation.

At the same time, as AI lowers the cost of aggregation and analysis, collecting large-scale human input becomes more feasible than before. For governments, this makes real-time policy input more efficient. PoH makes sure that participation can be trusted to come from real people and not bots. Depending on the context, additional credentials that help narrow the set of eligible participants (like country/location) might be needed.

2.4. Agent on Behalf of Human

As AI agents become more capable, an increasing share of human-like online activity will originate from bots rather than from humans. Communication, coordination, and economic actions will often be carried out asynchronously, at high frequency, and over long horizons by systems acting on someone's behalf.

To enable the same benefits PoH creates for human-to-human interaction, agent-mediated activity also needs to be anchored to real humans. This can be implemented as revocable delegation of someone's PoH to an agent. Actions taken by an agent can therefore be verified as occurring on behalf of a human, even when the execution is automated. Importantly, this doesn't enable someone to delegate PoH to a large number of bots but it preserves the core property of PoH: participation and influence are human-bounded and therefore rate-limited.

2.5. Protect Freedom of Speech

Freedom of speech is for humans, not bots. A public "town square" depends on people forming opinions based on engagement with other human participants. When humans cannot be distinguished from bots, individual contributions are harder to surface. Human verification is becoming necessary. PoH is the only option that is effective and preserves anonymity which is essential for free speech. Further, instead of drowning out human voices among bots leading to most humans being assumed to be a bot and people giving up to express their opinion, PoH empowers individuals and enables public debate.

2.6. Human Participation in AI Value / Preparedness for Individual AI Economic Impact Mitigation

Whether AI will create the need for broad redistribution systems, such as benefits for impacted individuals, universal basic compute, or comparable per-person allocations, is plausible but uncertain. However, if such mechanisms become important, there is currently no infrastructure capable of supporting this on a large scale without catastrophic failure modes—especially if it crosses geographical borders. Any system that distributes resources per person is immediately vulnerable to Sybil attacks. Without a reliable way to identify unique humans, redistribution collapses under unlimited duplication and resource drain. PoH therefore functions as preparedness infrastructure: it preserves the option to implement per-human allocation if needed.

3. How to Address Common Concerns and Misconceptions of Proof of Human

Poorly implemented PoH creates severe risks. However, in a rigorous implementation (see Section 4), those risks can be contained:

Concern: PoH is a privacy risk and enables surveillance.

Poorly designed PoH systems can be privacy invasive, enabling tracking across applications. However, a well-designed PoH implementation can be strongly privacy preserving by design. Without PoH, systems must infer legitimacy indirectly through continuous tracking: behavioral monitoring, device fingerprinting, cross-service correlation, and identity checks. These approaches require persistent visibility into user behavior and create strong incentives for surveillance. PoH shifts the model from invasive monitoring to a privacy-preserving proof. PoH can be implemented private-by-design. Secure multi-party computation protects privacy when establishing uniqueness, having multiple issuers minimizes the risk for censorship and zero-knowledge proofs and unlinkable nullifiers preserve anonymity when proving humanness to others (see Section 4.3). This way no cross-service profile can be established and surveillance can be prevented.

Concern: PoH requires a centralized database of people.

This is the case for traditional identity systems. However, for a well designed PoH system, uniqueness can be verified using encryption techniques that avoid exposing biometric or identifying data and distribute trust across multiple independent parties (see Section 4.3) such that there is no need for a centralized database.

Concern: PoH is a centralizing force.

PoH can be designed such that the opposite is the case. Without PoH, influence concentrates among actors who leverage bots, coordinated networks, or purchased accounts. This centralizes power in the hands of those with resources to manufacture participation. PoH inverts this dynamic by making participation human-bounded, which prevents authentic voices from being drowned out and empowers individuals. At the same time, for any PoH there will be an implementation and bootstrapping phase in which in almost all cases will lead to temporary centralization which needs to be iteratively eliminated over time. Initially, there is a small group of people that builds the first version and has decision-making authority. The class of PoH systems we are advocating for can be and need to be progressively protected against incompetence and malice of this initial group of people. The world decentralization whitepaper describes one potential implementation.⁹

PoH leads to a black market for credentials

When PoH becomes critical, malicious actors will be strongly motivated to amass credentials, which would undermine PoH's effectiveness. While complete prevention of PoH delegation to bad actors is likely impossible, several measures can significantly increase the difficulty and economic cost of such actions.

Key security and recovery mechanisms include:

- **Strong Authentication:** Using biometric verification methods to compare the person attempting to use the credential against an embedding tied to the credential ensuring that the credential can only be used by its rightful owner. This includes low-security phone-based face verification all the way to purpose-built secure hardware. Those checks can be performed locally in a private manner see section 4.4.

⁹<https://whitepaper.world.org/#advancing-decentralization>

- Recovery Mechanisms: Enable the legitimate owner to reclaim their PoH following theft or sale, reducing the long-term utility of illicit transfers (this requires careful design—see section 4.3)
- Geographic Association: Optionally disclosing the country of issuance of a PoH can help prevent arbitrage based on income disparities.

Furthermore, as long as each person can only acquire a single PoH, renting it out carries significant risk: if the malicious actor uses the PoH against the terms of use of applications the original owner risks being locked out of essential applications.

Ultimately, if PoH achieves the importance we anticipate, the delegation or misuse of another person's PoH may reasonably be made illegal, mirroring existing laws against the misuse of a passport.

Despite all measures there will always be some black market for PoH credentials. However, even if the measures were entirely ineffective, PoH would still significantly throttle malicious actors since there can only ever be eight billion PoH credentials and reducing fraudulent accounts from infinite to maximum 8 billion is a big reduction. Facebook is banning 1.3 billion spam accounts every three months. Since every PoH credential could only be used once until detected the scale of malicious influence can be significantly reduced. The bounded nature of PoH credentials also increases their price on any black market and makes fraud less lucrative.

Misconception: Face ID solves Proof of Human.

Different systems have different requirements. Authenticating a user via FaceID as the rightful owner of a phone is a very different task from verifying billions of people as unique. The main differences in requirements relate to accuracy and fraud resistance. With FaceID, biometrics are essentially used as a password, with the phone performing a liveness check and a single 1:1 comparison against a saved identity template, to determine if the user is who they claim to be. Establishing global uniqueness is much more difficult. The biometrics have to be compared against (eventually) billions of previously registered users in a 1:N comparison. If the system is not accurate enough, an increasing number of people will be incorrectly rejected due to an incorrect determination that they already have a PoH. Face biometrics are not distinctive enough and therefore would lead to double digit percentage false rejection rates and falsely rejecting billions of people. As previously discussed, uniqueness is not optional but essential. Without uniqueness, one person could pass many Face ID checks on different phones. As a result, they could pose as thousands or even millions of humans and delegate those proofs of human to bots, thereby evading the protections proof of human is meant to provide. Face ID and face biometrics therefore cannot be used as the basis for PoH.

