

THE STATE OF HUMAN EVIDENCE 2027 · EARLY FINDINGS

Leaders trust real work. Almost none ask for it.

When AI can produce the work, what should count as proof of what a person can do? In October 2026, we put that question to 148 U.S. hiring managers and executives.

WHAT WE MEAN BY HUMAN EVIDENCE

Proof of what a person can do themselves, as opposed to what a tool did for them.¹

In practice, it shows up when someone explains why they framed a problem the way they did, catches an error in the AI's output, or defends a call under questioning.

57%

say a 60-minute work task for finalists is worth it for most or every role

**6%**

already use one for most roles



Leaders said the task was worth it even though it costs each finalist an hour and delays the offer by about three days. In the evidence ranking, a work sample completed for the hiring process came out first of the 12 kinds we tested.

88%

Were involved in a hire that fell short.

At least one hire in the past year turned out weaker than the evidence suggested. For 41%, it was one in four hires or more.

12%

Accept the output alone as proof.

That is when AI may have helped. The most common answer, at 43%, was work the person can explain, defend and correct.

78%

Give more weight to reasoning.

They weigh how a candidate reasons through a problem more than they did two years ago; only 4% weigh it less.

42%

Unscored interviews were least trusted.

Picked as least convincing 42% of the time it appeared, the highest of the 12.³ Résumés (34%) and degrees (33%) were close behind.

Three moves the findings point to

These follow from what leaders told us, not from measured hiring outcomes.

- 1 Stop grading the output. Grade the explanation.**
Only 12% accept the output alone as proof. The doubt leaders raised most often was a candidate who cannot explain how they got the answer (39%).
- 2 Ask what they would push back on.**
Asked what tells them most about a person's own contribution, leaders put challenging an assumption (26%) and explaining how the problem was framed (24%) ahead of everything else.
- 3 Test finalists on real work.**
Most leaders (57%) see a 60-minute finalist task as worth it for most or every role, and only 6% use one today.

FOUR MORE FINDINGS WE COVER IN A BRIEFING

The paradox

More than a third said a résumé or years of experience would have most improved their confidence in a recent decision. Neither ranks in the top five overall.

The AI divide

Where AI does around half the work or more, leaders were 32 points more likely to back a finalist task (73% vs. 40%).

The skill map

What leaders say proves judgment, domain knowledge and critical thinking, and what makes them doubt it.

The surprise

One kind of evidence finished far lower than we expected. I would want any leader to see it before redesigning how they hire.

A NOTE FROM THE RESEARCH ARCHITECT

I started this study because of a question I could not answer: when AI can do the work, how do you know what a person can do? The findings make the most sense when they are applied to real roles, and I would like to walk you through them.

Dr. Michael Campbell

Research Architect, Human Evidence Project

See the full findings before they are published.

A private 30-minute readout at no cost, applied to the roles you hire for. You will see where your hiring evidence is strong and where it is thin.

Book your readout →

A full hour, or a workshop for your team, is a paid session; ask on the call.

About the study. An independent study by the Human Evidence Project, sponsored by SignalVerified, whose founder, Dr. Michael Campbell, is its Research Architect. Method and limits are on page 3 and sources on page 4. Full report: January 2027, humanevidenceproject.com.

Method, robustness and limits

Because 54% of respondents work in technology, we tested whether the results hold outside tech.

FINDING	ALL (148)	TECH (80)	OTHER (68)	RESULT
Involved in a past-year hire that fell short of the evidence	88%	88%	88%	Holds
Say 1 in 4 or more of those hires fell short	41%	38%	44%	Holds
Give more weight to reasoning than two years ago	78%	81%	74%	Holds
When AI may have helped, top answer: work the person can explain, defend and correct	43%	48%	38%	Holds
Finalist work task worth it for most or every role	57%	58%	57%	Holds
Already use one for most roles	6%	8%	4%	Holds
Top two kinds of evidence: work sample, track record	1st, 2nd	1st, 2nd	1st, 2nd	Holds
Accept the output alone as proof when AI may have helped	12%	18%	6%	Differs
AI does around half or more of the team's work	72%	80%	62%	Differs
Rank of the résumé (of 12)	10th	12th	6th	Differs

"Holds" means no significant gap between tech and other industries, so reweighting by industry would not change the result. With groups of this size, only gaps of about 16 points or more are detectable. Checks by role and by company size (under vs. over 1,000 employees) found no significant gap on the first five rows. A finer split by company size puts the work-task result between 35% and 72%, with no consistent pattern.

