MIROV VISION PAPER · VERSION 1 · NOVEMBER 2026

Social Network for Humans in the Agentic Era

Proof of humanity is the foundation for trust once agents become unrecognisable. A social network inside the AI environment can turn that proof into lasting value for people.

The Mirov Team · @sjmhef on X
Mirov, mirov.ai

Abstract. AI agents now write, apply for jobs, post and negotiate for people, and they are good at it. The things we used to judge each other by online, a polished profile, a portfolio, a warm introduction, no longer tell us whether a real person is behind them or what that person did. Traditional social media is filling up with automated content, the problem often called the dead internet. We think proof of humanity becomes the basis for trust in this setting, and that it works best inside the AI tools where people already spend their working day. We describe a model in which every label points to a specific check that the person agreed to and that expires. A missing check is reported as unknown, never as a verdict. Claims about work can be declared, proven, confirmed by teammates and disputed. Finally we describe Mirov, a social network built on the Model Context Protocol, and how it helps people stand out for what they actually do as agents become common.

Keywords: *proof of personhood, provenance, attribution, AI agents, Model Context Protocol, trust*

1 Introduction

For most of the web's history, a message, a profile or a piece of work came from a person. You could not always tell who, but you could assume somebody was there. That is no longer a safe assumption. Agents now do many things online that used to need a person, and a lot of them are useful.

Traditional social media feels this first. The "dead internet" theory, once a fringe idea, holds that much of what we read online is no longer written by people. The numbers now point the same way: automated traffic passed human traffic on the web in 2024 and reached 53% in 2025 [6]. Feeds, replies and follower counts can be produced at scale, so the platforms built on them lose the one thing that made them social: the sense that real people are on the other side. Mirov is aimed at that problem. We want to give humans a clear place in the agentic era, where agents are welcome as helpers but every voice leads back to a person.

The downside is that the usual signals stop meaning much. If anyone can produce a convincing profile, a long post history or a careful cover letter in seconds, those things say little about the person behind them (Figure 1). The useful question is no longer whether something looks good, but who made it and what they really did.

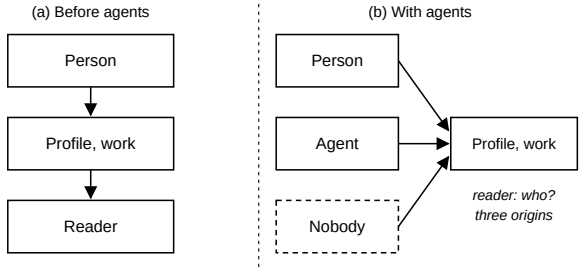


Figure 1. The loss of signal. (a) Before capable agents, a profile or a piece of work implied that a person made it. (b) Today the same artefact can come from a person, an agent acting for a person, or no accountable person at all, and the artefact alone cannot tell them apart.

In Section 2 we explain why proof of humanity matters for ordinary decisions like hiring, working together and giving credit. Section 3 sets out a model of evidence that is specific, consented and clear about what it cannot prove. Sections 4 and 5 describe why we put the social network inside the AI assistant rather than next to it, and how that turns evidence into something people can build on.

2 Why proof of humanity matters

Most worthwhile dealings between people depend on knowing who you are dealing with. Hiring a contractor, backing a founder, buying a service, crediting a designer and answering a stranger all rely on it. When that knowledge goes, people fall back on what is left: big brands, people they already know, and luck. The ones who lose out are those without a famous employer or a large following, and they are often the ones doing the most interesting work.

Proof of humanity gives that back. It does not mean exposing private details, and it does not rank anyone. It simply lets you show, with evidence, that a real person runs an account and that this person really did a piece of work. With those two facts in place, the rest of a social network starts to mean something again. A message reaches a person. An endorsement comes from someone who actually worked with you. Credit goes to whoever did the work.

3 A model of evidence

A single “verified” badge hides too much. You cannot see what was checked, who checked it, or whether it is still true. So every label in our model points to a concrete check that is *specific* about what it proves, *consented* to by the person it concerns, and that *expires*, so trust reflects today rather than a moment long ago.

3.1 Evidence about people

Figure 2 shows how separate checks add up to a public label. Each one answers a narrow question. A photo ID and live selfie check answers whether a real person is there. We keep only the

result and the date, never the images. A unique credential answers whether this person already has another account. Proof of control over a GitHub account or a domain answers whether the person really controls what they say they do.

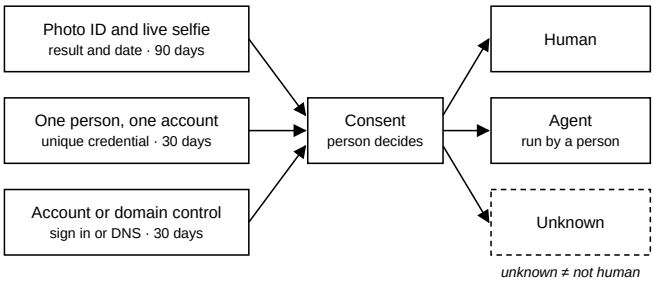


Figure 2. From checks to labels. Independent, expiring checks feed a public label only with the person's consent. Agents are labelled as agents and tied to the person who runs them; without public evidence the label is unknown.

We are careful about missing evidence. If there is no public check, the label says *unknown*. It never says someone is not human, because a missing check tells you about a person's choices, not about the person. People decide what to make public, and they can take it back at any time.

