

Contents

1	Introduction	2
2	The Challenge of Human Verification	2
3	Zero-Knowledge as an Enabler	3
3.1	How do Zero-Knowledge Proofs work in principle?	4
4	The Limitations of the Single-Identity Model	6

1 Introduction

Identity has always been a core part of the internet. In fact, there’s now a multi-billion dollar market around it [?] that offers personalized web experiences, from advertising networks that build detailed behavioral profiles to social platforms that tie participation to phone numbers.

Despite its scale, it remains inadequate, particularly in addressing the challenges posed by the increasing presence of bots as new consumers of the internet. With increasingly-capable Turing-complete AI agents, bots have become more difficult to distinguish from real users. Companies offering free server access intended for humans struggle with soaring compute and bandwidth costs. Know Your Customer Systems (KYC) were introduced to solve some of these problems. However, in many cases, they break some of the fundamental rules of secure software engineering: collect the minimum data required. There’s genuinely no need for a dating or alcohol retailing app to know my full government name, address and social security ID. Simply, verify I am an adult and not lying about my pictures and end the data collection process there. For privacy-lax users, you may be comfortable with numerous legally sound enterprises having access to all of your identity, but doing so only increases long-term exposure and liability. The “tea app” incident demonstrated how excessive personal data can become a privacy issue once it falls into the hands of bad actors.

Beyond privacy and compliance, there are other areas where existing identity systems fail. Bots are becoming increasingly capable, forcing us to think about how societies will shape with universal basic income. In such a world, verifying who is human will determine who can earn, participate, and access resources. This raises an important question: how can access to monetary and digital services remain exclusive to real humans while upholding the principle of least privilege?

Zero-knowledge proofs offer a new path by allowing individuals to prove their uniqueness or eligibility without disclosing raw identifiers, thus enhancing privacy and security in identity verification processes. Most proposed ZK-wrapped ID systems rest on a one person and one permanent identity model. This rigid structure cannot guarantee pseudonymity, leaves users vulnerable to coercion, and breaks under the pressure of large-scale deployment.

This white paper defines a plural identities framework, contrasts it with the incumbent single identity model, sets measurable criteria for privacy and robustness, and outlines practical directions for deployment and research. We also explore how this new framework balances accountability and freedom while preventing large-scale bot access without reverting to central control.

2 The Challenge of Human Verification

The emergence of advanced artificial intelligence entities necessitates a fundamental reevaluation of methods for verifying and distinguishing human users in digital spaces. Recent advancements in artificial general intelligence have enabled bots to emulate human behavior with growing sophistication, complicating the efforts of bot identification providers such as Cloudflare to discern humans from non-human activity online. These AI-driven agents can now autonomously establish accounts, generate substantial content, and interact online in ways that are increasingly indistinguishable from genuine human users, a development

that has contributed to phenomena such as the “dead internet theory” and the “digital imposter” dilemma. According to a 2025 cybersecurity report, automated bots accounted for more than 51% of all web traffic in 2024, illustrating the critical scope of the issue [?]. This proliferation poses a significant threat to the integrity of online environments and imposes considerable operational costs on platforms seeking to eliminate inauthentic users. In light of these challenges, it is imperative for us to develop robust, scalable systems for reliably differentiating human users from advanced bots, thereby safeguarding the authenticity of the online human experience.

Historically, digital platforms have employed mechanisms such as CAPTCHA, phone or SMS verification, and Know-Your-Customer (KYC) protocols to identify real humans. Nonetheless, these approaches present significant limitations: CAPTCHA, for example, is now often circumvented by advanced machine vision algorithms and operationalized click farms, rendering them less effective. Meanwhile, KYC procedures are criticized for their intrusiveness, frequently mandating the disclosure of extensive personal information—including legal names, addresses, and identification numbers—even in situations where only a single verification attribute, such as age or uniqueness, is required. This practice contravenes the principle of least privilege, which stipulates that systems should collect only the minimum data necessary for their specific function. For instance, social media platforms like Facebook require users to upload government-issued identification to recover locked accounts, even when the sole verification needed may be age or account ownership. As noted by Ethereum’s Vitalik Buterin, obligating individuals to disclose their entire legal identity to substantiate a single fact constitutes a “gross violation” of least privilege [5]. Unfortunately, many online services today do exactly that, amassing large identity datasets in the name of security or compliance.