Who took part

Respondents were 148 U.S. hiring managers and executives at organizations of 50 or more employees, each of whom made or influenced a hiring or advancement decision for a knowledge-work role in the past 12 months; 74% had been involved in six or more. Executives and founders made up 48% of the sample and team managers 52%, and 46% work at companies of 1,000 or more. Recruiter, talent acquisition and HR roles were screened out.

How the sample was built

We recruited through Cint, an opt-in online research panel, from October 6 to 9, 2026. Cint estimates that about 45,000 U.S. panelists in its marketplace match this profile.² Of the 673 panelists Cint reports entering from its side, 252 did not reach our questionnaire (pre-entry drop-off, which can include Cint's own screening). Of the 421 who did, 177 were screened out, 47 failed an attention check and 197 completed. Median completion time was 12.8 minutes. Results are unweighted, and because the sample is opt-in, no response rate can be calculated (Baker et al., 2013).

Two layers of quality checks

- **Before the survey (Cint):** an AI fraud score that ends suspect sessions, location and duplicate-entry checks, a deny list of past fraud, third-party VPN and IP fraud detection, completion-rate limits and speed checks (Cint, 2026a, 2026b).
- **Inside the survey (ours):** an attention check, followed by a review of every complete, which removed 45 of 197. We kept the first 148 of the remaining 152.

Cint's checks are taken from its published quality documentation, not from an audit of this study.

How precise the numbers are

A result near 50% could be about 8 points either way, and a result of 12% could plausibly fall between 8% and 18%. Scores from the best-worst ranking (Louviere et al., 2015) carry about ±9 points. These ranges reflect random variation only; they do not account for who chose to join the panel (Baker et al., 2013).

What this study does not show

It records what leaders say counts as proof, not what predicts performance on the job. It did not measure the cost of a weak hire, time to productivity, retention or promotion. The share of hires that fell short is each leader's own estimate.

Independence, sources and notes

Sponsor and independence

SignalVerified sponsors the study, and its Research Architect, Dr. Michael Campbell, founded SignalVerified. Respondents were told at the outset that the project is founded and funded by SignalVerified; no question names the company or its products. Respondents were also told that de-identified data will be published so others can check the work, and the questionnaire is available on request. Not every result favors the sponsor. In our scenario experiment, showing a candidate's reasoning, or their checking of the AI's work, did not measurably change whether leaders would advance them; between 90% and 96% would have done so in all four versions.

How to cite this brief

Human Evidence Project. (2026). *Leaders trust real work. Almost none ask for it. The State of Human Evidence 2027: Early findings*. <https://www.humanevidenceproject.com>

Sources

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., & Tourangeau, R. (2013). Summary report of the AAPOR Task Force on Non-probability Sampling. *Journal of Survey Statistics and Methodology*, 1(2), 90–143. <https://doi.org/10.1093/jssam/smt008>

Cint. (2026a). *The Cint ESOMAR questions*. Retrieved October 23, 2026, from <https://www.cint.com/esomar/>

Cint. (2026b). *Quality and standards: Data quality approach*. Retrieved October 23, 2026, from <https://www.cint.com/data-quality-approach/>

Human Evidence Project. (2026). *The State of Human Evidence 2027 questionnaire*. Available on request from customersupport@signalverified.net

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107337855>

Notes

1. The definition of human evidence follows the study's own open-ended question, asked before any options were shown: what tells you the most about what a person can do themselves, as opposed to what a tool did for them?
2. Panel counts (673 entries from Cint's side; about 45,000 matching U.S. panelists) were provided by Cint in correspondence with the Research Architect, October 2026.
3. Percentages are of all 148 respondents unless a subgroup is named. Figures from the evidence ranking (42%, 34% and 33%) are the share of times an item was picked as least convincing when it appeared.

About the Human Evidence Project

The Human Evidence Project is an independent research initiative founded and funded by SignalVerified. It studies which human abilities gain value as work is automated, what evidence employers accept as proof of capability, and how hiring and promotion decisions change when the origin of the work is uncertain. The State of Human Evidence 2027 is its first study; the full report, with methods and de-identified data, publishes in January 2027.

humanevidenceproject.com · customersupport@signalverified.net