

Position: Accountability Mechanisms Are Adopted Only When the Verifier Pays Nothing*

Evidence from a census of freelance job postings

Akshara Rajee
Product Manager

Abstract

Proposals for accountability in agent-mediated work—AI use statements, detection, watermarking, content provenance—presuppose that the parties who would rely on them want them. Whether that demand exists has not been measured in the markets where AI-mediated work is exchanged between strangers.

We report a complete enumeration of the active job-posting queues on two freelance platforms, coded against a preregistered scheme using an automated coder validated against hand codes and against injected positive controls. Requests for disclosure of AI use are close to absent, and no posting in either market names watermarking, content credentials, or provenance of any kind. Clients are nevertheless concerned with verification, and those most hostile to AI request it most often. The mechanisms they adopt share a property that the provenance stack does not: each is free to the client, and its cost falls on the applicant. Where the platform sells assurance products directly, uptake is negligible, and several such products are purchased by no client in the corpus. Demand for verification is also close to flat in the value of the contract.

Our position is that accountability mechanisms in such markets are selected on cost incidence rather than on design, and that any mechanism imposing a cost on the verifier will not be adopted there. We argue that who pays to verify should be treated as a first-class design constraint, and that demand should be measured rather than assumed.

Keywords

AI disclosure, provenance, watermarking, accountability, freelance work, cost incidence, challenge–response, position paper

1 What clients build instead

Accountability proposals for agent-mediated work generally take the form of a record: a statement of AI use, a detector output, a watermark, or a content credential. Each assumes a reader who wants the record and will act on it. That assumption is rarely examined, and it is most consequential in markets where work changes hands between strangers who will transact once, since these are the settings in which shared social context is least able to substitute for documentation.

A software company posting on Upwork opens its listing in capital letters: “DO NOT SEND AN AI GENERATED PROPOSAL. READ THROUGH THIS POST AND GIVE US YOUR HUMAN RESPONSE.” Further down, the role requires a developer who “treats Claude

Code as a peer on the keyboard,” who can “review agent diffs critically,” and who will “pair with Claude Code daily.” AI use is not merely permitted; it is the work. To apply, the client asks for a link to a recent pull request and a five-to-ten minute screen recording of the applicant working with Claude Code on a real task. A second client, hiring a virtual assistant, asks for something cheaper: “Include paprika in your response to ensure you’ve read the full post.”

Neither client asks for an AI use statement, a detector, or a provenance log. Each instead creates an occasion to observe a person, producing no record of what the agent did—a mechanism for witnessed contribution [6], improvised by someone who has almost certainly never read Suchman [10].

One reading is that these clients reached for the available tool rather than the appropriate one. We think the evidence supports a different explanation. Both mechanisms share a property that watermarking and content credentials do not: they cost the client nothing, and their cost is borne by the applicant, in unpaid effort and attention. **On this account the provenance stack is unused not because it is the wrong instrument but because using it would require the verifier to spend something, and in this market the verifier does not.**

2 A census of what clients request

Corpus. The first source is 92 publicly visible Upwork postings spanning writing, design, software development, and virtual assistance, hand-coded in full by one author. The second is a complete enumeration of the Freelancer.com *active* queue—every public posting open for bids at one instant, 2,942 in all, collected from the platform’s public API across matched categories in July–August 2026. Every posting in it is coded, so its counts are exact and involve no extrapolation. It is an enumeration of a stock, not a sample of the flow of postings: the open pool over-represents postings that linger in it, and we return in §6 to the selection this induces and to a direct test of whether it moves any reported quantity.

Coding. Keyword matching is not viable here: “AI” denotes Adobe Illustrator file formats, the subject matter of the deliverable, and the product the client is building at least as often as it denotes a tool the worker might use. We wrote an explicit codebook, froze it before coding the census, and preregistered four hypotheses together with the rule that a failure to replicate would be reported as such. An automated coder applies the codebook and must quote a verbatim evidence span for each positive code, checked programmatically against the source text, so an unsupported justification is detected rather than accepted.