Naturally, it did not take long for this noble quest for “security” to backfire spectacularly—the kind of poetic justice only the internet delivers.

Earlier this year (2025), the Tea social app—a “women-only” dating safety platform—suffered a breach that exposed 72,000 private images, including selfies and driver’s license ID scans. Users shared sensitive data for verification, but this KYC collection heightened their risk [4]. This raises a key question: how can we verify human attributes such as sex or age without creating risky data honeypots while making us indistinguishable from AI agents?

3 Zero-Knowledge as an Enabler

Zero-knowledge (ZK) proofs offer a promising solution to the human-verification dilemma. They enable individuals to authenticate essential characteristics, such as uniqueness and eligibility, without disclosing sensitive personal information [2]. They are implemented via cryptographic protocols that allow users to demonstrate compliance with predefined criteria (for instance, being from a specific country or satisfying an age or sex restriction) while preserving the confidentiality of their identity attributes, such as social security number, address, education, or children’s names. Rather than submitting traditional identification documents for each verification request, users instead generate a verifiable cryptographic proof that is assessed by the service provider.

Several projects, such as World ID (formerly Worldcoin) and Taiwan’s government-led

digital ID initiative, have gained traction implementing ZK-based ID protocols to verify user humanity or attributes using biometric data. World ID, in particular, uses an iris scan to establish uniqueness and then issues a cryptographic personhood credential that can be used “pseudonymously.” Currently, World ID has over 10 million registered users, mainly in the United States, and has been consistently growing, attracting the attention of the British government and the European Union because of the urgent need to verify age or citizenship with a digital ID under their new privacy laws [3].

3.1 How do Zero-Knowledge Proofs work in principle?

Firstly, users have to undergo an identification check. This might include biometric scans or government ID checks with a trusted issuer. Once the user has been verified as a unique entity, they(their device) store a cryptographic token that’s linked to their ID. The cryptographic token is the private key(never to be shared) and a paired public key(ZK Digital ID) is stored on a public registry. This registry could be on a blockchain or external database. Whenever an application wants to verify an individual, the user generates application specific identifiers using the secret(stored on device’s keychain) then uses a cryptographic attestation model to verify that the generated identifiers correspond to a verified user on the blockchain.

We can easily visualize the zero-knowledge verification pipeline in Figure 1.