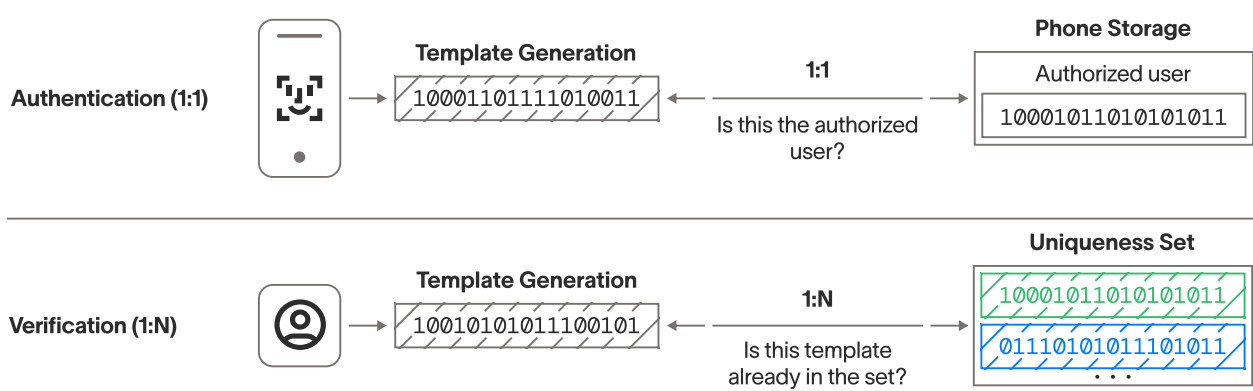


Figure 6. There are two different modes for biometrics. The simpler mode is 1:1 authentication, which involves comparing a user’s template against a single previously enrolled template. This is commonly used in technologies such as Face ID, which compares an individual against a single facial template. However, for global Proof of Human, 1:N verification is required. This mode involves comparing a user’s template against a large set of templates to ensure that there are no duplicate registrations.

Furthermore, phones have inadequate fraud detection capabilities because they often lack signed and/or multispectral imaging sensors which are needed for secure spoof detection. This vulnerability allows for fake images to be introduced. This can be done by replacing the camera hardware with a MIPI stream injector, or using a professional optical setup to bypass display detection algorithms. Such setups might cost hundreds of thousands of dollars but become cost-effective when amortized at scale.

Misconception: Government IDs solves Proof of Human

At first sight, government IDs are a convenient starting point for Proof of Human. One billion people already have verifiable documents and there are legal guardrails against document misuse. However, document-based PoH has severe limitations. Most importantly they create a surveillance risk through governments if recovery is possible. Further, they don’t solve for uniqueness because many people have the same name and one can acquire multiple documents (see section 4.2.2). It is also relatively easy to illicitly acquire real documents and spoof liveness checks which creates a black market for Proof of Human credentials. Additionally, document-based PoH replicates geographical boundaries and eight billion people don’t have verifiable government IDs so the majority of humanity would be excluded. Governments could also print fake IDs and undermine Proof of Human which creates risks like foreign influence operations. Similarly, since governments are (by design) in control of issuing IDs, this would enable strategically excluding certain populations and thereby excluding them from interacting on the internet as a human and therefore severely limit their freedom of speech to be heard by other humans. Therefore, document-based PoH can be a helpful addition while establishing a PoH that actually empowers people but it doesn’t actually solve it. Section 4.2.2 goes into more detail on a government document-based root of trust for PoH.

4. Building Proof of Human

The implementation of PoH can take many forms, leading to vastly different societal and individual consequences. These potential outcomes span a spectrum, from intrusive, Orwellian surveillance to systems that safeguard privacy and actively enable free expression. Consequently, the core system architecture, along with the resulting capabilities and incentives for all participants, must be designed with extreme care. This section outlines the key properties that we believe are essential

for creating the best possible future for humanity, separate from the World project. World ID is our effort to translate these properties into a working system.

4.1. Design Requirements that Maximize Individual Empowerment

To establish design requirements, we first need to define what we value most. We place the highest importance on individual empowerment because we think this leads to the most beneficial implementation for (human) society. This means, the design requirements should be defined such that they: prevent surveillance, maximize privacy, ensure broad participation and accessibility, and upholding freedom of expression and individual agency.

Importance of Uniqueness.

Counterintuitively, proof of human alone is insufficient because, without uniqueness it is vulnerable to relay attacks. In this scenario, a small group of humans could authenticate repeatedly to serve millions of automated agents—picture a “human call-center” dedicated solely to passing proof of human challenges for bots.

It is important for uniqueness to be strict: one, and only one, Proof of Human credential per person.¹⁰ Allowing individuals to acquire even a small number of credentials is not enough to protect integrity. If PoH becomes vital societal infrastructure, the incentive to bypass it becomes extremely high. If it is possible to acquire multiple credentials, some will find it lucrative to sell all credentials but one. As a result, if a malicious actor convinced just 1% of the US population to sell nine of their ten credentials, that actor could present as 31 million US individuals. Given only a small fraction of people usually participate in debates on any given social media platform this is sufficient for malicious actors to control the majority of participation in most situations. Elected governments are powerful which creates a significant incentive and threat to manipulate it. Especially during elections bots can pretend to be highly vocal individuals, drown out the voices of real people, manufacture apparent consensus and change voters’ opinions which could change election outcomes. Therefore, to maximize individual empowerment and protect the integrity of the society, uniqueness for the purpose of PoH must be absolute—exactly one credential for every human.

Proof of Human Design Requirements.

PoH is still nascent and the thinking around requirements will likely evolve from practical experience as adoption increases. Starting from first principles, valuing individual empowerment above all else leads to the following design requirements:

- Inclusive and scalable. It must be feasible for every single human on the planet to be able to participate: across geographies, economic conditions, varying levels of digital literacy, and levels of technical access.
- Unique and high-integrity. Individuals should be able to acquire exactly one Proof of Human, and not more. In order for uniqueness to be effective, the uniqueness “test” must be very hard to spoof / bypass.
- Person-bound. Credentials must be difficult to sell, or transfer at scale; otherwise, uniqueness collapses via a black market for credentials.
- Privacy-preserving. Proving that one is a unique human should not require revealing identity, and should support unlinkable use across contexts.

