

# **Who Watches the Watchmen?**

A Review of Subjective Approaches for  
Sybil-resistance in Proof of Personhood Protocols

## Who Watches the Watchmen?

~~A Review of Subjective Approaches for  
Sybil resistance in Proof of Personhood Protocols~~

a review of approaches to ensure that  
**each unique human only gets one identity in a domain**

**Why This Matters?**

# Universal Basic Income

# Peer-to-peer democracies

# **Who Watches the Watchmen?**

A Review of Subjective Approaches for  
Sybil-resistance in Proof of Personhood Protocols

# **Who Watches the Watchmen?**

A Review of Subjective Approaches for  
Sybil-resistance in Proof of Personhood Protocols

# Who Verifies the Verifier?

A Review of Subjective Approaches for  
Sybil-resistance in Proof of Personhood Protocols



# social media identity

**social media identity: the bright side**

# social media identity: the bright side

- **Identity is the secret sauce of Facebook, Twitter, Google:** orgs established themselves by achieving a sufficient level of consensus over their identity credentials.

# social media identity: the bright side

- **Identity is the secret sauce of Facebook, Twitter, Google:** orgs established themselves by achieving a sufficient level of consensus over their identity credentials.
- **Only former alternative able to reach such widespread use pertained to nation-states:** passports, licenses, and national ID cards. The monopoly was broken.

# social media identity: the bright side

- **Identity is the secret sauce of Facebook, Twitter, Google:** orgs established themselves by achieving a sufficient level of consensus over their identity credentials.
- **Only former alternative able to reach such widespread use pertained to nation-states:** passports, licenses, and national ID cards. The monopoly was broken.
- Global identification system outside of the strict control of nation-states has created a **networked social infrastructure, a new trust layer for society**

# social media identity: the bright side

- **Identity is the secret sauce of Facebook, Twitter, Google:** orgs established themselves by achieving a sufficient level of consensus over their identity credentials.
- **Only former alternative able to reach such widespread use pertained to nation-states:** passports, licenses, and national ID cards. The monopoly was broken.
- Global identification system outside of the strict control of nation-states has created a **networked social infrastructure, a new trust layer for society**
- **A new kind of participative politics,** with multiple, borderless, political and social movements, completely changing national and international politics over the past few years.

# **social media identity**

# **social media identity: the dark side**



# social media identity: the dark side

- **algorithms have an upper hand over humans:** private corporations turned prediction of behaviour into influence of behaviour. Emergence of Surveillance Capitalism.

# social media identity: **the dark side**

- **algorithms have an upper hand over humans:** private corporations turned prediction of behaviour into influence of behaviour. Emergence of Surveillance Capitalism.
- centralized architecture exposes society to **political manipulation.**

# social media identity: **the dark side**

- **algorithms have an upper hand over humans:** private corporations turned prediction of behaviour into influence of behaviour. Emergence of Surveillance Capitalism.
- centralized architecture exposes society to **political manipulation.**
- “likes” rarely have institutional impact, only serve editorial purposes:  
**illusion of participation may be worse than no participation**



"The one vulnerability being exploited across  
all systems is **Identity**"

Edward Snowden — Web3 2019 (Berlin)

# **Who Watches the Watchmen?**

A Review of Subjective Approaches for  
Sybil-resistance in Proof of Personhood Protocols

# Who Watches the Watchmen?

A Review of Subjective Approaches for  
**Sybil-resistance** in Proof of Personhood Protocols

# The Sybil Attack

## The Sybil Attack

John R. Douceur  
Microsoft Research  
johndo@microsoft.com

*“One can have, some claim, as many electronic personas as one has time and energy to create.”*

– Judith S. Donath [12]

**Abstract** – Large-scale peer-to-peer systems face security threats from faulty or hostile remote computing elements. To resist these threats, many such systems employ redundancy. However, if a single faulty entity can present multiple identities, it can control a substantial fraction of the system, thereby undermining this redundancy. One approach to preventing these “Sybil attacks” is to have a trusted agency certify identities. This paper shows that, without a logically centralized authority, Sybil attacks are always possible except under extreme and unrealistic assumptions of resource parity and coordination among entities.

