What do I think about biometric proof of personhood?

Special thanks to the Worldcoin team, the Proof of Humanity community and Andrew Miller for discussion.

One of the trickier, but potentially one of the most valuable, gadgets that people in the Ethereum community have been trying to build is a decentralized proof-of-personhood solution. Proof of personhood, aka the "unique-human problem", is a limited form of real-world identity that asserts that a given registered account is controlled by a real person (and a different real person from every other registered account), ideally without revealing which

real person it is.

There have been a few efforts at tackling this problem: Proof of Humanity, BrightID, Idena and Circles come up as examples. Some of them come with their own applications (often a UBI token), and some have found use in Gitcoin Passport to verify which accounts are valid for quadratic voting. Zero-knowledge tech like Sismo adds privacy to many of these solutions. More recently, we have seen the rise of a much larger and more ambitious proof-of-personhood project: Worldcoin.

Worldcoin was co-founded by Sam Altman, who is best known for being the CEO of OpenAI. The philosophy behind the project is simple: AI is going to create a lot of abundance and wealth for humanity, but it also may kill very many people's jobs and make it almost impossible to tell who even is a human and not a bot, and so we need to plug that hole by (i) creating a really good proof-of-personhood system so that humans can prove that they actually are humans, and (ii) giving everyone a UBI. Worldcoin is unique in that it relies on highly sophisticated biometrics, scanning each user's iris using a piece of specialized hardware called "the Orb":

The goal is to produce a large number of these Orbs and widely distribute them around the world and put them in public places to make it easy for anyone to get an ID. To Worldcoin's credit, they have also committed to decentralize over time. At first, this means technical decentralization: being an L2 on Ethereum using the Optimism stack, and protecting users' privacy with ZK-SNARKs and other cryptographic techniques. Later on, it includes decentralizing governance of the system itself.

Worldcoin has been criticized for privacy and security concerns around the Orb, design issues in its "coin", and for ethical issues around some choices that the company has made. Some of the criticisms are highly specific, focusing on decisions made by the project that could easily have been made in another way - and indeed, that the Worldcoin project itself may be willing to change. Others, however, raise the more fundamental concern of whether or not biometrics - not just the eye-scanning biometrics of Worldcoin, but also the simpler face-video-uploads and verification games used in Proof of Humanity and Idena - are a good idea at all. And still others criticize proof of personhood in general Risks include unavoidable privacy leaks, further erosion of people's ability to navigate the internet anonymously, coercion by authoritarian governments, and the potential impossibility of being secure at the same time as being decentralized.

This post will talk about these issues, and go through some arguments that can help you decide whether or not bowing down and scanning your eyes (or face, or voice, or...) before our new spherical overlords is a good idea, and whether or not the natural alternatives - either using social-graph-based proof of personhood or giving up on proof of personhood entirely - are any better.

What is proof of personhood and why is it important?

The simplest way to define a proof-of-personhood system is: it creates a list of public keys where the system guarantees that each key is controlled by a unique human. In other words, if you're a human, you can put one key on the list, but you can't put two keys on the list, and if you're a bot you can't put any keys on the list.

Proof of personhood is valuable because it solves a lot of anti-spam and anti-concentration-of-power problems that many people have, in a way that avoids dependence on centralized authorities and reveals the minimal information possible. If proof of personhood is not solved, decentralized governance (including "micro-governance" like votes on social media posts) becomes much easier to capture by very wealthy actors, including hostile governments. Many services would only be able to prevent denial-of-service attacks by setting a price for access, and sometimes a price high enough to keep out attackers is also too high for many lower-income legitimate users.

Many major applications in the world today deal with this issue by using government-backed identity systems such as credit

cards and passports. This solves the problem, but it makes large and perhaps unacceptable sacrifices on privacy, and can be trivially attacked by governments themselves.

How many proof of personhood proponents see the two-sided risk that we are facing[Image source](#).