*Position paper accepted to the workshop “Nobody Did This”: Contribution, Originality, and Accountability in Agent-Mediated Collaboration [6] at the 2026 ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW ’26), Salt Lake City, UT, USA, October 2026.

Validating an instrument for a null result. A null claim depends on *sensitivity*, which is precisely what a gold set containing no positives cannot establish: a coder that never returns a positive would score perfectly. We therefore built positive controls, inserting disclosure and detection language drawn from observed marketplace phrasing into genuine postings and running the coder blind on the mixture. Sensitivity was 12/12 for disclosure and 11/12 for detection, with no false positives on unmodified negative controls, and a second model from a different family agrees exactly ($\alpha = 1.00$) on the three constructs introduced here (Appendix A.1).

Table 1: Cross-platform comparison. Both columns model-coded with the same codebook, so the contrast is between markets rather than coders. p from Fisher’s exact test, Holm-adjusted across the eight constructs.

Construct	Upwork	Freelancer	Holm p
Prohibits AI in the work	3.3%	1.6%	0.922
Signals human-only, no ban	4.3%	2.7%	1
Mentions detection or verification	0.0%	0.4%	1
Requests disclosure of AI use	0.0%	0.1%	1
Expects or requires AI fluency	12.0%	3.2%	0.00181
Bans AI in the application only	2.2%	0.2%	0.171
Challenge–response human test	8.7%	1.5%	0.00106
Requests evidence of competence	48.9%	46.7%	1

2.1 No posting requests provenance

No posting in either market asked a freelancer to declare AI use in a form that would produce a usable record, and none named watermarking, content credentials, SynthID, or provenance (Table 1). The two null results are statistically indistinguishable between platforms. Among the 14 postings that name any verification instrument, 71% name plagiarism infrastructure developed for a different problem a decade earlier, and 0 name the provenance stack.

This shapes how the null should be interpreted. Accountability is not unwanted; rather, the demand that exists was already met by a cheaper and better integrated substitute. Education demanded AI detection and was served by Turnitin, and freelance clients reach for the same tooling—poorly matched to the task, since it identifies copying rather than generation, but already available and free at the point of use.

2.2 Clients concerned about AI verify more, not less

Clients hostile to AI request verification substantially more often than others (62% against 46%, Fisher $p = 0.00141$), so inattention is an unlikely explanation. One outreach posting specifies CAN-SPAM compliance, Do Not Call scrubbing, and autodialer consent, regulating the worker’s tooling in considerable detail while saying nothing about AI. These clients considered verification and selected a different instrument.

The instrument they converge on is a challenge–response test: the applicant is asked to reproduce an arbitrary token, and the stated purpose is generally explicit, so *I know you read this post*. Such tests appear in 8 of 92 Upwork postings against no mentions

of detection ($p = 0.00781$), and the ordering replicates on the census across all four categories ($p = 0.00126$). The token conveys nothing about competence. It is an inexpensive test that a particular person processed a particular document, and it costs the client one sentence.

The two markets differ in composition—budget type, category mix, posting length—so the cross-platform gaps could in principle be artifacts of what each market is made of. They are not. Holding those factors fixed in a penalized logistic regression (Appendix A.2), the challenge–response test remains far more common on Upwork (adjusted odds ratio 4.44, Holm $p = 0.0112$), as does expectation of AI fluency (3.62, $p = 0.00449$), while demand for evidence of competence does not differ between platforms (1.23, $p = 0.254$), as one would expect of a general screening practice.

2.3 Uptake of paid assurance products is negligible

Freelancer.com already sells assurance. At the moment of posting, a client may purchase a non-disclosure agreement, sealed bids, qualified-bidder screening, an IP contract, or a non-compete. Across the census, 1.2% of clients purchase any upgrade: 0.7% a non-disclosure agreement, 0.5% sealed bids, and 5 of the products, including qualified-bidder screening and IP contracts, are purchased by no client at all. Clients hostile to AI are no more likely to purchase than others (0.8% against 1.2%).