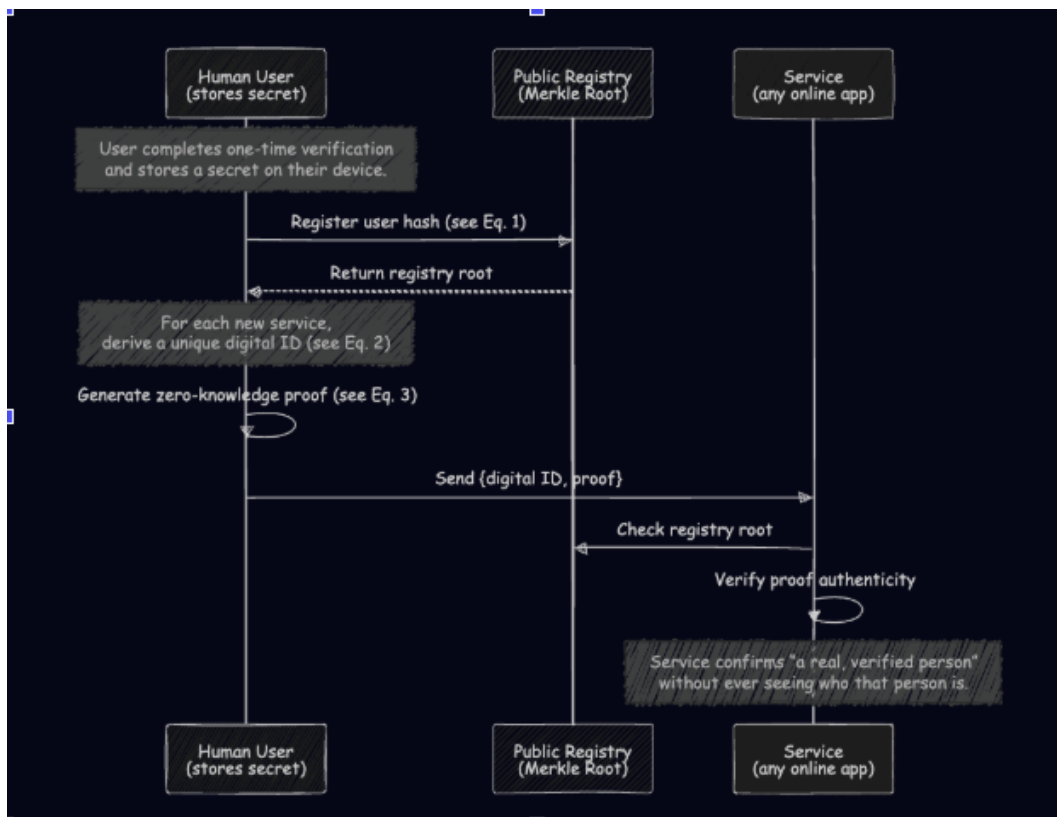


Figure 1: Diagrammatic process of zero-knowledge digital identity verification.

The process unfolds in three core phases: registration, identity derivation, and zero-knowledge verification. Each step enables the user to prove they are a real, previously verified human without revealing any sensitive identity attributes. The cryptographic operations behind this process are outlined below.

1. Registration to the Public Registry Once a user is verified (biometrically or via an official ID authority), a unique secret s is securely stored on their device. From this secret, the user computes a hash:

$$h = H(s) \quad (1)$$

This hashed commitment h is submitted to a public registry, typically implemented as a Merkle tree committed on a blockchain or decentralized network. The resulting Merkle root serves as the canonical record of all valid, verified human commitments.

2. Deriving Application-Specific Pseudonyms When the user interacts with a new service, they generate a service-specific pseudonym or digital identifier, unlinkable to other applications:

$$\text{ID}_{\text{app}} = H(s, \text{app_id}) \quad (2)$$

This process ensures contextual pseudonymity: each application sees a different identifier, preventing cross-service correlation or tracking, even by the identity system itself.

3. Zero-Knowledge Proof Generation and Attestation To authenticate with the service, the user generates a zero-knowledge proof π , attesting that:

- They know a secret s such that its hash $H(s)$ exists in the public registry (that is, it is a valid human commitment).
- The application-specific identifier ID_{app} is correctly derived from this same secret and the application's identifier.

The formal statement proven in zero-knowledge is:

$$\pi = \text{ZK-Proof}\left(s \mid H(s) \in \mathcal{R} \wedge \text{ID}_{\text{app}} = H(s, \text{app_id})\right) \quad (3)$$

Here, \mathcal{R} is the current Merkle root of the public registry.

This proof π and the pseudonymous identifier ID_{app} are then sent to the requesting application. The application independently verifies the proof's validity and ensures that:

- The commitment hash exists in the registry's current Merkle root.
- The proof was constructed using a valid witness (i.e. the secret s) without ever revealing it.

4. The Result is Verification Without Identity Exposure Upon successful verification, the service can confidently treat the user as a unique, real human, without learning who they are, their legal name, or any sensitive demographic information. This enables selective disclosure and privacy-preserving access control—critical for emerging contexts such as decentralized voting, on-chain benefits, or access to digital services requiring human-only participation.

This mechanism preserves unlinkability, minimal disclosure, and resistance to coercion, and forms the foundational architecture for private digital personhood in a post-password, bot-saturated internet.

4 The Limitations of the Single-Identity Model

Most proposed zero-knowledge wrapped digital identification systems adopt an implicit assumption: each human being is represented by a single, permanent, globally unique identifier. In practice, this “one person, one identifier” model is appealing because it simplifies many policy and engineering questions. It promises strong Sybil resistance, straightforward uniqueness guarantees, and an apparently clean mapping between legal identity and digital credentials. However, this simplicity comes at a structural cost. When examined as an infrastructural layer for the next generation of the internet, the single-identity model introduces severe limitations around privacy, coercion resistance, failure recovery, and long-term governance.

First, a single global identifier destroys contextual separation. If the same identifier or derivable token is reused across financial services, social media, voting platforms, and employment systems, it becomes trivial for state or corporate actors to correlate behavior across domains. This dramatically strengthens the already powerful tracking capabilities described in analyses of online profiling and advertising, where cross-site identifiers enable large-scale behavioral surveillance [1]. Even when the underlying identifier is wrapped in zero-knowledge, if the system design practically encourages repeated use of the same credential, the resulting pattern of proofs can still enable linkage at the application layer.

