

Zero-Knowledge Plurality: Private and Robust Human Verification for the Next Internet

By Munachi Ernest-Eze

George Brown College

Blockchain Development WIP Paper

November 17, 2025

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	7
5	Plural Identities: Conceptual Framework	9
5.1	Design Goals and Threat Model	9
5.2	A Minimal Formal Model of Plural Identities	10
5.3	Cost Curves and Quadratic Scaling	10
5.4	Explicit and Implicit Plurality	11
5.5	Compatibility with Absolute Identity Requirements	12
5.6	Summary	13

1 Introduction

Identity has always been a core part of the internet. In fact, there's now a multi-billion dollar market around it [1] 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 divulging sensitive personal information [2]. They are implemented using cryptographic protocols that allow users to demonstrate compliance with predefined criteria (for example, 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 child’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 identification 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 an external database. Whenever an application wants to verify an individual, the user generates application specific identifiers using the secret(stored on the device 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.

There are three core phases, as described earlier: registration, identity derivation, and zero-knowledge verification. Each phase enables the user to prove that they are a real, previously verified human without revealing any sensitive identity attributes. In this subsection we make the process precise enough that it can be implemented directly.

We assume two fixed hash functions:

- An in-circuit hash H_{reg} (for example, Poseidon) used for commitments and the Merkle tree.
- An external hash H_{app} (for example, Secure Hash Algorithm number two hundred and fifty six or Blake three) used to derive application-facing identifiers.

1. Registration to the Public Registry Once a user is verified (biometrically or via an official identity authority), their device samples a secret seed

$$s \xleftarrow{\$} \{0, 1\}^\lambda, \quad (1)$$

and computes a commitment

$$h = H_{\text{reg}}(s). \quad (2)$$

This hashed commitment h is submitted to a public registry, typically implemented as a Merkle tree committed on a blockchain or decentralized network. The registry stores a set of leaves $\{h_i\}$ and publishes the Merkle root

$$\mathcal{R} = \text{MerkleRoot}(h_1, \dots, h_n), \quad (3)$$

which serves as the canonical record of all valid, verified human commitments.