¹⁰This makes it important that the PoH is basically impossible to lose and steal to make sure nobody is left without a PoH.

Taking the above requirements seriously requires investing in a very sophisticated system, which requires substantial resources to establish but seems likely to be the best outcome for humanity.

How to Build Proof of Human to Maximize Individual Empowerment and Human Agency

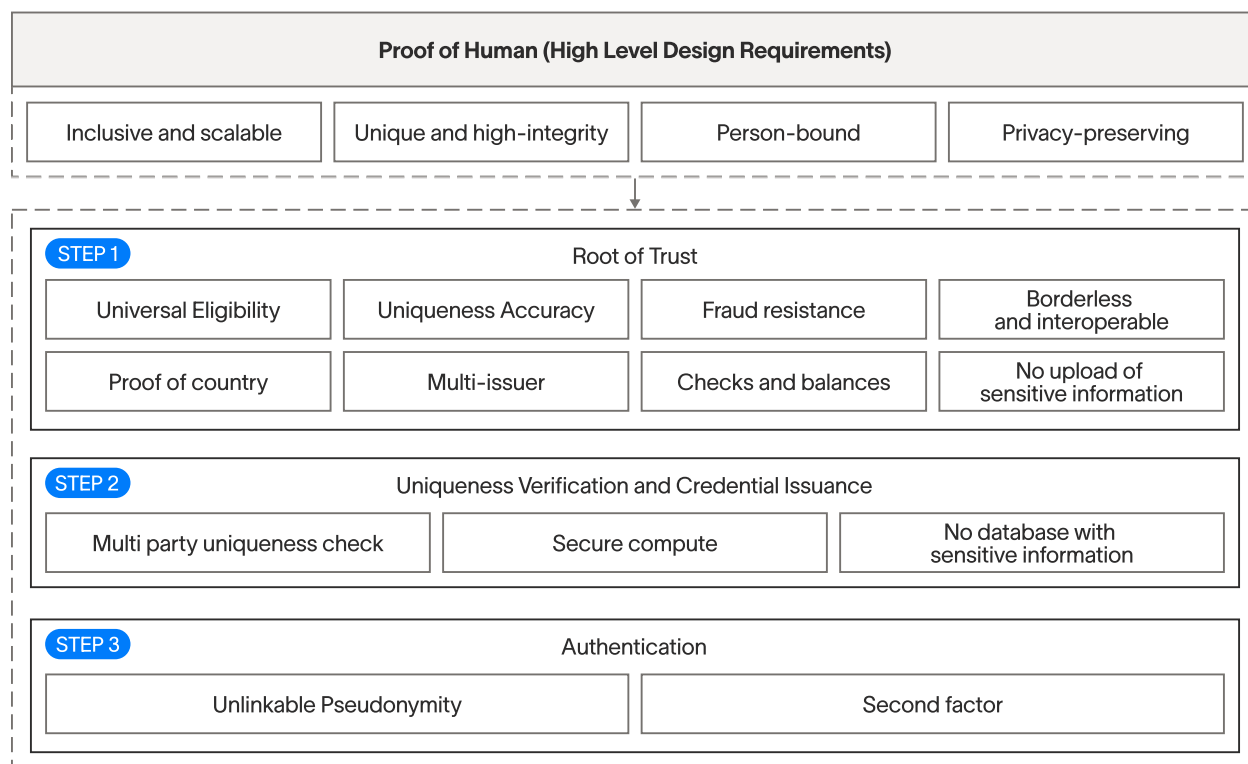


Figure 7. Building PoH in a way that maximizes human agency leads to high level design requirements. Those can be broken down into derivative requirements for the three core components of PoH: root of trust, uniqueness verification and credential issuance, and authentication. These requirements are costly to implement in practice but important to empower individuals and prevent privacy-invasive tracking.

The process of proving unique humanness involves three distinct stages:

- Stage 1: Root of Trust—Proving an individual is human and acquiring verifiable, high-integrity information. This information is then used in the second stage to establish uniqueness.
- Stage 2: Uniqueness Verification and Credential Issuance—Verifying uniqueness based on the previously acquired information and issuing a unique human credential.
- Stage 3: Authentication—Using the issued credential to prove one’s unique human status to other parties.

Based on the initial design requirements, we can deduce derivative design requirements for each of these three stages:

Design Requirements for Root of Trust.

- Universal Eligibility. Anyone needs to be eligible
- Uniqueness Accuracy. In order to be able to establish uniqueness, there needs to be enough entropy to distinguish between $\mathcal{O}(10B)$ humans without falsely rejecting a single person.
- Fake Human Resistance. In order to prevent bypassing uniqueness, the humanness and uniqueness test needs to be very hard to bypass or spoof. This includes the ability for the issuer to be able to act quickly if a compromise becomes known.

- Borderless and interoperable. Inclusivity means that citizens from different countries need to be able to prove to each other that they are humans; High integrity uniqueness means countries need to be able to trust the PoH from citizens of other countries to not be bots to prevent influence operations.
- Proof of Country. In order to trust PoH in certain high stakes scenarios with national context (e.g. opining on presidential candidates) and to prevent global income arbitrage to acquire Proof of Human credentials, it is important for people to be able to prove in which country their credential was issued.
- Multi-issuer. In order to ensure inclusivity and make sure no entity has the power to exclude someone from receiving a PoH, there need to be multiple entities as a root of trust—but importantly, they all need to tie into the same uniqueness set.
- Checks and balances. In order to prevent an issuer from printing identities and subverting the uniqueness requirement it needs to be possible to validate one’s PoH via another issuer, which can make it game-theoretically uneconomical to create fake identities.
- No upload of sensitive information. In order to preserve privacy, no sensitive information should be uploaded.

Design Requirements for Uniqueness Verification.

- Multi-party uniqueness check. In order to enable inclusivity and prevent any party from being able to block anyone from acquiring a PoH, the uniqueness check should be performed by multiple parties.
- Secure compute.¹¹ In order to enable high integrity uniqueness, it must be difficult as possible for any single party to adversarially inject fake identities or deviate from the comparison protocol in any way.
- No centralized database with sensitive information. To preserve privacy, sensitive information must not be aggregated or stored in any centralized database; instead, the system should minimize data exposure and avoid single points of compromise.
- Recovery: Enable the legitimate owner to reclaim their PoH following theft or sale, reducing the long-term utility of illicit transfers.