In many proof-of-personhood projects - not just Worldcoin, but also Proof of Humanity, Circles and others - the "flagship application" is a built-in "N-per-person token" (sometimes called a "UBI token"). Each user registered in the system receives some fixed quantity of tokens each day (or hour, or week). But there are plenty of other applications:

- Airdrops for token distributions
- Token or NFT sales that[give more favorable terms to less-wealthy users](#)
- [Voting in DAOs](#)
- A way to "seed" graph-based reputation systems
- [Quadratic voting](#) (and funding, and attention payments)
- Protection against bots / [sybil attacks](#) in social media
- An alternative to captchas for preventing [DoS attacks](#)

In many of these cases, the common thread is a desire to create mechanisms that are open and democratic, avoiding both centralized control by a project's operators and domination by its wealthiest users. The latter is [especially important in decentralized governance](#). In many of these cases, existing solutions today rely on some combination of (i) highly opaque AI algorithms that leave lots of room to undetectably discriminate against users that the operators simply do not like, and (ii) centralized IDs, aka "KYC". An effective proof-of-personhood solution would be a much better alternative, achieving the security properties that those applications need without the pitfalls of the existing centralized approaches.

What are some early attempts at proof of personhood?

There are two main forms of proof of personhood: social-graph-based

and biometric

. Social-graph based proof of personhood relies on some form of vouching: if Alice, Bob, Charlie and David are all verified humans, and they all say that Emily is a verified human, then Emily is probably also a verified human. Vouching is often enhanced with incentives: if Alice says that Emily is a human, but it turns out that she is not, then Alice and Emily may both get penalized. Biometric proof of personhood involves verifying some physical or behavioral trait of Emily, that distinguishes humans from bots (and individual humans from each other). Most projects use a combination of the two techniques.

The four systems I mentioned at the beginning of the post work roughly as follows:

- [Proof of Humanity](#)

: you upload a video of yourself, and provide a deposit. To be approved, an existing user needs to vouch for you, and an amount of time needs to pass during which you can be challenged. If there is a challenge, a [Kleros decentralized court](#) determines whether or not your video was genuine; if it is not, you lose your deposit and the challenger gets a reward.

- [BrightID](#)

: you join a video call "verification party" with other users, where everyone verifies each other. Higher levels of verification are available via [Bitu](#), a system in which you can get verified if enough other Bitu-verified users vouch for you.

- [Idena](#)

: you play a captcha game at a specific point in time (to prevent people from participating multiple times); part of the captcha game involves creating and verifying captchas that will then be used to verify others.

- [Circles](#)

: an existing Circles user vouches for you. Circles is unique in that it does not attempt to create a "globally verifiable ID"; rather, it creates a graph of trust relationships, where someone's trustworthiness can only be verified from the perspective of your own position in that graph.

How does Worldcoin work?

Each Worldcoin user installs an app on their phone, which generates a private and public key, much like an Ethereum wallet. They then go in-person to visit an "Orb". The user stares into the Orb's camera, and at the same time shows the Orb a QR code generated by their Worldcoin app, which contains their public key. The Orb scans the user's eyes, and uses complicated hardware scanning and machine-learned classifiers to verify that:

1. The user is a real human
2. The user's iris does not match the iris of any other user that has previously used the system

If both scans pass, the Orb signs a message approving a specialized hash of the user's iris scan. The hash gets uploaded to a database - currently a centralized server, intended to be replaced with a decentralized on-chain system once they are sure the hashing mechanism works. The system does not store full iris scans; it only stores hashes, and these hashes are used to check for uniqueness. From that point forward, the user has a "World ID".

A World ID holder is able to prove that they are a unique human by generating a ZK-SNARK proving that they hold the private key corresponding to a public key in the database, without revealing which

key they hold. Hence, even if someone re-scans your iris, they will not be able to see any actions that you have taken.

What are the major issues with Worldcoin's construction?

There are four major risks that immediately come to mind:

- Privacy

. The registry of iris scans may reveal information. At the very least, if someone else

scans your iris, they can check it against the database to determine whether or not you have a World ID. Potentially, iris scans might reveal more information.