### 1. Introduction

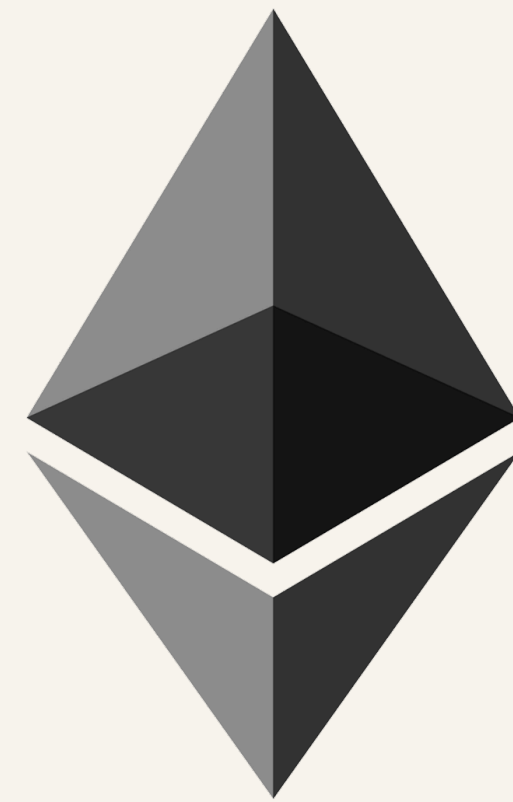
We\* argue that it is practically impossible, in a distributed computing environment, for initially

If the local entity has no direct physical knowledge of remote entities, it perceives them only as informational abstractions that we call *identities*. The system must ensure that distinct identities refer to distinct entities; otherwise, when the local entity selects a subset of identities to redundantly perform a remote operation, it can be duped into selecting a single remote entity multiple times, thereby defeating the redundancy. We term the forging of multiple identities a *Sybil attack* [30] on the system.

It is tempting to envision a system in which established identities vouch for other identities, so that an entity can accept new identities by trusting the collective assurance of multiple (presumably independent) signatories, analogous to the PGP web of trust [37] for human entities. However,

The abuse of an online, **peer-to-peer system** by creating many illegitimate virtual personas.

# proof of work & proof of stake



- resource-based sybil protection mechanisms
- proof of work = **one-CPU-one-vote**
- proof of stake = **one-dollar-one-vote**
- **inherently plutocratic**



# Who Watches the Watchmen?

A Review of Subjective Approaches for  
Sybil-resistance in **Proof of Personhood Protocols**

# Proof of Personhood

- **each unique human only gets one identity in a domain**
- first generation of solutions mostly emerged in **2019**
- **objective incentive - UBI**
- instead of \$ or CPU, **human input (subjective)**