Design Requirements for Authentication.

- Unlinkable Pseudonymity. In order to preserve privacy, it should not be possible to identify who someone is or to track someone across different contexts.
- Illicit transfer prevention. In order to prevent anyone from acquiring other people’s credentials, the system needs to ensure that the person using the PoH credential is the one that it was issued to. Therefore, beyond the credential, there needs to be a person-bound second factor that can be used to authenticate the credential holder as the rightful owner of the credential on a periodic basis to prevent illicit transfer.

4.2. Ideal Root of Trust for PoH

The following sections walk through a first principles approach of choosing an ideal root of trust for PoH based on the above requirements.

4.2.1. Overview

When evaluated against the design requirements outlined above, most candidate PoH mechanisms fail for structural reasons. Approaches based on online accounts, social graphs, webs of trust, or financial credentials can be mimicked by AI systems, and ultimately require accepting multiple

¹¹“Maliciously secure” cryptographic protocols are those that enable multiple parties to jointly compute a function over private inputs, while guaranteeing privacy and security, including if some parties behave “maliciously” (i.e. they try to cheat or otherwise deviate from the protocol).

identities per person. Identity and document-based approaches, on the other hand, suffer from poor inclusivity (vast majority of the global population excluded when only considering documents that have sufficient integrity), aren't cross-verifiable across issuers which gives governments the ability to exclude people from being able to prove they are human and not all governments are trusted equally (for more details see section 4.2.2). Biometric verification is the only class of mechanisms that can simultaneously satisfy all design requirements at global scale when implemented correctly (see section 4.2.3).

Among biometric modalities, iris recognition uniquely satisfies the accuracy, scalability, and privacy requirements of global uniqueness verification. Iris biometrics achieve false match rates on the order of 10^{-14} .¹² Additionally, the iris texture is formed through random morphogenesis during gestation, is highly stable over time, and is difficult to alter, resulting in essentially uncorrelated patterns even between identical twins. Other modalities exhibit fundamental limitations: fingerprints degrade and are vulnerable to combinatorial attacks; facial recognition lacks sufficient accuracy at global scale; and DNA, while accurate, is not spoof resistant, difficult to scale, and reveals extensive additional personal information. As summarized in Figure 8, iris recognition is therefore the only biometric modality that currently balances accuracy, privacy, and scalability for global PoH. This requirement is fundamentally different from consumer biometric authentication systems such as Face ID, which perform local 1:1 authentication against a single stored template; global PoH instead requires highly accurate 1:N uniqueness verification against billions of prior registrations.

Biometric Modalities				
	Fingerprint	Face	DNA	Iris
Privacy	Possible	Possible	Hard	Possible
Accuracy for global scale	Not enough	Not enough	Sufficient	Sufficient
Scalability	High	High	Low	High
Integrity	Low	Medium	High	High

Figure 8. Overview of how different biometric modalities impact key considerations such as privacy, accuracy, scalability, and integrity (red indicates insufficient or problematic). Iris biometrics is the only modality that enables all of them.

Meeting the design requirements outlined above necessitates guarantees that cannot be achieved through software or general-purpose hardware alone. Reliable verification of human uniqueness depends on compute integrity, ensuring that image capture and processing occur within a tamper-resistant environment and cannot be emulated, replayed, or modified. This requires hardware-backed signing, secure execution, and protections against compromised or simulated enrollment flows.

Robust enrollment further depends on multispectral imaging and active liveness detection to defend against presentation attacks, such as images or videos displayed to a camera. Single-sensor consumer devices lack the signal diversity needed to reliably distinguish genuine human presentations from high-quality spoofs in adversarial settings.

Finally, high-resolution infrared imaging is required to capture sufficient iris entropy across eye colors. Visible-spectrum cameras suffer from reflections and low contrast, particularly for darker

¹²For more information on how World's iris recognition inference system can achieve false match rates that scale to billions of people, please refer to this blog post.

irises, resulting in noisy measurements and increased error rates.

Taken together, these constraints motivate the use of a purpose-built device rather than reliance on general-purpose hardware. Based on this conclusion, Tools for Humanity (TFH) built a high-security open-source camera called the Orb, which anonymously issues an AI-safe¹³ PoH credential. The Orb is purpose-built to verify humanness and ensure uniqueness in a fair and inclusive way.

4.2.2. Why a Document-based Root of Trust is Not Ideal for Individual Empowerment

In a future shaped by highly capable machine intelligence, the foundations of economic and social power shift. As automation increases, control over bots as well as the ability to determine who counts as a unique human will become an increasing source of power. Any entity that controls PoH gains significant influence over access to platforms, economic participation, and collective decision-making. Therefore, any incentive misalignment between issuer and participants can lead to catastrophic failures. The inherent nature of how documents are issued suggests that any PoH system based on them will result in long-term incentive misalignment.

In short: Despite its benefits, such as offering legal incentives against issuer hacking (as a hacker would face prosecution; although this could be implemented through new policies for other PoH roots of trust as well) and the fact that one billion people already have a verifiable document, the structural limitations of document-based PoH make it unviable on a global scale. The following section discusses the reasoning in more detail.

Properties of Document-Based Proof of Human	
Universal Accuracy	Not possible
Universal Eligibility	Not possible
Fraud resistance	Possible
Borderless & interoperable	Not possible
Proof of country	Possible
Multi-issuer	Not possible
Checks & balances	Not possible
No upload of sensitive information	Possible

Figure 9. Properties of document-based PoH evaluated against core design requirements for a root of trust that maximizes individual empowerment. Five of eight requirements (red) cannot be met by document-based approaches.