- Accessibility

. World IDs are not going to be reliably accessible unless there are so many Orbs that anyone in the world can easily get to one.

- Centralization

. The Orb is a hardware device, and we have no way to verify that it was constructed correctly and does not have backdoors. Hence, even if the software layer is perfect and fully decentralized, the Worldcoin Foundation still has the ability to insert a backdoor into the system, letting it create arbitrarily many fake human identities.

- Security

. Users' phones could be hacked, users could be coerced into scanning their irises while showing a public key that belongs to someone else, and there is the possibility of 3D-printing "fake people" that can pass the iris scan and get World IDs.

It's important to distinguish between (i) issues specific to choices made by Worldcoin, (ii) issues that any biometric proof of personhood

will inevitably have, and (iii) issues that any proof of personhood in general

will have.

For example, signing up to Proof of Humanity means publishing your face on the internet. Joining a BrightID verification

party doesn't quite

do that, but still exposes who you are to a lot of people. And joining Circles publicly exposes your social graph. Worldcoin is significantly better

at preserving privacy than either of those. On the other hand, Worldcoin depends on specialized hardware, which opens up the challenge of trusting the orb manufacturers to have constructed the orbs correctly - a challenge which has no parallels in Proof of Humanity, BrightID or Circles. It's even conceivable that in the future, someone other than Worldcoin will create a different

specialized-hardware solution that has different tradeoffs.

How do biometric proof-of-personhood schemes address privacy issues?

The most obvious, and greatest, potential privacy leak that any proof-of-personhood system has is linking each action that a person takes to a real-world identity. This data leak is very large, arguably unacceptably large, but fortunately it is easy to solve with zero knowledge proof

technology. Instead of directly making a signature with a private key whose corresponding public key is in the database, a user could make a ZK-SNARK proving that they own the private key whose corresponding public key is somewhere

in the database, without revealing which specific key they have. This can be done generically with tools like Sismo (see here for the Proof of Humanity-specific implementation), and Worldcoin has its own built-in implementation. It's important to give "crypto-native" proof of personhood credit here: they actually care about taking this basic step to provide anonymization, whereas basically all centralized identity solutions do not.

A more subtle, but still important, privacy leak is the mere existence

of a public registry of biometric scans. In the case of Proof of Humanity, this is a lot of data: you get a video of each Proof of Humanity participant, making it very clear to anyone in the world who cares to investigate who all the Proof of Humanity participants are. In the case of Worldcoin, the leak is much more limited: the Orb locally computes and publishes only a "hash" of each person's iris scan. This hash is not a regular hash like SHA256; rather, it is a specialized algorithm based on machine-learned Gabor filters that deals with the inexactness inherent in any biometric scan, and ensures that successive hashes taken of the same person's iris have similar outputs.

Blue: percent of bits that differ between two scans of the same person's iris. Orange: percent of bits that differ between two scans of two different people's irises.

These iris hashes leak only a small amount of data. If an adversary can forcibly (or secretly) scan your iris, then they can compute your iris hash themselves, and check it against the database of iris hashes to see whether or not you participated in the system. This ability to check whether or not someone signed up is necessary for the system itself to prevent people from signing up multiple times, but there's always the possibility that it will somehow be abused. Additionally, there is the possibility that the iris hashes leak some amount of medical data (sex, ethnicity, perhaps medical conditions), but this leak is far smaller than what could be captured by pretty much any other mass data-gathering system in use today (eg. even street cameras). On the whole, to me the privacy of storing iris hashes seems sufficient.

If others disagree with this judgement and decide that they want to design a system with even more privacy, there are two ways to do so:

1. If the iris hashing algorithm can be improved to make the difference between two scans of the same person much lower (eg. reliably under 10% bit flips), then instead of storing full iris hashes, the system can store a smaller number of error correction bits for iris hashes (see: fuzzy extractors). If the difference between two scans is under 10%, then the number of bits that needs to be published would be at least 5x less.
2. If we want to go further, we could store the iris hash database inside a multi-party computation (MPC) system which could only be accessed by Orbs (with a rate limit), making the data unaccessible entirely, but at the cost of significant protocol complexity and social complexity in governing the set of MPC participants. This would have the benefit that users would not be able to prove a link between two different World IDs that they had at different times even if they

wanted to.