# Proof of Personhood

**easy for humans, hard for AI**

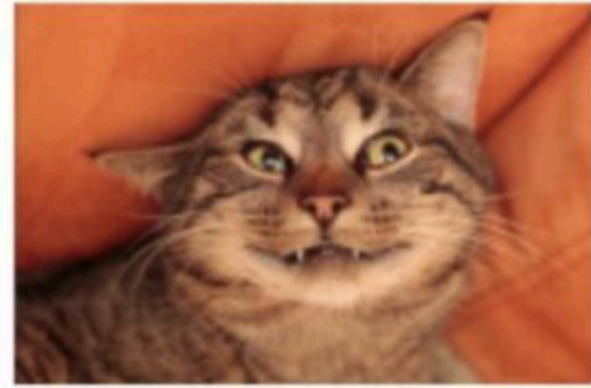
# Proof of Personhood

**easy for humans to join once,  
hard or impossible to join twice**

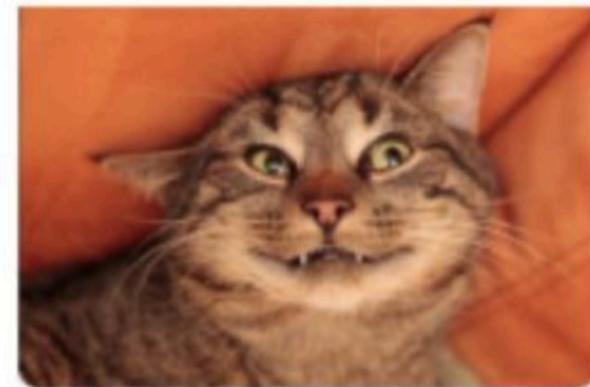
**how can we distinguish unique  
humans from duplicates or AI?**

# Reverse Turing Tests

Meaningful  
story



Meaningless  
story



"The vase was knocked over by the cat, so the puddle had to be cleaned with a sponge."

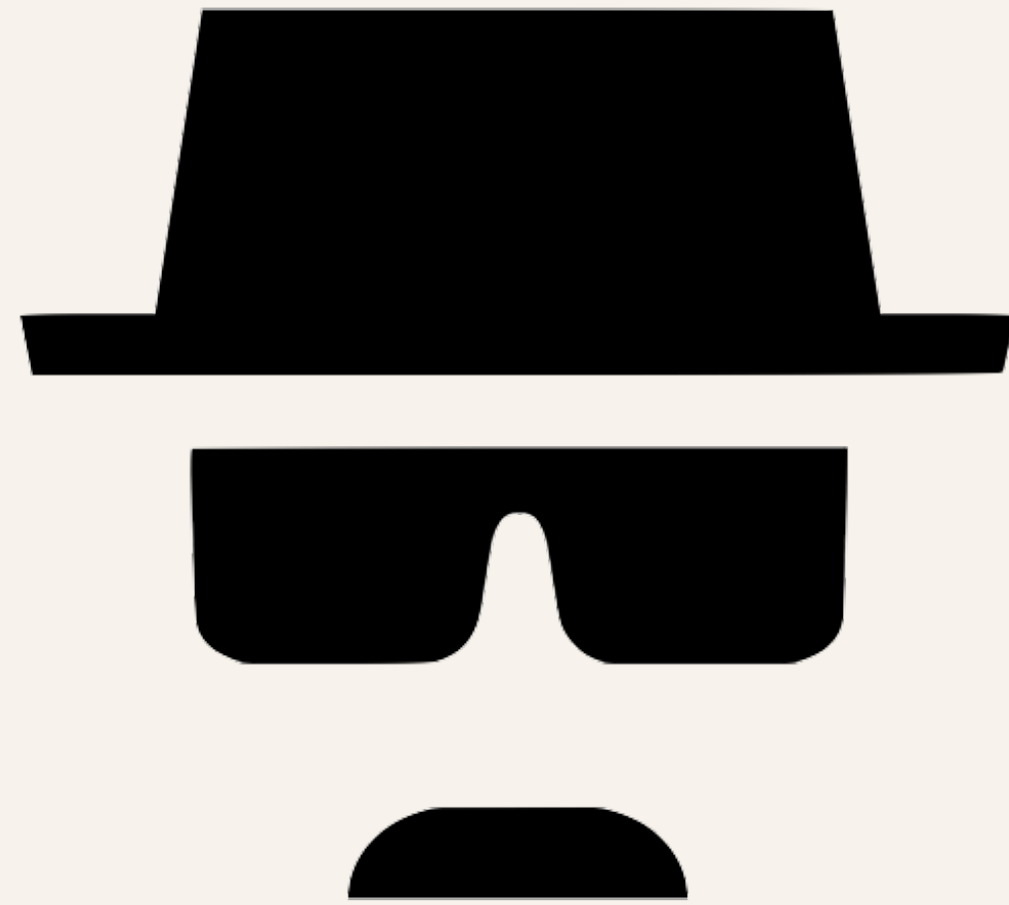
This sequence conveys no discernible storyline.