In the following, we analyze a document-based root of trust against the design requirements:

Universal Accuracy: Establishing uniqueness using government documents is challenging. Names are insufficient for deduplication due to high collision rates (e.g., nearly 40,000 James Smiths in the US), and the birthday problem further complicates uniqueness checks. It is possible for more than 10 people to share the same name, birthday and birth year, and all but one of them would be excluded if uniqueness were strictly required. It would be possible to create a unique identifier

¹³In this context, AI-safe refers to a process that is hard for AI models to bypass. For example, spoofing the Orb is significantly harder for AI than performing a CAPTCHA because the Orb exists in the real (e.g. not digital) world and uses multiple sensors to confirm that the images being captured are coming from the real world.

based on name, birthday and the face image. However, this would require being able to recover the PoH credential using the passport. Given governments have a central database with passport information it would create a very powerful surveillance tool for governments to see past activity and is a significant risk to individual freedom. Therefore, a document’s numerical identifier is the only viable uniqueness signal since it doesn’t require recovery but a new passport could be used to create a new credential. This creates a vulnerability where individuals could acquire multiple PoH simply by reporting their document “lost” and obtaining a new one. If PoH turns out to be as important as we think it will be, this creates a high price for PoH credentials and thereby incentivizes individuals to acquire multiple PoH and sell all but one. Therefore, it enables well resourced actors to acquire a large number of PoH. If 1% of the US population participated, this could easily lead to tens of millions of “authentic human” bot accounts which could be quite effective in influence operations.

Universal Eligibility: Only about one in eight people possess documents that can be cryptographically verified, which is required to prevent deep-faked registrations. Therefore, basing PoH on documents would exclude many billions of people, leading to a system that may function within certain jurisdictions but is globally suboptimal.

Fake Human Resistance. Some documents are cryptographically verifiable which makes them relatively fraud resistant. While there are still attack vectors, this makes it relatively difficult for attackers to create malicious PoH credentials. However, Document-based PoH incentivizes the theft of physical documents like passports (e.g., at hotels or from the mail). Stolen documents can be used to generate PoH or, in some implementations, even be cloned without the owner’s knowledge. While face authentication via the user’s phone can raise the bar, the achievable spoof and compute integrity using phones is limited, especially since the bar is defined by the weakest phone. As AI capabilities grow, the importance of PoH will also increase. This progression will lead to a rapid evolution in both situations where PoH is useful and necessary as well as potential attack vectors. As a result, the issuing infrastructure will need increasingly advanced capabilities, including securing the root of trust, anomaly detection, verifying uniqueness across issuers to increase integrity, revocation mechanisms, and dynamic expiration dates. Developing these sophisticated capabilities is likely better suited to several publicly auditable and mutually verifiable companies, rather than governments, as it aligns more closely with their core competencies.

Borderless & Interoperable. Achieving interoperability is relatively straightforward since verifiable documents are already based on standards and public keys of the certificate signing authorities are publicly accessible. However, the fact that governments could create fake documents and therefore inject fake PoH credentials make it hard to trust foreign credentials since different governments may be misaligned in their incentives (e.g. due to influence operations). Having the ability to prove which country a PoH credential was issued by helps to mitigate the most severe risks but leads to a replication of geographical borders on the internet and may in many cases even lead to the exclusion of free exchange between people from different countries. Importantly, this is not an inherent property, but just a result of issuer incentives. In the case of a biometric root of trust, governments don’t have special privileges to issue PoH credentials (preventing hacking is a different challenge and can be made harder by having a diverse set of hardware devices that can be cross-checked against each other) which makes it much less likely to have malicious influence operations (but not impossible) and therefore leads to less segregation of the internet.

Proof of Country. Documents include the issuing country which makes a proof of location straightforward.

Multi-issuer. Wherever PoH becomes a prerequisite for interaction (e.g., platform access, financial systems, governance mechanisms, or public services) the issuer gains leverage through the ability to grant, deny, or revoke credentials. In practice, this gives the issuer a powerful control surface: loss of PoH will eventually imply exclusion from large parts of the internet including significant limits on freedom of speech (since one would be assumed to be a bot and therefore not be taken seriously). This can be mitigated if no single entity or institution has a monopoly over the issuance of PoH in any given region, which is not possible for government documents.

Checks & Balances. The ability to issue PoH credentials directly translates to power. Any entity that can mint PoH credentials can create artificial participants, enabling influence operations, manipulation, or resource capture at scale. As a result, the security and integrity of the entire system is constrained by its weakest issuer. In the case of government documents, this means the country with the lowest standards defines the security bar. To prevent an issuer from generating fake PoH identities (e.g. for influence operations), there must be a mechanism to validate integrity across different issuers. For government-based documents, no second issuer exists for cross-checking and validation. For biometric based PoH there can be a diverse set of hardware devices from different manufacturers and with different supply chains that can be cross-checked against each other.

No upload of sensitive information. It is possible to verify the integrity of cryptographically verifiable documents without uploading sensitive information.

The risks of document-based PoH become more significant in scenarios involving large-scale redistribution of economic resources. Many models of AI progress assume a meaningful change in human labor¹⁴ and the need to distribute resources, such as money or compute, on a per-person basis. Benefits access would likely need to be mediated via PoH (especially if it is cross-country) to prevent fraudulent depletion of resources. Therefore, the issuer gains the ability to exclude individuals from these systems or print identities and capture resources. For these reasons, PoH infrastructure should not be controlled by a single centralized entity (not even governments) but instead be mediated by maximally transparent, decentralized infrastructure with end-to-end verifiable issuers—like auditable biometric hardware that enables cross-checks of PoH across different hardware manufacturers to ensure authenticity which can make fraudulent behavior game-theoretically uneconomical.¹⁵

4.2.3. Why Iris-based Root of Trust Maximizes Individual Empowerment

To our knowledge, a set of secure hardware devices from different companies (that are based in different countries) and issue a root of trust based on the entropy of the iris, is the root of trust that maximizes individual empowerment and is the only root of trust that fulfills all design requirements. Note: This is not where World ID is today; but it should get there eventually. Iris-based hardware—while being able to fulfill all requirements—also comes at the cost of being very capital intensive and operationally complex to scale.

¹⁴See Bostrom, *Superintelligence* (2014).

¹⁵Refer to World’s decentralization whitepaper here, in particular the section on verification devices.

Advantages of Iris-based Root of Trust	
Universal Accuracy	Possible
Universal Eligibility	Possible
Fraud resistance	Possible
Borderless & interoperable	Possible
Proof of country	Possible
Multi-issuer	Possible
Checks & balances	Possible
No upload of sensitive information	Possible

Figure 10. Unlike document-based approaches (Fig. 9), an iris-based root of trust satisfies all core design requirements for PoH that maximizes individual empowerment.