Second, the single-identity model is fragile under coercion. As Buterin argues in his critique of digital identification, even zero-knowledge protected credentials can be weaponized if adversaries can force users to prove statements on demand or to delegate control of their core identity secret [2]. A single canonical identity key becomes a pressure point: a government, employer, or abusive partner needs to compromise only that one secret (or the associated device) to demand a wide range of proofs, link otherwise separate personas, or block access to critical services. By design, a one-to-one identity architecture centralizes power in that single secret.

Third, single-identity systems concentrate technical risk. If the global identity secret is lost, the user faces a catastrophic failure mode: either they are permanently locked out of their digital life, or the system must implement highly privileged recovery pathways that reintroduce trusted intermediaries. If the secret is stolen, the attacker effectively becomes the user in all relying applications. Biometric-based personhood systems exacerbate this tension. Projects like World Coin and related personhood protocols tie uniqueness to one-time biometric enrollment [3]. Biometric traits cannot be rotated like passwords, so large-

scale compromise or subtle correlation bugs can have irreversible consequences for affected users.

Fourth, any scalable one-person-one-identifier scheme tends to reintroduce central authorities. Douceur’s classic result on the Sybil attack shows that, in open networks, it is impossible to prevent participants from creating many identities without assuming some form of centralized trust or out-of-band verification [?]. In practice, this means that global uniqueness claims rely on governments, large corporations, or tightly controlled enrollment devices. These authorities then control who is recognized as a “real” human, who can be excluded, and under what conditions revocation occurs. Such concentration of gatekeeping power conflicts with the plural, multi-jurisdictional reality of the internet.

Fifth, the single-identity model amplifies the impact of data breaches. Centralized or logically centralized registries that bind one canonical identity to rich personal attributes become high-value targets. Breaches like the Tea application incident, in which a women-only safety platform leaked thousands of sensitive images and identification documents, illustrate what happens when verification data is collected in bulk and later compromised reuters2025. A global digital identity system built around permanent, strongly linked records risks producing even larger, more damaging honeypots, especially where biometric and civil registry data are combined.

Finally, the single-identity model fails to reflect how humans actually live. Individuals inhabit many roles: citizen, employee, pseudonymous creator, activist, patient, and more. For many of these roles, it is both socially and politically important that they remain separable. Forcing all of them through one global digital identity channel not only chills participation in sensitive contexts, it also conflicts with the pluralistic vision of zero-knowledge based identification that emphasizes role-specific, context-specific credentials buterin2025plural. In short, the single-identity model optimizes for administrative convenience rather than human autonomy.

These structural limitations motivate the need for a different design space. Instead of binding each human to one canonical digital persona, a plural identity framework maintains strong guarantees of uniqueness and Sybil resistance while allowing individuals to hold and selectively reveal multiple unlinkable identities across contexts. The next section develops this plural identity model and shows how it can be realized using zero-knowledge friendly architectures.

References

- [1] Deighton, J., & Kornfeld, L. (2020, February). *The socioeconomic impact of internet tracking*. Interactive Advertising Bureau. <https://www.iab.com/wp-content/uploads/2022/02/The-Socio-Economic-Impact-of-Internet-Tracking.pdf>
- [2] Buterin, V. (2025, June 28). *Does digital ID have risks even if it's ZK-wrapped?* Vitalik.eth. <https://vitalik.eth.limo/general/2025/06/28/zkid.html>
- [3] Tools for Humanity. (2024, November). *TFH's World App passes 10 million users*. World. <https://world.org/blog/announcements/tfh-world-app-passes-10-million-users>
- [4] Reuters. (2025, July 26). *Women's dating app Tea reports 72,000 images stolen in a security breach*. Reuters. <https://www.reuters.com/sustainability/boards-policy-regulation/womens-dating-app-tea-reports-72000-images-stolen-security-breach-2025-07-26/>
- [5] Buterin, V. (2025, June 28). *Pluralistic ZK digital IDs are the best realistic solution to preserve privacy*. CryptoSlate. <https://cryptoslate.com/vitalik-buterin-says-pluralistic-zk-digital-ids-are-the-best-realistic-solution-to-p>