# Reverse Turing Tests

- AI-hard
- Created by humans
- Based on images
- Great strategy against AI attacks, but doesn't prevent human attacks



# Pseudonym Parties



**each human can only be  
in one place at a time**

# Idena Network — Synchronous Reverse Turing Tests.

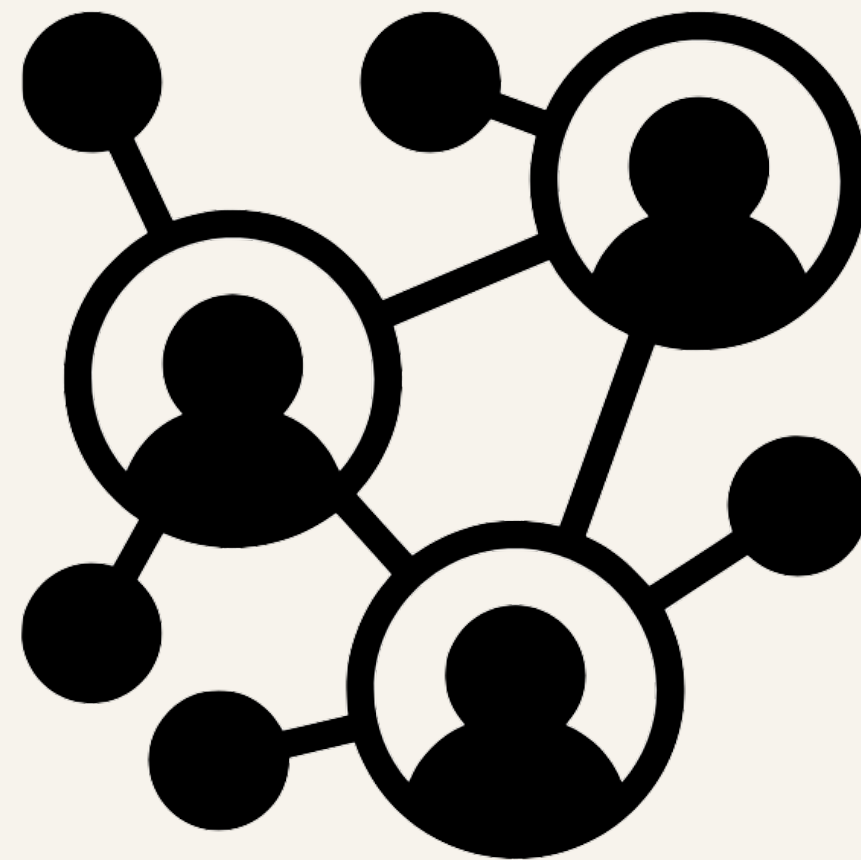


I D E N A

Idena implements a synchronous event held over the entire network where participants are required to solve Turing tests that are hard for Machine Learning systems to solve.

