Right, but Unproven

What five AI models said about one person, and what an evidence-bounded assessment could establish

Sergei Ponomarev · Andrey Ekhmenin · September 2026

On 23 September 2026 we gave five AI models the same task: say who Andrey Ekhmenin is, what he does, and whether there is any reason not to work with him. Each model had his full name, his role, his company and the address of his public LinkedIn profile. Each searched the web live.

One model found nothing it could tie to him and said so, three times. Another linked him to historical business records in a different country and a different script. After the results were disclosed, Andrey confirmed that those records were his. But nothing in the answer showed that. A reader with no way to ask him could not have told a correct link from a mix-up with a namesake. The three remaining models described him recognisably, in the main correctly, and incompletely.

That is the finding. An AI model can be factually right about a person and still give you no reliable reason to believe it. Below we describe how we checked this, what each side found, what we changed in the product as a result, and what follows for anyone who asks an AI assistant about a person.

1. Why we ran this

Before a meeting, a deal or an interview, people now ask an AI assistant about the person they are about to meet. The answer arrives in seconds, in confident prose, with a few links. What is not visible is how much of it the model found, how much it repeated from the question, and how much it inferred.

In August we examined a customer-facing AI agent through two independently applied approaches and published the result as *Silence Rather Than Error*. This time the object was not an agent but a person, and the person was one of us. Andrey volunteered his own public identity as the test object. That gave us something a third-party subject could not: after both results were frozen, the subject could say which of the models' claims were true. It also created a limitation we return to at the end.

2. How the experiment was built

The order of work was the same as in August and did not change once started.

1. The object was frozen: the public identity of Andrey Ekhmenin, with his public LinkedIn profile as the single entry point.
2. Each side worked alone. Sergei ran the retail AI Person Scan in the same form in which it is publicly sold, with no manual intervention. Andrey ran an ANDEKS™ assessment of what the public evidence could establish about the same person.
3. Each result was written to a file, and the SHA-256 checksum of the file was sent to the other side before either file was shared. A checksum is a fingerprint of a file: change one character and it changes. It proves after the fact that nothing was edited once the other side's result had been seen.

4. The files were exchanged and compared only after both checksums were on record. Clarifications that arose after disclosure, including what Andrey said about his own history, were recorded in a separate layer and did not alter either frozen result.

The comparison was written jointly and agreed by both sides as a Joint Comparison Record. It is the methodological appendix to this article and is linked at the end.

The first line of work: AI Person Scan (S. Ponomarev). The scan sends three questions to five AI models, each through its programming interface with live web search on: OpenAI gpt-4.1-mini, Anthropic claude-haiku-4-5, Google gemini-3.5-flash-lite, Perplexity sonar and xAI grok-4.3, the versions available on the day. The questions are the ones people actually ask: who is this person, what do they do, are there red flags about working with them. The fifteen answers are recorded word for word with the pages each model cited. A summary is then written by a model from those answers only, checked by fixed rules, and delivered as a PDF. The scan records what the models said. It does not verify it and does not claim any of it is true.

The second line of work: ANDEKS™ (A. Ekhmenin). This method does not ask what models say. It asks what can be established. ANDEKS™ independently assesses what can be evidentially established within a defined evidence boundary. It qualifies the strength and attribution of available public evidence, separates current from historical states, and marks where a conclusion would go beyond what the evidence supports. Its output is a bounded reconstruction: what is established, with what strength, and where the limit of a permissible conclusion lies.

The two lines were not assumed to be equivalent. One shows what a person asking an AI is told. The other shows what a careful assessor could stand behind.

3. What the two results agreed on

Both results placed Andrey in the same city, connected him to ANDEKS™ and to its public registration with a governance registry, and named the same independent assessments he had carried out and the same joint publication. Four of the five models named his current organisation and role.

We do not count that as a confirmation, and this is the first lesson of the experiment. The models were given his name, his role and his organisation in the question. A model that repeats them has found nothing. What the models added beyond the question was smaller: the city, the registry entry, two named assessments, a publication trail, a second business. In the report those rows now read "given in the question" rather than "said by four of five". Agreement with the input is not discovery.

The ANDEKS™ assessment made the same point from the other side. It established the current cluster, city plus ANDEKS™ plus assessment activity, with public evidential support, while separately qualifying the strength and independence of that support.

4. Where they diverged

Not found. One model, Anthropic's, could not tie any answer to the person given and said so for all three questions. Its answers were held back from the report under the scan's identity rules. For one model in five, at that moment, he was not identifiable.

Found, but unproven. The xAI model linked the profile to older business records in another country and another script. It gave a register number and named companies. In its first answer it allowed that this

"could be the same individual"; in its second it stated the link as fact. OpenAI's model, separately, placed him at an address in a third country from a record dated 2015.

The frozen ANDEKS™ assessment could not independently establish either link and therefore kept both outside the established core. The name alone, in another script, was not enough to tie records to the person without further identifiers, and the assessment said so. After disclosure Andrey stated that both sets of records were his: he had lived in that country, and the older business history was his own. According to that post-disclosure attestation, the models were right. But neither answer contained anything that would let a reader check the link, and the assessor working from the same public evidence could not establish it. The agreed record puts it in one line: factually true is not the same as evidentially established.