Walking through the design requirements:

Universal Accuracy: Global uniqueness requires false match rates on the order of 10^{-20} in order to make sure that no single person on earth would mistakenly not be issued a PoH credential and thereby be excluded from participating in the human internet. This is a very low error rate and very high bar to achieve. Iris-based uniqueness algorithms are the most accurate and achieve error rates on the order of 10^{-14} today. Those can be further improved. If the improvements aren't sufficient, it can be combined with face-based entropy which in combination would already today achieve error rates below 10^{-20} .

Universal Eligibility: Iris biometrics far surpass the inclusivity of other root of trust alternatives like verifiable documents by many orders of magnitude. It is important to note that many health conditions, like cataracts to a certain degree, do not impede iris biometrics. However, if PoH becomes essential for society, it is important that not only 99% of people are eligible but that eventually every single person can verify if they want to. Although not currently established, there could be specialized verification centers to facilitate alternative means of verification for individuals with eye conditions, via e.g. facial biometrics. Additionally, hardware devices would need to be accessible to everyone which is a hard operational challenge and requires significant capital which are serious hurdles but surmountable.

Fake Human Resistance: The biometric camera should include multispectral cameras, a TEE, verifiable software, as well as multiple secure elements in order to make spoofing it very expensive. However, no hardware system interacting with the physical world can achieve perfect security even with ongoing red teaming and deploying of patches. It is expected that hardware devices may get spoofed or compromised by determined actors. PoH can be designed with this threat in mind: any root of trust issued by a particular hardware device should be possible to be revoked through a governance mechanism. Implementing suitable incentive mechanisms for decentralized audits of all hardware devices in operation can help raise the bar far beyond what hardware security alone could achieve in isolation, especially for scalable attacks. Buying PoH is an issue for any PoH implementation. In order to make it as costly as possible to acquire or rent PoH credentials on the black market, hardware devices make it possible to re-authenticate on a periodic basis with very high integrity to establish that the holder of a PoH credential is indeed the legitimate owner. In combination, those mechanisms make it conceivable to establish a very high security bar.

Borderless & Interoperable: Biometrics and hardware are inherently borderless and it is relatively straightforward to build the supporting infrastructure in a way that it is interoperable between countries.

Proof of Country: Hardware devices can include means to verify their location via e.g. celltower and GPS connectivity as well continuous audits of devices and ensuring they are in the country they are expected to be. This way, the root of trust can include a country credential. However, anyone can travel to a different country in order to verify with a hardware device there and spoof the country credential this way. Although this is possible, with ongoing re-authentication it becomes expensive to continuously travel for this. The fraction of people that manipulate their country credential this way is likely low. If this turned out to be an issue, one could add document-based countries for certain use cases and accept reduced inclusivity.

Multi-issuer. Multiple companies across different jurisdictions can build auditable and verifiable hardware devices to the same specifications.

Checks & Balances: Hardware devices from different companies can issue iris-based roots of trust to the same specifications which makes it possible to cross-check between them. If hardware manufacturers e.g. staked a security deposit, it would align incentives with integrity and with the right design of periodic re-authentication across hardware devices from different companies and security deposits it can become game-theoretically uneconomical for a hardware manufacturer to engage in injecting fake identities and it can make it hard to compromise the integrity of the root of trust for a malicious actor since it would require compromising different hardware devices at the same time.

No upload of sensitive information: It is possible to issue an iris-based root of trust without uploading sensitive information.

Deploying hardware for an iris-based root of trust for PoH in an adversarial environment is unprecedented in both scale and complexity.

4.3. How to Privately Verify Uniqueness on a Global Scale?

Determining whether a person has already verified requires global information from all prior verifications to compare against, so this process cannot happen locally on verification devices. Instead, a global uniqueness check service is required, complementing the verification hardware in the PoH issuance process. We discussed three properties in the design requirements section that are critical for maximizing the alignment of the uniqueness check with individual empowerment:

- **Multi-party uniqueness check.** In order to enable inclusivity and prevent any party from being able to block anyone from acquiring a PoH, the uniqueness check should be performed by multiple parties.
- **Secure compute.** In order to enable high integrity uniqueness, it must be as difficult as possible for any single party to adversarially inject fake identities or deviate from the comparison protocol in any way.
- **No centralized database with sensitive information.** To preserve privacy, sensitive information must not be aggregated or stored in any centralized database; instead, the system should minimize data exposure and avoid single points of compromise.
- **Recovery:** Enable the legitimate owner to reclaim their PoH following theft or sale, reducing the long-term utility of illicit transfers.

One way to implement these requirements is through a secure multi-party computation (SMPC) protocol that verifies uniqueness in an anonymous manner without revealing biometric data to any entity. At enrollment, biometric information is processed locally in a TEE on a custom biometric camera and transformed into encrypted, statistically random fragments by verifiable software. Those fragments are then sent to the user's phone. The user can then choose to send them to multiple independent node operators as shown in Figure 11. Those nodes have private state and jointly determine whether someone has verified before — crucially, in such a manner where no party learns any statistically meaningful information about the underlying data whatsoever, except whether the entry is unique.