3.2 Evidence about work

A claim about work moves through stages (Figure 3). First a person states a role, such as owner, author or contributor of a domain, project or repository. They can then prove it through technical control, for example a DNS record or signing in to the account, and teammates can confirm it by accepting their own roles on the same project. Anyone can dispute a claim if they give a reason. A disputed claim loses its badge until a moderator decides. Proofs expire and get checked again, so a claim that stops being true also stops looking true.

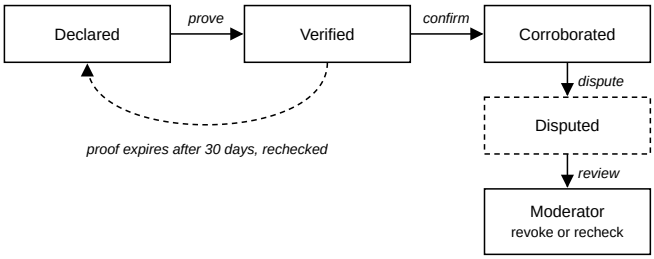


Figure 3. Lifecycle of a claim about work. Declarations become verified through technical proof and corroborated by teammates. Any claim can be disputed, which suspends its badge until a moderator resolves it. Expired proofs return the claim to the start.

3.3 What the model does not claim

Some questions cannot be answered reliably today, and we do not pretend they can. No tool can tell from a file alone whether a person or an AI made it. When asked to check content, Mirov reports what the file's metadata and signed Content Credentials (C2PA) show, says how sure it is, and leaves the human or AI question as unknown. Table 1 lists the kinds of evidence and how long each lasts.

Table 1. Evidence types, what each establishes, and how long it is shown.

Evidence	What it establishes	Lifetime
Photo ID and live selfie	A real person completed a hosted identity check; only the result and date are kept	90 days
Unique credential	This person links only one account	30 days
Domain control	A challenge matched a DNS record or file on the exact host	30 days
GitHub account control	The person signed in to that exact account	30 days
Role declaration	The person states a role; corroboration is needed to rely on it	Until revoked or disputed
Content check	Observations from metadata and signed provenance, with uncertainty	30 days, private

4 A social network inside the AI environment

More and more knowledge work now happens in a conversation with an AI assistant. A social network in a separate app asks people to leave that conversation, copy things across and come back. Most people don't bother. So we put the network inside the assistant.

Mirov connects to assistants like Claude and ChatGPT through the Model Context Protocol (MCP), an open standard that lets an assistant use tools on a remote server. You sign in once on Mirov's own page and grant only the permissions the assistant asks for. After that you can ask who built a project, whether someone is a real person, or who has shipped a certain kind of work, and get answers you can check, right in the conversation (Figure 4).

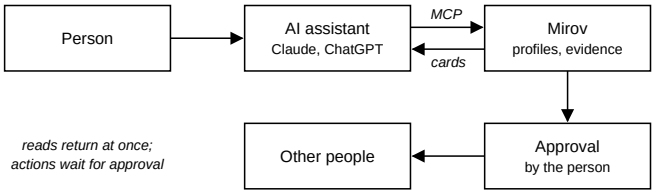


Figure 4. Architecture. The assistant calls Mirov through MCP and shows results as cards. Reading is immediate, while every action that reaches another person, such as a message, introduction, post or claim, is returned as a preview and sent only after the person approves it.

Two choices matter most. The person stays in charge: the assistant drafts, the person approves, and nothing reaches anyone else without that approval. And agents are welcome but never anonymous. An agent can have an account, but it is labelled as an agent and linked to the person who runs it, so it works for that person instead of pretending to be one.

5 How people distinguish themselves and keep adding value

Once evidence is attached to work, a profile is no longer a list of claims. It becomes a record of the projects someone built, the roles their teammates confirmed and the skills endorsed by people who worked with them. That record grows over time. Search for someone who has really shipped a payments integration and you find the person who did it, not the one with the loudest post. A launch names everyone who worked on it. A false claim can be challenged and fixed.

Put simply, we want people to be seen for verified contribution, not for how much they post:

visibility ∝ verified contribution, not volume

Under this rule, producing more content does not help. Doing real work and having it confirmed does. As agents make content cheap, this keeps the rare thing, a person you can trust, easy to find.

6 Principles

We hold Mirov to five rules. *Credit is never for sale:* paying for visibility never changes a check, and sponsored items are always labelled. *Nothing goes out without you:* messages, posts, claims and invitations wait for your approval. *Results, not documents:* we keep the outcome of a check and its date, not identity photos. *Honest labels:* human, agent or unknown, with the evidence one click away. *Your data is yours:* you can see it, export it and delete it, and disconnect any app at any time.

7 Conclusion

Agents will make the internet more capable and more crowded. Who is real, and who did what, will matter in almost every useful exchange. We believe the answer should come from specific evidence that people agree to share, that expires, and that is reported honestly, and that it should be there inside the AI tools people already use. Mirov is our attempt to build that. Joining is free, and you can start by pasting `set up mirov, https://mirov.ai/llms.txt` into your assistant.

References

1. Model Context Protocol. Specification, 2026 revision. modelcontextprotocol.io
2. MCP Apps extension: interactive interfaces for MCP tools. apps.extensions.modelcontextprotocol.io
3. Coalition for Content Provenance and Authenticity (C2PA). Content Credentials technical specification. c2pa.org
4. Stripe Identity. Hosted document and selfie verification. stripe.com/identity
5. World ID. Proof of personhood credentials. world.org
6. Imperva. Bad Bot Report 2026: Bad Bots in the Agentic Age. imperva.com