Then, not now. According to Andrey's post-disclosure attestation, the 2015 address and the current city are both true. They are true of different years. The scan printed them as a contradiction about where he is based. The ANDEKS™ assessment had already qualified business registers and old listings as sources that describe different moments in time and warned against merging them into one synchronous profile. A true historical attribute, presented as a current one, misleads.

Narrowed. The scan's summary described Andrey through his current work: ANDEKS™, AI governance, his business. The ANDEKS™ assessment listed further lines of activity that the summary did not carry. Each of those lines appears in only one of the fifteen answers, and the summary correctly leaves out what a single model says. The result is a description that is accurate and narrower than the person. Consensus among models around the most findable cluster does not show that the cluster is the whole picture.

Positive from nothing. Asked about red flags, one model found no adverse material and concluded that the reputation "appears to be positive" and that there was "no apparent reason to be cautious". The other models were more careful: little independent evidence, standard due diligence advised. The scan's own summary drew no reputation conclusion, because the identity was ambiguous. The agreed record states the rule: absence of discovered adverse evidence is not positive reputation established.

Search noise. One model placed two unrelated organisations with similar names next to his, one in liquidation and one from a leaked offshore database, and in the same breath said the connection was not established. The scan's rules were not, at that point, holding such an answer back. It appeared in the appendix.

5. What the product got wrong, and what changed

The scan's central safeguard worked. Because one model could not identify the person and another brought up what looked like a namesake, the report drew no conclusion about reputation and said why. The summary did not repeat the register number, the address or the unrelated organisations.

The appendix did. The report prints every answer word for word as the record, and that record contained a register number, a street address and company names that the scan had no way to tie to the person. In this case, on Andrey's attestation, they were his. For another buyer, about another person, they could have belonged to a stranger. The frozen report was left exactly as it was, as the evidence of what the product did on that day. The product was changed the same day.

The scan is now designed to hold back, from the appendix and from the summary, any answer that attaches a person's register entry to the profile without a place that matches it, any answer in which the model itself doubts it has the right person, any answer that quotes a street address, and any answer that puts a liquidated company or a leaked database next to the name without establishing a link. Register numbers and addresses are cut from the answers that remain. The line that explains a withheld answer no longer says the model confused the person with a namesake. It says the scan could not tie the answer to the profile, and that such an answer may be about a namesake or about the person's own past under another spelling or in another country.

That last point is the second finding of the experiment. On this pilot the new rule would have withheld the very answers Andrey later confirmed as his own. A scan that cannot tell a person's past from a stranger's record has to choose, and it chooses not to print either. Safety reduces completeness. We think that is the right trade for a product sold to anyone about anyone, and we think it should be said out loud rather than discovered.

6. What this means if you ask an AI about a person

Separate what you gave from what it found. If you typed the name, the role and the employer, the model repeating them tells you nothing. Look for what it added, and look at where it says it found it.

A confident link is not a proven one. The most valuable thing a model found in this experiment came with no way to check it. Treat any connection across countries, scripts or decades as a lead, not a fact, until something in the sources ties it to the same person.

Ask when, not only what. A model that has read a register entry from 2015 will present it in the present tense. Two answers that contradict each other about where someone is may both be true.

"No red flags found" is the floor, not the ceiling. It means the search turned up nothing adverse. It does not mean the search was wide, and it does not mean the reputation is good.

Ask more than one model. Not because five are wiser than one, but because their disagreement is the most honest signal you will get. In this experiment the range ran from "cannot find this person" to a reconstruction that he later confirmed as his earlier history, on the same input, on the same day.

7. What the experiment does not establish

- It is one person, on one date, with the models available that day. A rerun may produce different answers, and the prevalence of what we saw across other people is not established.
- The subject and the author of the ANDEKS™ assessment are the same person. Freezing both results before disclosure prevented either side from adapting to the other, but it does not make his confirmation of his own history an independent fact. In the agreed record such statements are marked as attested by the subject, not as established.
- The ANDEKS™ assessment did not include a full registry tying each of its conclusions to a source. This limits how far a third party can reproduce its individual steps, and is noted as a development point.
- Some of the pages the models cited are published on aibusiness.vc, the site of one of the authors. An external source is not automatically an independent one.
- The scan does not verify what models say, by design, and this experiment did not test whether it should.

8. The experiment and the authors

The experiment was carried out on 23 September 2026. Both results were frozen with checksums before exchange. The comparison was agreed by both sides and fixed as a PDF with its own checksum. The frozen results were not altered afterwards.

Sergei Ponomarev, PhD in political science, author of the test purchase methodology for AI agents and of AI Person Scan and AI Company Scan. Seven years of test purchases and service quality assessment; doctoral research on e-government. Portal aibusiness.vc. Profile: linkedin.com/in/sergei-ponomarev.

Andrey Ekhmenin, founder of ANDEKS™, an independent assessment framework for complex AI, digital and governance systems. Profile: linkedin.com/in/andrey-ekhmenin-eu.

Methodological appendix: AI Person Scan × ANDEKS™, Joint Comparison Record v1.0, agreed 23 September 2026. Published by ANDEKS™: docs.google.com/document/d/1dyHpLxbIGzkm0c4lH6LL5ZbODvOtG3md3ZV5PKutj98.

This article describes a private research experiment on one person, who took part with his consent. It is not a background check, a rating of any AI model, or a general assessment of either method. Personal records that the models found are described here only in general terms and are not reproduced.