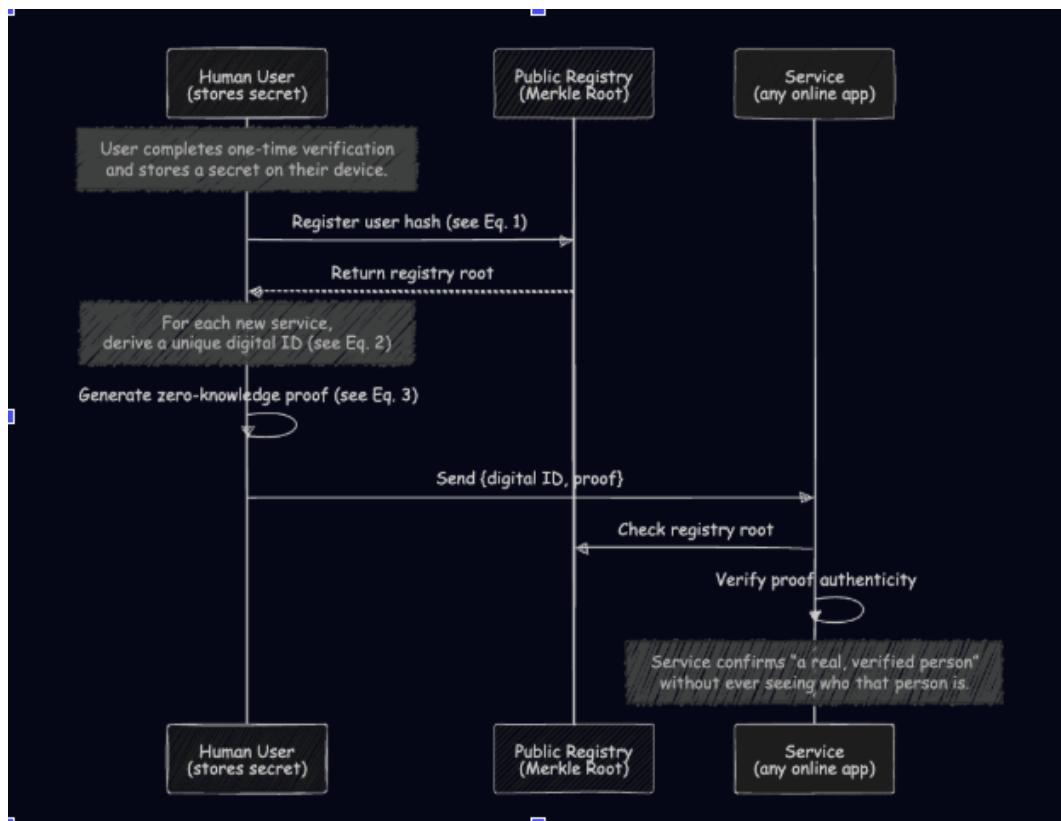


Figure 1: How zero-knowledge identity principle works

2. Deriving Application-Specific Pseudonyms When the user interacts with a new service identified by a byte string app_id , their device generates a service-specific pseudonym or digital identifier that is unlinkable to other applications:

$$\text{ID}_{\text{app}} = H_{\text{app}}(\text{"APP_ID"} \parallel s \parallel \text{app_id}), \quad (4)$$

where \parallel denotes byte concatenation and “APP_ID” is a domain-separation tag.

This construction ensures contextual pseudonymity: each application sees a different identifier, and even if two applications collude, they cannot link their views of the same user without learning the secret s .

3. Zero-Knowledge Proof Generation and Attestation To authenticate with a service, the user generates a zero-knowledge proof π over a simple statement.

The *public inputs* to the verifier are:

- the current Merkle root \mathcal{R} of the registry,
- the application identifier app_id ,
- the application pseudonym ID_{app} .

The *private witness* held only by the user device consists of:

- the secret s ,
- a Merkle authentication path path showing that $h = H_{\text{reg}}(s)$ is one of the leaves in the tree with root \mathcal{R} .

Inside the zero-knowledge circuit, the following checks are encoded as arithmetic constraints:

1. Recompute the leaf from the secret:

$$h' = H_{\text{reg}}(s). \quad (5)$$

2. Fold the Merkle authentication path to recover a root:

$$\mathcal{R}' = \text{MerkleRoot}(h', \text{path}). \quad (6)$$

3. Recompute the expected application pseudonym from the same secret:

$$\text{ID}'_{\text{app}} = H_{\text{app}}(\text{"APP_ID"} \parallel s \parallel \text{app_id}). \quad (7)$$

The circuit enforces the equalities

$$\mathcal{R}' = \mathcal{R} \quad \wedge \quad \text{ID}'_{\text{app}} = \text{ID}_{\text{app}}. \quad (8)$$

The formal statement proven in zero knowledge can be written as

$$\pi = \text{ZK-Proof}\left(s, \text{path} \mid \text{MerkleRoot}(H_{\text{reg}}(s), \text{path}) = \mathcal{R} \wedge \text{ID}_{\text{app}} = H_{\text{app}}(\text{"APP_ID"} \parallel s \parallel \text{app_id})\right). \quad (9)$$

The user sends the pair $(\text{ID}_{\text{app}}, \pi)$ to the application. The application then runs the verifier

$$b = \text{Verify}(\mathcal{R}, \text{app_id}, \text{ID}_{\text{app}}, \pi) \in \{0, 1\}. \quad (10)$$

If $b = 1$, the service accepts that there exists a registered human with secret s in the public registry such that the presented pseudonym ID_{app} is correctly derived from the same secret, without ever learning s or any other underlying identity attributes. ,

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.

5 Plural Identities: Conceptual Framework

The single digital identity model assumes a function

$$f : \mathcal{H} \rightarrow \mathcal{I}$$

from the set of humans \mathcal{H} to a set of digital identities \mathcal{I} , where each human $h \in \mathcal{H}$ is associated with exactly one globally recognized identifier $f(h)$. This model underlies many zero knowledge wrapped identity systems described in recent work by Buterin and others [2, 5]. It is attractive because it gives a clean notion of “one human, one account” for voting, universal basic income distributions, and anti-sybil protection. However, as Buterin argues, many of the deepest risks of digital identity are not solved by zero knowledge wrapping alone; they arise from this very one-identity-per-person property itself [5].