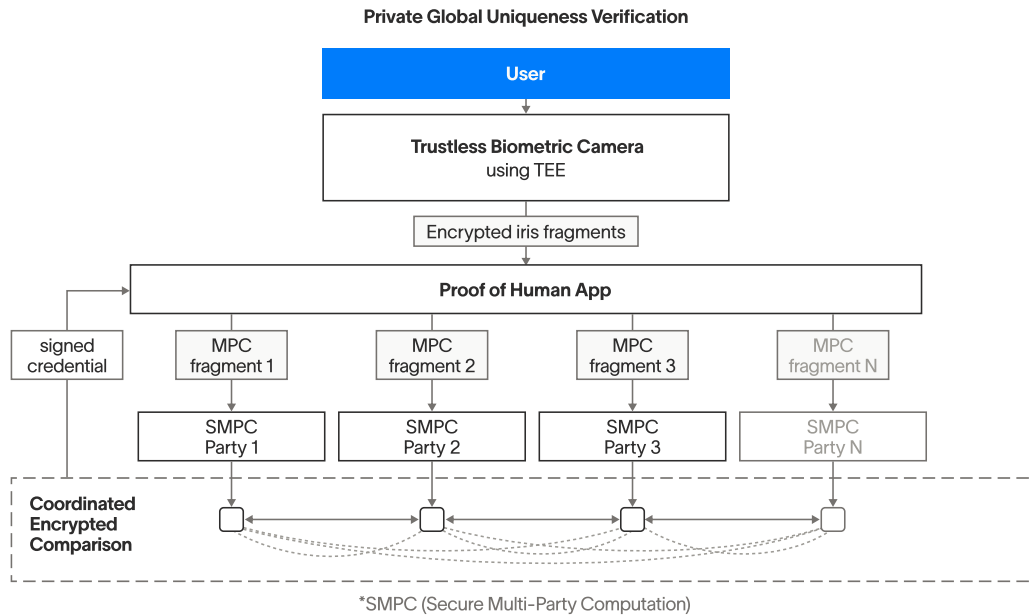


Figure 11. When a new person goes to a trustless biometric camera, their iris data is converted into a unique code and split into multiple encrypted, statistically random fragments and sent to the user's app. The user can then send those fragments to different SMPC parties that independently compare the new fragment against their set of existing encrypted fragments, and together, determine the existence of potential duplicates without revealing the underlying data to any party. If all SMPC results indicate that the fragments are unique, each party stores its respective fragment and a signed PoH credential is returned to the user. This process ensures that each Proof of Human corresponds to a single, unique human while preserving user anonymity.

If the uniqueness check was successful, the encrypted secret share fragments are stored in the respective SMPC node and a signed Proof of Human credential is returned to the user. Importantly, the credential should not be used directly to prove humanness in order to prevent tracking across applications, as discussed in section 4.4.

As PoH becomes more widely relied upon for platform access, economic participation, and public discourse, losing access to one's credential becomes a serious problem. However, enabling recovery via issuing a new PoH credential would conflict with the uniqueness of PoH, making credential recovery challenging.

Any recovery mechanism must meet strict criteria: it must preserve the owner's privacy, and only the legitimate owner can be allowed to trigger it. This places significant requirements on the root of trust for recovery. Relying solely on a document like a passport is insufficient, as it would be too

easy to impersonate the owner.

Crucially, recovery must deactivate the old access and issue new access without resetting the PoH’s unique property. This is necessary to prevent individuals from fraudulently presenting as multiple people or circumventing a block issued due to misuse.

One way to implement this recovery is to store the PoH key in an SMPC system, accessible via user-stored authentication keys. If these authentication keys are compromised, the user can undergo a trusted verification process (e.g., via a secure biometric camera). This process would temporarily enable the user to deactivate the compromised keys, add new ones, and thus regain access.

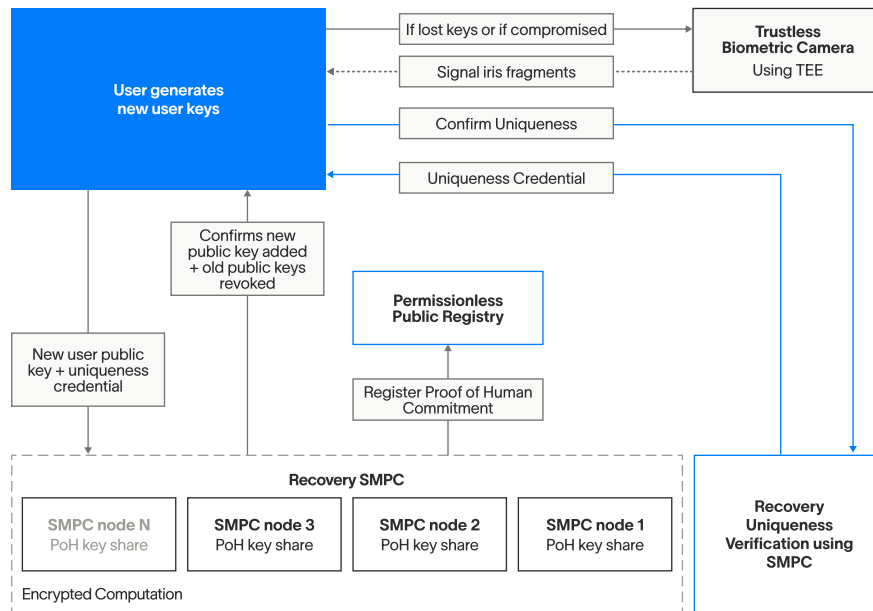


Figure 12. Recovery flow using biometric re-verification and SMPC-based key revocation. The SMPC nodes can act as a recovery agent for the user’s identity; if compromised or lost, a trusted biometric camera enables key rotation and thereby recovery.

4.4. Authenticating Using Proof of Human

A robust authentication mechanism is needed to prove that an action originates from a human. Further, this mechanism must be privacy-preserving and prevent cross-context tracking of individuals across services. There are two requirements that describe these features:

- **Unlinkable Pseudonymity.** In order to preserve privacy, it should not be possible to identify who someone is or to track someone across different contexts.
- **Illicit transfer prevention.** In order to prevent anyone from acquiring other people’s credentials, the system needs to ensure that the person using the PoH credential is the one that it was issued to. Therefore, beyond the credential, there needs to be a person-bound second factor that can be used to authenticate the credential holder as the rightful owner of the credential on a periodic basis to prevent illicit transfer.

One way to address unlinkable pseudonymity is through a combination of self-custody (user-held key authenticating materials) and zero-knowledge proofs. Credentials are held locally by users and

can be presented without revealing identity or linking activity across contexts. A public, tamper-resistant registry of these credentials enables verification and revocation without exposing personal data.