Unfortunately, these techniques are not applicable to Proof of Humanity, because Proof of Humanity requires the full video of each participant to be publicly available so that it can be challenged if there are signs that it is fake (including AI-generated fakes), and in such cases investigated in more detail.

On the whole, despite the "dystopian vibez" of staring into an Orb and letting it scan deeply into your eyeballs, it does seem like specialized hardware systems can do quite a decent job of protecting privacy. However, the flip side of this is that specialized hardware systems introduce much greater centralization concerns. Hence, we cypherpunks seem to be stuck in a bind: we have to trade off one deeply-held cypherpunk value against another.

What are the accessibility issues in biometric proof-of-personhood systems?

Specialized hardware introduces accessibility concerns because, well, specialized hardware is not very accessible. Somewhere between 51% and 64% of sub-Saharan Africans now have smartphones, and this seems to be projected to increase to 87% by 2030. But while there are billions of smartphones, there are only a few hundred Orbs. Even with much higher-scale distributed manufacturing, it would be hard to get to a world where there's an Orb within five kilometers of everyone

.

But to the team's credit, they have been trying!

It is also worth noting that many other

forms of proof of personhood have accessibility problems that are even worse. It is very difficult to join a social-graph-based proof-of-personhood system unless you already know someone who is in the social graph. This makes it very easy for such systems to remain restricted to a single community in a single country.

Even centralized

identity systems have learned this lesson: India's Aadhaar ID system is biometric-based, as that was the only way to quickly onboard its massive population while avoiding massive fraud from duplicate and fake accounts (resulting inhuge cost savings), though of course the Aadhaar system as a whole is far weaker on privacy than anything being proposed on a large scale within the crypto community.

The best-performing systems from an accessibility perspective are actually systems like Proof of Humanity

, which you can sign up to using only a smartphone - though, as we have seen and as we will see, such systems come with all kinds of other tradeoffs.

What are the centralization issues in biometric proof-of-personhood systems?

There are three:

1. Centralization risks in the system's top-level governance (esp. the system that makes final top-level resolutions if different actors in the system disagree on subjective judgements).
2. Centralization risks unique to systems that use specialized hardware.
3. Centralization risks if proprietary algorithms are used to determine who is an authentic participant.

Any proof-of-personhood system must contend with (1), perhaps with the exception of systems where the set of "accepted" IDs is completely subjective. If a system uses incentives denominated in outside assets (eg. ETH, USDC, DAI), then it cannot be fully subjective, and so governance risks become unavoidable.

[2] is a much bigger risk for Worldcoin than for Proof of Humanity (or BrightID), because Worldcoin depends on specialized hardware and other systems do not.

[3] is a risk particularly in "logically centralized" systems where there is a single system doing the verification, unless all of the algorithms are open-source and we have an assurance that they are actually running the code that they claim they are. For systems that rely purely on users verifying other users (like Proof of Humanity), it is not a risk.

How does Worldcoin address hardware centralization issues?

Currently, a Worldcoin-affiliated entity called Tools for Humanity is the only organization that is making Orbs. However, the Orb's source code is mostly public: you can see the hardware specs in this github repository, and other parts of the source code are expected to be published soon. The license is another one of those "shared source but not technically open source until four years from now" licenses similar to the Uniswap BSL, except in addition to preventing forking it also prevents what they consider unethical behavior - they specifically list mass surveillance and three international civil rights declarations.

The team's stated goal is to allow and encourage other organizations to create Orbs, and over time transition from Orbs being created by Tools for Humanity to having some kind of DAO that approves and manages which organizations can make Orbs that are recognized by the system.