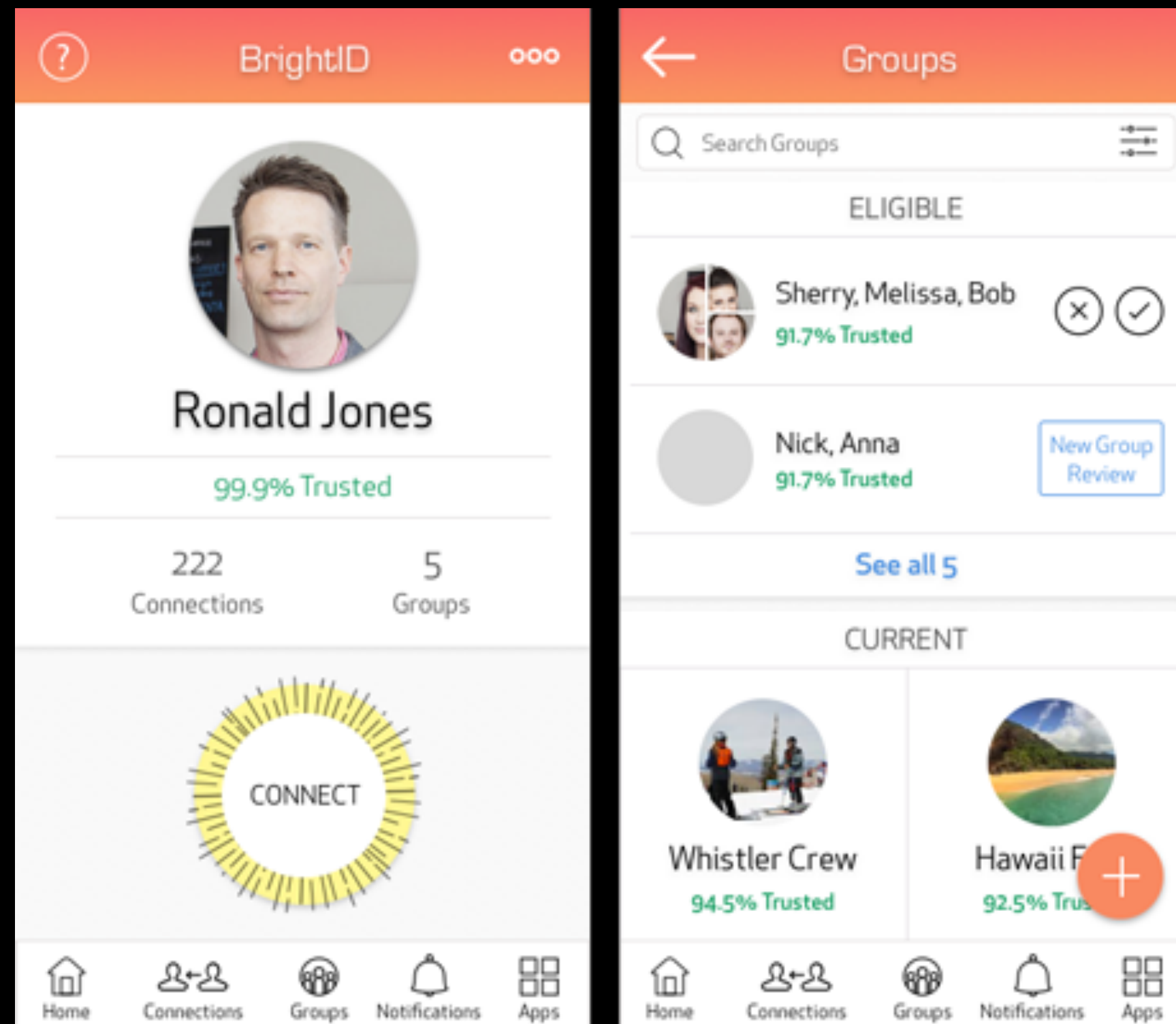
This provides a proof of personhood assuming the tests cannot be captured by existing AI.