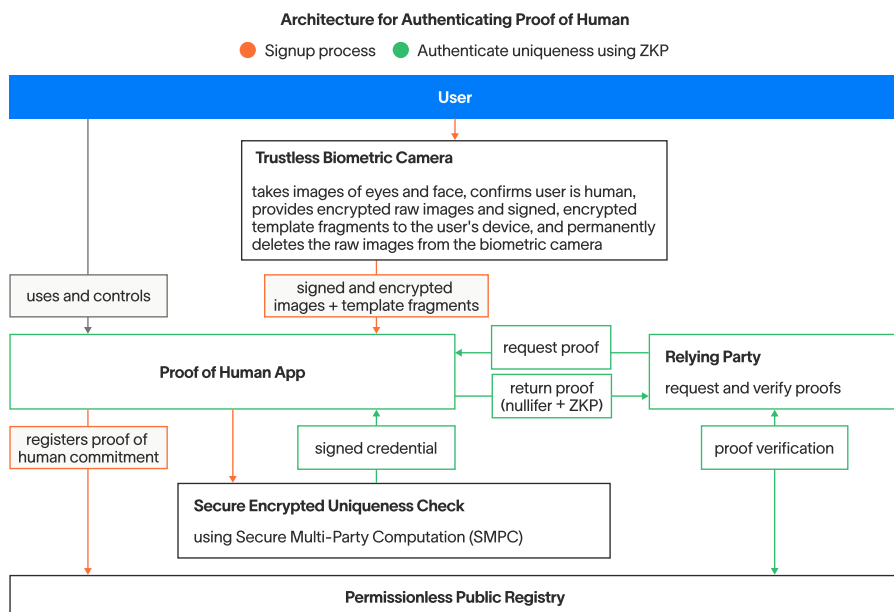


Figure 13. During enrollment (orange), a biometric camera captures images to confirm the user is a real human, then provides signed and encrypted template fragments to the user’s device before permanently deleting the images. A SMPC network performs an encrypted uniqueness check to ensure the user has not previously enrolled, and the app publishes the proof of human commitment to a public registry. During verification (green), a relying party requests proof of humanness from the user’s app, which generates and returns a zero-knowledge proof (ZKP) paired with a nullifier. The relying party verifies this proof against the public registry, confirming the user is a unique human without revealing their identity.

Additionally, a second factor is required. The initial verification alone is insufficient to maintain integrity over time. A PoH credential must remain person-bound in ongoing use to prevent selling, stealing, and renting of credentials. Without continuous authentication, credentials could be transferred to malicious actors—voluntarily or not. Continuous authentication like face-based checks against embeddings from the initial verification can be performed locally on the user’s device to ensure that the person presenting a PoH credential is the same person to whom it was issued. By requiring frequent re-authentication, the original owner cannot hand off their credential for extended periods without either being physically present or risking loss of access. For high stakes use-cases, users can go back to a purpose-built secure biometric camera for a high assurance authentication (think “anonymous notary”).

5. World ID: Bootstrapping Proof of Human

World is an effort at implementing Proof of Human as described above, with the design goal to maximize individual empowerment.

5.1. Proactively Implementing Proof of Human

Bureaucracies tend to address major problems only after significant damage has occurred, which helps focus on what matters most. However, a reactive approach to PoH is highly undesirable and could lead to surveillance. This not only results in avoidable negative consequences but also

increases the likelihood of a less considered and simpler implementation, which would likely impact efficacy, privacy and freedom of speech. By the time a crisis makes PoH seem necessary—potentially involving extreme outcomes like meaningful increases in societal unrest or other threats to democracy like outcome-altering election interference—the resulting PoH would most likely be implemented in haste. A rush would likely lead to prioritizing speed and functionality above all else, compromising critical elements such as individual empowerment, privacy and resilience to adversarial actors, making the final system far less beneficial than a proactive solution in ways that cannot be addressed iteratively afterwards (local optimum).

The World project is a proactive effort to address these challenges. In order to bootstrap PoH ahead of time and avoid some of the negative consequences, World employed one particular approach to try to counteract the described dynamics: It used its token as an economic bootstrapping mechanism while PoH is new and adoption is nascent and provides an incentive to verify. Analogous mechanisms have been used historically to grow networks at a smaller scale: for example, in PayPal’s early years, the company invested in user incentives to rapidly expand user adoption and achieve network scale, which was critical to its transition from a niche startup to a global platform. The token further gives all network participants a native ownership share for their participation, because the overall system becomes more useful as more people are verified.¹⁶

5.2. Multiple Uniqueness Credentials

In practice, supporting multiple forms of uniqueness can be useful in the short term to accelerate adoption and improve coverage. Therefore, World ID supports not only PoH credentials through the Orb but also credentials through passports and face based verification.

However, when evaluated against the requirements for global PoH, these alternatives exhibit structural limitations. Cryptographically verifiable identity documents with embedded chips (e.g., NFC-enabled IDs) depend on inhomogeneous availability, heterogeneous security infrastructure, and document issuance processes rather than the person themselves. Uniqueness degrades through issuance by multiple authorities (e.g. dual citizenship), re-issuance or replacement, and robust recovery is incompatible with privacy without introducing avenues for serious government surveillance. In more extreme cases, not all governments might be trusted to not create fake IDs (to potentially disseminate disinformation in other countries or even their own).

Similarly, face-based verification on consumer devices provides limited assurance and does not scale to global uniqueness. Camera quality and sensor diversity limit entropy and therefore the ability to distinguish lookalikes. This makes it practically impossible to distinguish real people from high resolution displays showing deep fakes. Compute integrity is also insufficient to prevent attacks.

For these reasons, while multiple credential types should coexist as transitional or complementary mechanisms, purpose-built biometric hardware appears to be the most beneficial long term approach that satisfies the full set of requirements for robust and privacy-preserving PoH.

5.3. World ID Today

Most of the security and privacy measures outlined previously—such as the secure hardware device (the Orb), Secure Multi-Party Computation (SMPC), Zero-Knowledge Proofs (ZKPs), and face authentication—are already integrated into World ID. This is complemented by numerous other privacy-enhancing features and ongoing development to further strengthen the system, including

¹⁶For more information on WLD, please see this.

World ID 4.0.

World ID is progressing towards global adoption. At the time of this writing, there are about 40M people on World App (the first of ideally many World ID authenticator apps) and about 18M people verified as human with an Orb. At the same time, current projections for advanced machine intelligence require accelerating adoption in order to enable the benefits of PoH when they are needed.

5.4. Looking Ahead

Recognizing PoH as critical infrastructure still requires a meaningful shift in public perception which is likely the biggest limiting factor for proactive adoption and safeguarding humans in a world with advanced machine intelligence. Although awareness of scalable automation is increasing, PoH is usually framed as a narrow anti-bot measure rather than as a foundational layer for human coordination and societal stability. Legitimate concerns around surveillance and privacy (which can be solved) often lead to conflating privacy-preserving PoH with centralized identity systems that collect and monitor personal data slow down adoption. At the same time, AI capabilities continue to advance in both scale and coordination, and as machine capabilities compound, the absence of widely adopted, high-integrity PoH becomes more consequential. Adoption must accelerate in parallel with AI progress.

Importantly, the implementation should be correct from the outset: decentralized in structure, resilient to concentration of power, and aligned with individual empowerment rather than control. While it is by far the most advanced implementation, World ID still has a lot of room for improvement on these dimensions. Recent work is advancing World ID's Proof of Human further on these properties, including the request for comment for World ID 4.0.

If this is exciting to you, and you want to help build Proof of Human to maximally empower people in a world with increasingly capable machine intelligence, we invite you to take a look at Tools for Humanity's careers page.