There are two ways in which this design can fail:

1. It fails to actually decentralize

. This could happen because of the common trap of federated protocols: one manufacturer ends up dominating in practice, causing the system to re-centralize. Presumably, governance could limit how many valid Orbs each manufacturer can produce, but this would need to be managed carefully, and it puts a lot of pressure on governance to be both decentralized and

monitor the ecosystem and respond to threats effectively: a much harder task than eg. a fairly static DAO that just handles top-level dispute resolution tasks.

1. It turns out that it's not possible to make such a distributed manufacturing mechanism secure

. Here, there are two risks that I see: * Fragility against bad Orb manufacturers

: if even one Orb manufacturer is malicious or hacked, it can generate an unlimited number of fake iris scan hashes, and give them World IDs.

- Government restriction of Orbs

: governments that do not want their citizens participating in the Worldcoin ecosystem can ban Orbs from their country. Furthermore, they could even force their citizens to get their irises scanned, allowing the government to get their accounts, and the citizens would have no way to respond.

1. Fragility against bad Orb manufacturers

: if even one Orb manufacturer is malicious or hacked, it can generate an unlimited number of fake iris scan hashes, and give them World IDs.

1. Government restriction of Orbs

: governments that do not want their citizens participating in the Worldcoin ecosystem can ban Orbs from their country. Furthermore, they could even force their citizens to get their irises scanned, allowing the government to get their accounts, and the citizens would have no way to respond.

- Fragility against bad Orb manufacturers

: if even one Orb manufacturer is malicious or hacked, it can generate an unlimited number of fake iris scan hashes, and give them World IDs.

- Government restriction of Orbs

: governments that do not want their citizens participating in the Worldcoin ecosystem can ban Orbs from their country. Furthermore, they could even force their citizens to get their irises scanned, allowing the government to get their accounts, and the citizens would have no way to respond.

To make the system more robust against bad Orb manufacturers, the Worldcoin team is proposing to perform regular audits on Orbs, verifying that they are built correctly and key hardware components were built according to specs and were not tampered with after the fact. This is a challenging task: it's basically something like the IAEA nuclear inspections bureaucracy but for Orbs. The hope is that even a very imperfect implementation of an auditing regime could greatly cut down on the number of fake Orbs.

To limit the harm caused by any bad Orb that does

slip through, it makes sense to have a second mitigation. World IDs registered with different Orb manufacturers, and ideally with different Orbs, should be distinguishable from each other

. It's okay if this information is private and only stored on the World ID holder's device; but it does need to be provable on demand. This makes it possible for the ecosystem to respond to (inevitable) attacks by removing individual Orb manufacturers, and perhaps even individual Orbs, from the whitelist on-demand. If we see the North Korea government going around and forcing people to scan their eyeballs, those Orbs and any accounts produced by them could be immediately retroactively disabled.

Security issues in proof of personhood in general

In addition to issues specific to Worldcoin, there are concerns that affect proof-of-personhood designs in general. The major ones that I can think of are:

1. 3D-printed fake people

: one could use AI to generate photographs or even 3D prints of fake people that are convincing enough to get accepted by the Orb software. If even one group does this, they can generate an unlimited number of identities.

1. Possibility of selling IDs

: someone can provide someone else's public key instead of their own when registering, giving that person control of their registered ID, in exchange for money. This seems to be happening already. In addition to selling, there's also the possibility of renting IDs

to use for a short time in one application.

1. Phone hacking

: if a person's phone gets hacked, the hacker can steal the key that controls their World ID.

1. Government coercion to steal IDs

: a government could force their citizens to get verified while showing a QR code belonging to the government. In this way, a malicious government could gain access to millions of IDs. In a biometric system, this could even be done covertly: governments could use obfuscated Orbs to extract World IDs from everyone entering their country at the passport control booth.

[1] is specific to biometric proof-of-personhood systems. [2] and [3] are common to both biometric and non-biometric designs. [4] is also common to both, though the techniques that are required would be quite different in both cases; in this section I will focus on the issues in the biometric case.