Plural identity intentionally relaxes that assumption. Instead of a single mapping f , each human h is allowed to maintain a set of digital identities

$$S(h) = \{i_{h,1}, i_{h,2}, \dots, i_{h,n(h)}\} \subseteq \mathcal{I},$$

possibly spanning different issuers, modalities, and cryptographic systems. The system does not attempt to collapse these into a single “true” identity. Instead, it focuses on two weaker, but more realistic, guarantees:

1. For given applications (for example, voting in a particular system), the influence a single human can exert through multiple identities is bounded by an explicit cost curve.
2. No fixed small set of credentials can plausibly be required by a coercer to reconstruct a complete global view of a person’s activity.

In other words, plural identity explicitly trades global uniqueness for a controlled many-to-many relation between humans and identities, with carefully designed economic and social friction.

5.1 Design Goals and Threat Model

We consider three interacting design goals, distilled from the arguments in [2, 5]:

- **Sybil resistance.** Large, well-funded actors should not be able to cheaply create an unbounded number of identities and dominate governance or drain universal distributions.
- **Pseudonymity.** Individual humans must be able to maintain multiple personas that are not easily linkable in practice, as a protection against social risk, harassment, and future political drift.
- **Coercion resistance;** It should be difficult for governments, employers, or abusive partners to coerce people into revealing a single secret that deanonymizes their entire online life.

In a plural system, adversaries can be of three broad types:

1. **Capital-rich adversaries**, who can invest significant money to generate many identities.
2. **Coercive adversaries**, who can force users to reveal secrets or hand over devices.
3. **Centralizing adversaries**, who attempt to make a single identity provider so dominant that refusing to use it becomes socially or economically impossible.

A single identity model is strong against capital-rich adversaries (at least in principle), but weak against coercive and centralizing ones. Plural identity aims for a more balanced equilibrium.

5.2 A Minimal Formal Model of Plural Identities

Let \mathcal{A} denote the set of applications (for example, social networks, on-chain voting systems, research studies, and payment platforms). For each human h and application $a \in \mathcal{A}$, we define a set of usable identities

$$S(h, a) \subseteq S(h),$$

representing the personas that h might legitimately present to a . For many systems, it is acceptable (and even desirable) to enforce that at most one identity per human is usable for a particular high-stakes application:

$$|S(h, a)| \leq 1.$$

However, crucially, that constraint is *local* to the application. It does not imply that there exists a global identity $i \in \mathcal{I}$ such that i is used everywhere.

From the perspective of a given application a , we primarily care about the induced equivalence classes on \mathcal{I} . Two identities i and i' are considered non-colluding for application a if the system cannot easily prove that they belong to the same human under its privacy assumptions. Plural identity aims to make it *expensive* for a single human to obtain many identities that are non-colluding from the perspective of a specific high-impact system, while still permitting multiple personas in other contexts.

5.3 Cost Curves and Quadratic Scaling

Following Buterin's argument [5], we model the cost for a human to obtain N usable identities in a given sybil-sensitive context as a function $C(N)$. Two qualitative constraints are desirable:

1. There should be no hard, easily legible cap on the number of identities. If everyone is known to have exactly one identity (or even a small fixed number), coercers can simply demand that they reveal them all.
2. The marginal cost of additional identities should grow faster than linearly, in order to dampen the advantage of large-scale attackers over small-scale participants.

A particularly clean target is a quadratic cost curve:

$$C(N) = \alpha N^2, \quad (11)$$

for some system-chosen scale parameter $\alpha > 0$. Under such a curve, a participant who wishes to control N identities must invest resources proportional to N^2 . This has two useful consequences:

- **Fairness across scales.** If a small participant controls one identity and a large participant controls N identities, then the large participant has invested αN^2 resources, whereas the small one has invested α . In governance settings where each identity carries roughly one unit of voting power, the marginal voting power per unit cost falls as N grows.
- **Protection of universal distributions.** In universal basic distribution scenarios (for example, token airdrops or universal basic services), the benefit of registering one more fake identity grows linearly with N , but the cost grows quadratically. Beyond some point, sybil creation ceases to be rational.