Demand for verification is also close to flat in the value of the contract (Figure 1A). Across a range of budgets spanning more than two orders of magnitude, requests for evidence of competence move only from 43% to 50%, and hostility to AI does not trend at all. Whatever governs verification here, it does not appear to be a risk-priced investment.

Counting postings understates the asymmetry, because a requirement written into a posting is discharged by every applicant who answers it while a purchased product is paid for once. Expressed as incidence, requirements satisfied by applicants outnumber assurance products bought by clients by more than three orders of magnitude (Figure 1B).

2.4 Where the platform already watches, the AI ban disappears

The account also predicts what happens when verification the client did not ask for is supplied for free. Freelancer.com works its 787 hourly contracts through a monitored desktop application that tracks time and takes periodic work-diary screenshots, while its 2,155 fixed-price contracts give the client no such view of how the work was produced. If clients police AI because the deliverable is otherwise a black box, the ones already handed a window onto the process should police it less. They do, but only where the window is relevant. Holding job category fixed, an hourly posting is markedly less likely to prohibit AI in the work than a fixed-price one (0.9% against 1.8%; adjusted odds ratio 0.37, $p = 0.0161$; Appendix A.3). The monitoring substitutes for the ban specifically: it does not lower demand for evidence of competence, which is slightly *higher* under hourly (49.9% against 45.5%), because watching the work in progress verifies effort, not the applicant’s fitness to be hired in the first place. The two mechanisms answer different

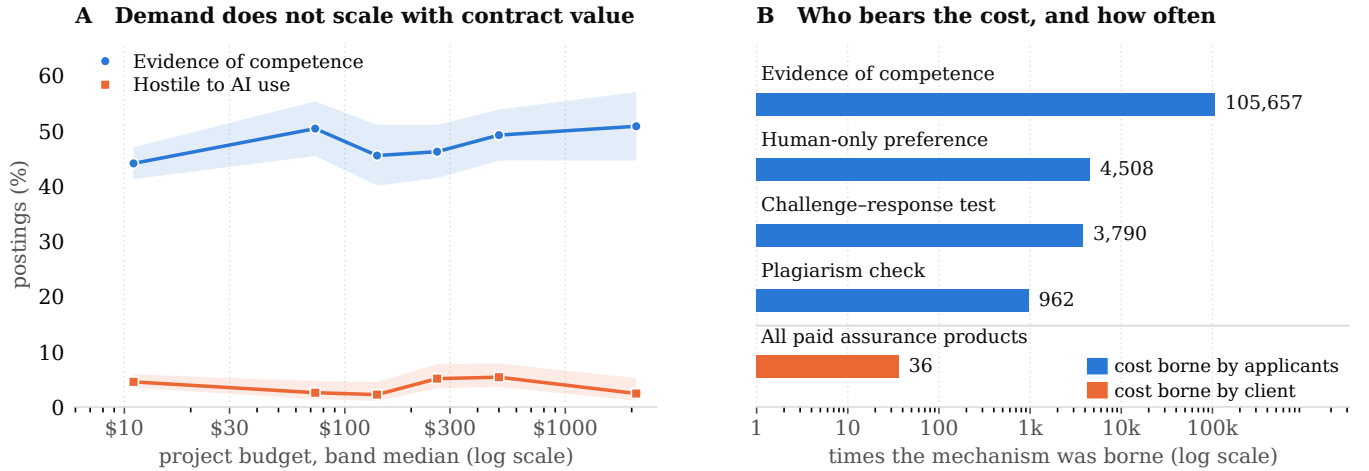


Figure 1: (A) Share of postings requesting evidence of competence, and share hostile to AI use, against the value of the contract; bands are 95% intervals. Across budgets spanning more than two orders of magnitude, neither quantity moves appreciably. (B) How often each mechanism is actually borne, on a logarithmic scale. The units differ deliberately between the two groups, and the difference is the point: a requirement written into a posting is satisfied once by every applicant, of which the median posting attracts 46, whereas a purchased product is paid for once by the client. The applicant counts are observed mid-life in an open-queue snapshot (median posting age 64 hours), so they undercount the bids each posting will finally receive and the asymmetry is larger than shown.

questions, and only the one the monitoring answers is displaced by it.