These are pretty serious weaknesses. Some already have been addressed in existing protocols, others can be addressed with future improvements, and still others seem to be fundamental limitations.

How can we deal with fake people?

This is significantly less of a risk for Worldcoin than it is for Proof of Humanity-like systems: an in-person scan can examine many features of a person, and is quite hard to fake, compared to merely deep-faking a video. Specialized hardware is inherently harder to fool than commodity hardware, which is in turn harder to fool than digital algorithms verifying pictures and videos that are sent remotely.

Could someone 3D-print something that can fool even specialized hardware eventually? Probably. I expect that at some point we will see growing tensions between the goal of keeping the mechanism open and keeping it secure: open-source AI algorithms are inherently more vulnerable to adversarial machine learning. Black-box algorithms are more protected, but it's hard to tell that a black-box algorithm was not trained to include backdoors. Perhaps ZK-ML technologies could give us the best of both worlds. Though at some point in the even further future, it is likely that even the best AI algorithms will be fooled by the best 3D-printed fake people.

However, from my discussions with both the Worldcoin and Proof of Humanity teams, it seems like at the present moment neither protocol is yet seeing significant deep fake attacks, for the simple reason that hiring real low-wage workers to sign up on your behalf is quite cheap and easy

.

Can we prevent selling IDs?

In the short term, preventing this kind of outsourcing is difficult, because most people in the world are not even aware of proof-of-personhood protocols, and if you tell them to hold up a QR code and scan their eyes for $30 they will do that. Once more people are

aware of what proof-of-personhood protocols are, a fairly simple mitigation becomes possible: allowing people who have a registered ID to re-register, canceling the previous ID

. This makes "ID selling" much less credible, because someone who sells you their ID can just go and re-register, canceling the ID that they just sold. However, getting to this point requires the protocol to be very

widely known, and Orbs to be very

widely accessible to make on-demand registration practical.

This is one of the reasons why having a UBI coin integrated into a proof-of-personhood system is valuable: a UBI coin provides an easily understandable incentive for people to (i) learn about the protocol and sign up, and (ii) immediately re-register if they register on behalf of someone else

. Re-registration also prevents phone hacking.

Can we prevent coercion in biometric proof-of-personhood systems?

This depends on what kind of coercion we are talking about. Possible forms of coercion include:

- Governments scanning people's eyes (or faces, or...) at border control and other routine government checkpoints, and using this to register (and frequently re-register) their citizens
- Governments banning Orbs within the country to prevent people from independently re-registering
- Individuals buying IDs and then threatening to harm the seller if they detect that the ID has been invalidated due to re-registration
- (Possibly government-run) applications requiring people to "sign in" by signing with their public key directly, letting them see the corresponding biometric scan, and hence the link between the user's current ID and any future IDs they get from re-registering. A common fear is that this makes it too easy to create "permanent records" that stick with a person for their entire life.

All your UBI and voting power are belong to us. Image source.

Especially in the hands of unsophisticated users, it seems quite tough to outright prevent these situations

. Users could leave their country to (re-)register at an Orb in a safer country, but this is a difficult process and high cost. In a truly hostile legal environment, seeking out an independent Orb seems too difficult and risky.

What is

feasible is making this kind of abuse more annoying to implement

and detectable

. The Proof of Humanity approach of requiring a person to speak a specific phrase when registering is a good example: it may be enough to prevent hidden

scanning, requiring coercion to be much more blatant, and the registration phrase could even include a statement confirming that the respondent knows that they have the right to re-register independently and may get UBI coin or other rewards. If coercion is detected, the devices used to perform coercive registrations en masse could have their access rights revoked. To prevent applications linking people's current and previous IDs and attempting to leave "permanent records", the default proof of personhood app could lock the user's key in trusted hardware, preventing any application from using the key directly without the anonymizing ZK-SNARK layer in between. If a government or application developer wants to get around this, they would need to mandate the use of their own custom app.