Plural identity does not require the cost function to be exactly quadratic in a strict mathematical sense, but it aims to approach this shape in practice. This can be realized by combining heterogeneous identity sources with different acquisition frictions: government passports, social graph attestations, biometric credentials, and reputation in specific communities [5]. Each additional high-quality identity often requires joining a new social or institutional domain, which naturally increases marginal cost.

5.4 Explicit and Implicit Plurality

Following [5], we can distinguish two ways plural identity arises:

Explicit plural identity. Systems that intentionally use overlapping social graphs and attestations to construct identity. For example, a protocol may require attestations from a set of peers who themselves must be well-connected to a global graph. The resulting identity weight is a function of graph connectivity rather than a single root authority.

Implicit plural identity. The de facto situation in today’s internet, where multiple identity providers coexist: email providers, government documents, social platforms, and device bound cryptographic wallets. No single one covers the entire population, and most applications accept several of them to reduce onboarding friction.

In explicit plural systems, we can formally model the identity weight of a persona i in an application a as a function

$$w_a(i) = F_a(G, i),$$

where G is a social graph of attestations. The exact form of F_a can encode sybil-resistant properties such as resistance to tightly knit collusion clusters and preference for identities embedded in diverse neighborhoods. The cost for an attacker to create N identities with

significant weight then depends on the difficulty of embedding those identities into the global graph in a natural way.

In implicit plural systems, the cost function emerges from the heterogeneity of providers. To create many meaningful identities, an attacker must:

- open bank accounts and credit histories,
- obtain passports or national documents in different jurisdictions,
- build social media accounts with organic looking activity,
- accumulate on-chain transaction histories that do not obviously correlate.

While none of these channels alone is sybil-proof, together they approximate a super-linear cost curve. Plural identity frameworks aim to formalize and strengthen this emergent behavior rather than collapse it into a single root identity that could be demanded, leaked, or compromised.

5.5 Compatibility with Absolute Identity Requirements

Plural identity does not deny that some domains require hard uniqueness. A clinical trial may need to guarantee that each physical human participates only once; a border control system may need to verify that a given traveler is the same person who was previously granted a visa; a high-risk financial compliance system may be legally obligated to bind transactions to a legal identity.

The key point is that these demands for absolute identity are *domain constrained*. We can model them as special applications a^* for which the policy explicitly enforces

$$|S(h, a^*)| = 1$$

and where the identity used for a^* is, by design, tightly linked to a legal or biometric anchor.

Plural identity frameworks therefore recommend two separation principles:

1. High-risk, high-authority systems may require one tightly anchored identity per human, but should be narrowly scoped and clearly separated from everyday interaction.
2. Most other systems, including social networks, content platforms, and low-stakes voting, should rely on plural identity primitives that preserve pseudonymity while enjoying sybil resistance through cost curves and graph structure.

The danger highlighted in [5] is precisely that one anchored system becomes universal, so that the identity used for border control or research enrollment becomes the same one silently used to log into every social platform. Plural identity frameworks treat this outcome as a failure mode, not a goal.

5.6 Summary

In summary, the plural identity framework replaces the idealized mapping

$$\text{one human} \longleftrightarrow \text{one global identity}$$

with a more nuanced structure:

- each human has a set of identities across different issuers and social contexts,
- applications locally decide how many identities per human are tolerable and how much weight each receives,
- a deliberately super-linear cost curve discourages mass fabrication of high-impact identities,
- no single credential is both necessary and sufficient to reconstruct the full picture of a person's online life.

This framework complements the cryptographic machinery of zero knowledge proofs. The cryptography ensures that specific claims are verifiable with minimal disclosure. Plural identity ensures that those claims are embedded in a social and economic structure that respects pseudonymity, resists coercion, and remains robust in the face of highly capable automated agents.

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>