

Sybil Identity Farms

NETWORK	Base mainnet (chain 8453)
---------	---------------------------

LICENSE	CC BY 4.0
---------	-----------

SEVERITY	HIGH
----------	------

Summary

On-chain agent identity is cheap to mint and trivial to clone. On the ERC-8004 registry deployed to Base mainnet (chain 8453) we find that **roughly a quarter of all registered agents belong to “identity farms”**: single owner addresses that mint hundreds or thousands of agents sharing one identical `agentURI`. Our corroborated detector flags **44** such farms, with registry state reconstructed as of block 48,425,370, the last indexed registry block before 10 July 2026 (00:00 UTC).

The important finding is not that farms exist. It is that **raw agent count does not distinguish a farm from a legitimate at-scale platform**. A legitimate registrar and a farm can both emit thousands of Registered events. What separates them is a feature that costs nothing to read but is expensive to fake at organic pace: **metadata diversity**.

Background

ERC-8004 lets anyone register an agent identity: an `agentId`, an `owner` address, and an `agentURI` pointing to the agent’s metadata. Registration is permissionless and near-free. That is by design, and it is also the attack surface. An identity that is cheap to create is cheap to create *en masse*.

Downstream systems (payment rails, agent marketplaces, reputation layers) increasingly treat “has an ERC-8004 identity” as a weak trust signal. A single actor controlling thousands of identities can manufacture the appearance of a diverse, populated ecosystem: inflate agent counts, seed reputation between its own agents, or present many “independent” counterparties that are in fact one entity.

The pattern

An **identity farm** is a single owner address that registers a large number of agents which are not meaningfully distinct from one another. In the clearest cases every agent points to the **same agentURI** : the metadata was written once and cloned across the whole batch.

The naive detector (“rank owners by agent count”) is wrong, and dangerously so: it surfaces the ecosystem’s most valuable legitimate participants (registrars that onboard agents on behalf of many real clients) while a moderately-sized farm sits below the threshold.

The discriminator: metadata diversity, not volume

The feature that actually separates the two is the ratio of **distinct agentURI values** to agents under one owner. A legitimate registrar gives each agent its own metadata; a farm reuses one template. On the live registry the largest legitimate registrars and the largest farms overlap almost completely in *size* and diverge completely in *signature*. Side by side, the top owners make the point better than any threshold:

OWNER (ANONYMIZED)	AGENTS	DISTINCT URIS	URI DIVERSITY	READING
Registrar A	ten thousand plus	one per agent	near-total	Legitimate at-scale registrar
Registrar B	over a thousand	one per agent	near-total	Legitimate
Registrar C	over a thousand	almost one per agent	near-total	Legitimate
Farm A	2,289	1	near-zero	Farm
Farm B	1,931	1	near-zero	Farm
Farm C	1,683	1	near-zero	Farm
Farm D	over a thousand	1	near-zero	Farm

The largest single owner on the registry is **not** the biggest problem: it is a legitimate platform assigning every agent a unique identity. The farms sit just below it in volume and are invisible to a count-only filter. Owner identifiers are withheld here pending legal review; they are reproducible from public chain data with the published method.

Naive heuristic versus corroborated detector

A first pass at this problem is a raw threshold: owners above some agent count whose distinct-URI ratio sits below some floor. That heuristic is directionally right but noisy, and its exact count swings with where the two knobs are set. It is not the number we publish.

The **44** farms above come from a corroborated detector. It requires the scale signal to be backed by at least one independent piece of evidence (near-zero metadata diversity, or an inter-registration cadence too fast and too regular for human onboarding) before an owner is flagged at all. The corroborated count is the smaller, harder-to-dispute figure, and it is the one we stand behind. When a headline number can be inflated by loosening a threshold, we publish the tighter one.

Why it matters

For anyone consuming ERC-8004 identity as a trust input:

- **Counterparty risk.** A payment rail settling an agent-to-agent transaction cannot assume its counterparty is an independent actor. One farm owner can present thousands of “distinct” agents.
- **Reputation laundering.** Farmed identities are the raw material for manufactured reputation: an owner’s agents can vouch for one another to fabricate a track record. (See our reputation pattern.)
- **Metric pollution.** Ecosystem dashboards that count “agents” overcount reality by roughly a quarter on this registry today.

Detection (recognition, not replication)

The signal is a first-order, cheap-to-read discriminator computed entirely from public registry data:

1. Group Registered events by `owner`.
2. For each owner, compute agent count and the count of **distinct** `agentURI` values.
3. A high agent count with near-zero URI diversity is the farm signature; high diversity at scale is a legitimate registrar. We flag only when raw scale is corroborated by low diversity or mechanical cadence.

We deliberately publish the *discriminator*, not an evasion recipe. We also note its ceiling below.

Limitations & honest caveats

- **This is a first-order signal.** A sophisticated operator can defeat “distinct-URI implies legitimate” by generating unique URIs per agent at an organic pace. Distinct URIs raise the cost of faking; they are not proof of legitimacy. Deeper signals (cross-owner metadata similarity, funding-source graphs, registration-burst timing) are where this goes next, and they stay a generation ahead of what we publish.
- **“Not flagged” is not “trustworthy.”** Absence of a farm signature means no farm signal, not a positive trust attestation.
- **Recognition, not accusation.** Owner addresses are public on-chain facts. We report the pattern and the method; we do not infer intent beyond what the data shows.

Reproducibility

Every figure here is derived from public ERC-8004 events on Base mainnet and is independently verifiable against the registry, with state reconstructed as of block 48,425,370 (the last indexed registry block before 10 July 2026, 00:00 UTC). Methodology and thresholds are published above; detection code is open (Apache 2.0).

Verifeth monitors the integrity of the on-chain agent economy. Every published figure is frozen at a stated block and independently reproducible from public data.