# Web of Trust



**establishes proof of personhood by  
connection to trusted seed identities**

# BrightID — Graph based sybil detection.

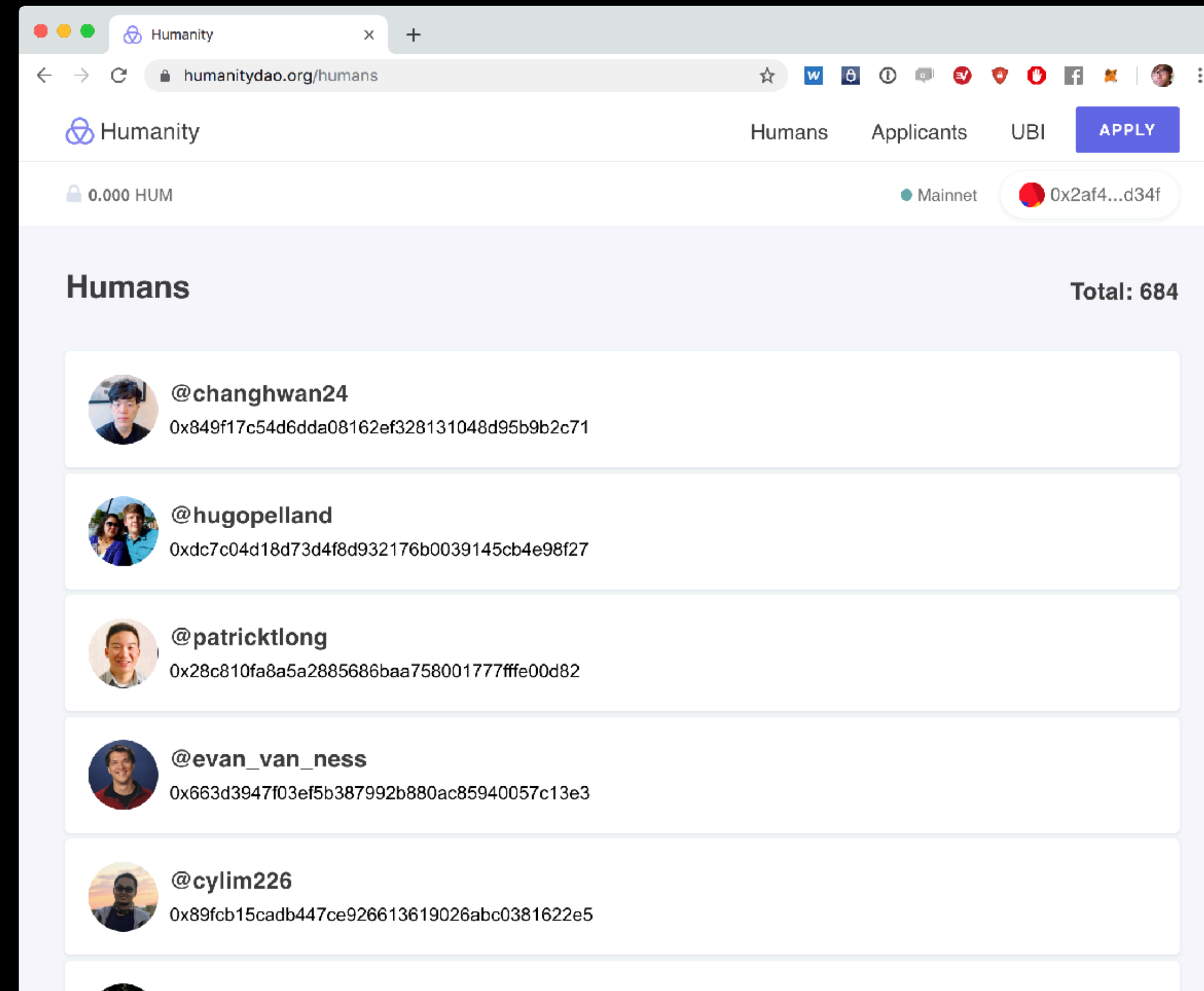


BrightID uses graph analysis of our social connections, with the additional input of a limited number of trusted seed identities. The interconnectivity of the graph should reveal which members are real and which are fake (aka "Sybils"), based on their position in relation to the trusted seeds.