With a combination of these techniques and active vigilance, locking out those regimes that are truly hostile, and keeping honest those regimes that are merely medium-bad (as much of the world is), seems possible. This can be done either by a project like Worldcoin or Proof of Humanity maintaining its own bureaucracy for this task, or by revealing more information about how an ID was registered (eg. in Worldcoin, which Orb it came from), and leaving this classification task to the community.

Can we prevent renting

IDs (eg. to sell votes)?

Renting

out your ID is not prevented by re-registration. This is okay in some applications: the cost of renting out your right to collect the day's share of UBI coin is going to be just the value of the day's share of UBI coin. But in applications such as voting, easy vote selling is a huge problem.

Systems like MACI can prevent you from credibly selling your vote, by allowing you to later cast another vote that invalidates your previous vote, in such a way that no one can tell whether or not you in fact cast such a vote. However, if the briber controls which key you get at registration time

, this does not help.

I see two solutions here:

1. Run entire applications inside an MPC

. This would also cover the re-registration process: when a person registers to the MPC, the MPC assigns them an ID that is separate from, and not linkable to, their proof of personhood ID, and when a person re-registers, only the MPC would know which account to deactivate. This prevents users from making proofs about their actions, because every important step is done inside an MPC using private information that is only known to the MPC.

1. Decentralized registration ceremonies

. Basically, implement something like this in-person key-registration protocol that requires four randomly selected local participants to work together to register someone. This could ensure that registration is a "trusted" procedure that an attacker cannot snoop in during.

Social-graph-based systems may actually perform better here, because they can create local decentralized registration processes automatically as a byproduct of how they work.

How do biometrics compare with the other leading candidate for proof of personhood, social graph-based verification?

Aside from biometric approaches, the main other contender for proof of personhood so far has been social-graph-based verification. Social-graph-based verification systems all operate on the same principle: if there are a whole bunch of existing verified identities that all attest to the validity of your identity, then you probably are valid and should also get verified status.

If only a few real users (accidentally or maliciously) verify fake users, then you can use basic graph-theory techniques to put an upper bound on how many fake users get verified by the system. Source: https://www.sciencedirect.com/science/article/abs/pii/S0045790622000611.

Proponents of social-graph-based verification often describe it as being a better alternative to biometrics for a few reasons:

- It does not rely on special-purpose hardware

, making it much easier to deploy

- It avoids a permanent arms race

between manufacturers trying to create fake people and the Orb needing to be updated to reject such fake people

- It does not require collecting biometric data

, making it more privacy-friendly

- It is potentially more friendly to pseudonymity

, because if someone chooses to split their internet life across multiple identities that they keep separate from each other, both of those identities could potentially be verified (but maintaining multiple genuine and separate identities sacrifices network effects and has a high cost, so it's not something that attackers could do easily)

- Biometric approaches give a binary score

of "is a human" or "is not a human", which is fragile: people who are accidentally rejected would end up with no UBI at all, and potentially no ability to participate in online life. Social-graph-based approaches can give a more nuanced numerical score

, which may of course be moderately unfair to some participants but is unlikely to "un-person" someone completely.

My perspective on these arguments is that I largely agree with them! These are genuine advantages of social-graph-based approaches and should be taken seriously. However, it's worth also taking into account the weaknesses of social-graph-based approaches:

- Bootstrapping

: for a user to join a social-graph-based system, that user must know someone who is already in the graph. This makes large-scale adoption difficult, and risks excluding entire regions of the world that do not get lucky in the initial bootstrapping process.

- Privacy

: while social-graph-based approaches avoid collecting biometric data, they often end up leaking info about a person's social relationships, which may lead to even greater risks. Of course, zero-knowledge technology can mitigate this (eg. see this proposal by Barry Whitehat), but the interdependency inherent in a graph and the need to perform mathematical analyses on the graph makes it harder to achieve the same level of data-hiding that you can with biometrics.

- Inequality

: each person can only have one biometric ID, but a wealthy and socially well-connected person could use their connections to generate many IDs. Essentially, the same flexibility that might allow a social-graph-based system to give multiple pseudonyms to someone (eg. an activist) that really needs that feature would likely also imply that more powerful and well-connected people can gain more pseudonyms than less powerful and well-connected people.

- Risk of collapse into centralization

: most people are too lazy to spend time reporting into an internet app who is a real person and who is not. As a result, there is a risk that the system will come over time to favor "easy" ways to get inducted that depend on centralized authorities, and the "social graph" that the system users will de-facto become the social graph of which countries recognize which people as citizens - giving us centralized KYC with needless extra steps.

Is proof of personhood compatible with pseudonymity in the real world?

In principle, proof of personhood is compatible with all kinds of pseudonymity. Applications could be designed in such a way that someone with a single proof of personhood ID can create up to five profiles within the application, leaving room for pseudonymous accounts. One could even use quadratic formulas: N accounts for a cost of $N². But will they?

A pessimist, however, might argue that it is naive to try to create a more privacy-friendly form of ID and hope that it will actually get adopted in the right way, because the powers-that-be are not privacy-friendly, and if a powerful actor gets a tool that could

be used to get much more information about a person, they will

use it that way. In such a world, the argument goes, the only realistic

approach is, unfortunately, to throw sand in the gears of any

identity solution, and defend a world with full anonymity and digital islands of high-trust communities.

I see the reasoning behind this way of thinking, but I worry that such an approach would, even if successful, lead to a world where there's no way for anyone to do anything to counteract wealth concentration and governance centralization, because one person could always pretend to be ten thousand. Such points of centralization would, in turn, be easy for the powers-that-be to capture. Rather, I would favor a moderate approach, where we vigorously advocate for proof-of-personhood solutions to have strong privacy, potentially if desired even include a "N accounts for $N²" mechanism at protocol layer, and create something that has privacy-friendly values and

has a chance of getting accepted by the outside world.

So... what do I think?

There is no ideal form of proof of personhood. Instead, we have at least three different paradigms of approaches that all have their own unique strengths and weaknesses. A comparison chart might look as follows:

What we should ideally do is treat these three techniques as complementary, and combine them all

. As India's Aadhaar has shown at scale, specialized-hardware biometrics have their benefits of being secure at scale. They are very weak at decentralization, though this can be addressed by holding individual Orbs accountable. General-purpose biometrics can be adopted very easily today, but their security is rapidly dwindling, and they may only work for another 1-2 years. Social-graph-based systems bootstrapped off of a few hundred people who are socially close to the founding team are likely to face constant tradeoffs between completely missing large parts of the world and being vulnerable to attacks within communities they have no visibility into. A social-graph-based system bootstrapped off tens of millions of biometric ID holders, however, could actually work. Biometric bootstrapping may work better short-term, and social-graph-based techniques may be more robust long-term, and take on a larger share of the responsibility over time as their algorithms improve.

A possible hybrid path.

All of these teams are in a position to make many mistakes, and there are inevitable tensions between business interests

and the needs of the wider community, so it's important to exercise a lot of vigilance. As a community, we can and should push all participants' comfort zones on open-sourcing their tech, demand third-party audits and even third-party-written software, and other checks and balances. We also need more alternatives in each of the three categories.

At the same time it's important to recognize the work already done: many of the teams running these systems have shown a willingness to take privacy much more seriously than pretty much any government or major corporate-run identity systems, and this is a success that we should build on.

The problem of making a proof-of-personhood system that is effective and reliable, especially in the hands of people distant from the existing crypto community, seems quite challenging. I definitely do not envy the people attempting the task, and it will likely take years to find a formula that works. The concept of proof-of-personhood in principle seems very valuable, and while the various implementations have their risks, not having any proof-of-personhood at all has its risks too: a world with no proof-of-personhood seems more likely to be a world dominated by centralized identity solutions, money, small closed communities, or some combination of all three. I look forward to seeing more progress on all types of proof of personhood, and hopefully seeing the different approaches eventually come together into a coherent whole.