# Token Curated Registries

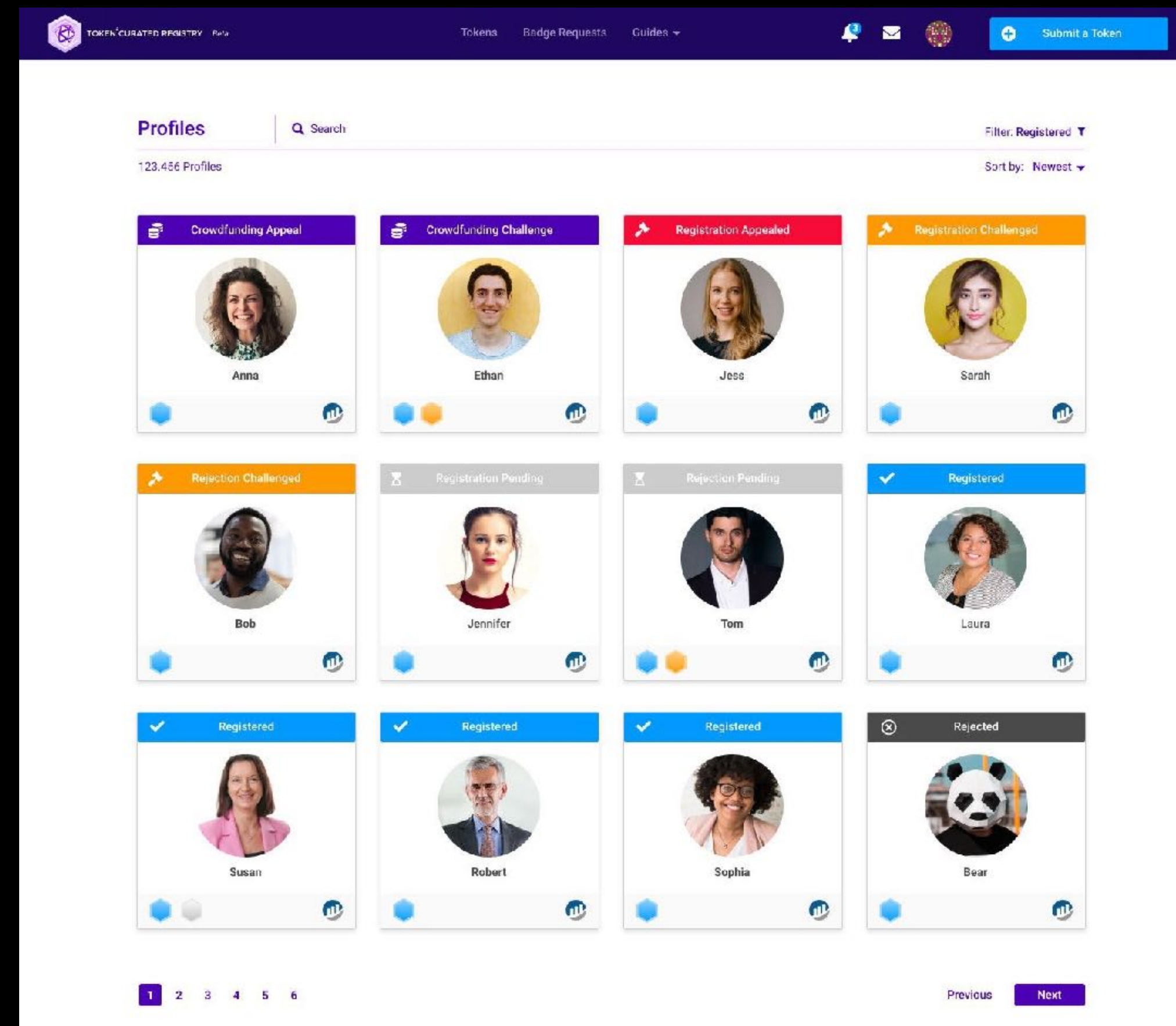
**Token Curated Registries aim to create economic incentives for decentralized list curation. Members of a list hold tokens associated with the list, which may increase in value if they are able to maintain its quality, legitimacy or popularity.**

# HumanityDAO — Token Curated Registry.



Stake to approve or challenge identities based on using Twitter Profiles as proof.

# Kleros — Web of Trust TCR with Video Proofs.



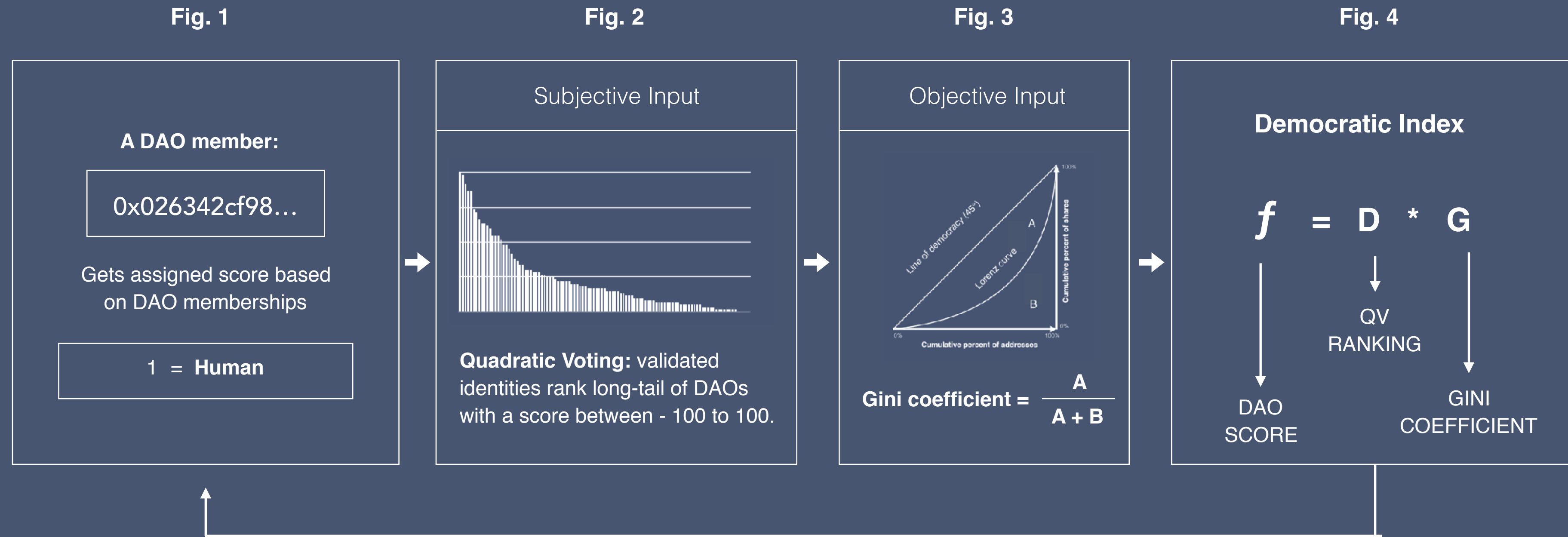
Use a Kleros TCR that randomly elects jurors that verify video of candidate IDs.

# DAOs

**Decentralized Autonomous Organizations. DAOs are a class of smart contract devised to automate the execution of organizational governance and fund allocation.**



# Equality protocol — Democracy Earth



**Fig. 1** Every address that belongs to a DAO will be weighted by with a percentage of the Democratic Index obtained from the intersection of DAOs that constitute the identity of such address.

**Fig. 2** In order to counter weight false positives on the Gini Coefficient, addresses that achieve a high PASS score will be granted the right to rank the different DAOs analyzed by the PASS oracle, according the their corresponding ability to ensure that no single Human controls more than one identifier within its domain.

**Fig. 3** The Gini coefficient for democracy ranges from 0 to 1, with 0 representing perfect totalitarianism and 1 representing perfect democracy. It computes a score that measures the share distribution by each segment of addresses belonging to a DAO.

**Fig. 4** A Democratic Index is calculated for each DAO, as a function of their position in the Quadratic Voting rank, and their Gini Coefficient.

# Upala — Price-of-Personhood.



- Upala provides a digital identity uniqueness score, represented by a price for which an identity can be sold at any time.
- Implements the social responsibility where members are incentivized to only invite trusted members to their DAOs.
- By “exploding” an identity, one must expect that the costs to rejoin the group will increase, creating an additional protection to groups over time.

# Takeaways

- There is still much work to be done!

# Takeaways

- There is still much work to be done!
- No "best approach".

# Who Watches the Watchmen?

A Review of **Subjective** Approaches for  
Sybil-resistance in Proof of Personhood Protocols

Humans are the best at recognizing other humans.



Paula Berman

@\_paulaberman

[paulamlberman@gmail.com](mailto:paulamlberman@gmail.com)