3 Cost incidence as the selection mechanism

Taken together these observations suggest a single mechanism. Every verification practice we observe is free to the principal, and its cost falls on the agent. A challenge-response test costs the client one sentence and every applicant a careful reading. A portfolio request or a narrative interrogation costs nothing to issue and hours to satisfy, for many applicants who will not be hired. An unpaid demonstration costs nothing to require and yields a real work sample. The paid trial is the only mechanism the client funds, and it is the least common.

Watermarking and content provenance fall on the other side of this line. A watermark must be checked. A content credential must be verified and must survive intermediaries that strip it. A disclosure statement must be read and interpreted by someone with an established norm for reading it [4]. Each imposes a cost on the verifier, in a market whose clients decline to spend a few dollars on a non-disclosure agreement.

The same account explains the concealment equilibrium without attributing bad faith to either party. Disclosure is neither required nor rewarded, and thirteen experiments find that actors who disclose AI use are trusted less than those who do not [8]. For the client, reading a disclosure costs attention and returns a signal that is inexpensive to produce [2, 9], and neither party has an incentive to establish the regime.

The worker’s side of that equilibrium is visible in the forums where freelancers talk to each other. We coded 1,472 posts from ten freelancing subreddits against a preregistered worker-side scheme, and the asymmetry that governs the client side reappears: concern

about *detection* is four times as common as any mention of a disclosure norm (16.1% of posts against 4.1%; paired $p = 8e - 27$), the mirror of clients naming detection but never asking for a declaration, and where a post reports an actual penalty it attaches to being caught rather than to having disclosed (Appendix A.5). What the data do *not* show is workers uniformly hiding: among the 384 posts that discuss using AI in the work, concealment and disclosure are discussed about equally (13% against 14%, $p = 0.85$). The equilibrium is better described as detection anxiety in the absence of a disclosure norm than as a universal strategy of silence. These are self-selected posts, so we read them as illustrative of the mechanism rather than as a second census.

4 Position and call to action

We hold that accountability mechanisms in markets of this kind are selected on cost incidence rather than on design, and that who pays to verify should be a first-class design constraint rather than a deployment detail. Three commitments follow.

Measure demand, and report who bears the cost. Work on watermarking and detection reports robustness against paraphrase attack. It does not report who requested the capability, what those parties were already doing, or what verification would cost them. We have shown that these questions are answerable from public data. They are also timely: Article 50 of the EU AI Act became applicable on 2 August 2026 [3], requiring machine-readable marking of synthetic content, while reported C2PA adoption shows signing considerably ahead of verification [1]. A marking obligation places the burden on providers and the reading burden on a market that declines to purchase sealed bids.

Design for zero marginal cost to the verifier. Turnitin was adopted because institutions absorbed the cost centrally and the individual verifier paid nothing at the point of use. Accountability technology should be evaluated on whether it can be made automatic and free—a platform default rather than a tool the principal must invoke—and on whether its cost can be shifted to the party seeking the engagement.

Study process witnessing alongside artifacts. Screen recordings, narrative interrogation, and challenge–response tokens are accountability infrastructure built by people who needed it, under time pressure, without a vocabulary for what they were building. They supervise a process rather than certify an output, and are closer in spirit to verifying reasoning steps [7] than to labelling artifacts. Each has a characterisable failure mode: demonstrations can be rehearsed, portfolios can be borrowed, and challenge tokens fail once public.¹

5 Alternative views

Absence of demand says nothing about efficacy. Clients may decline to request provenance because it would not help them, which is the workshop’s thesis reached by another route. We think this is inconsistent with the pattern: these clients actively construct verification, and those most concerned about AI construct the most. They are addressing the problem with a different instrument rather than neglecting it.

Uptake of non-disclosure agreements says nothing about AI tooling. The upgrades we observe are legal and visibility products rather than accountability products, so our estimate is a revealed-preference *upper bound* on willingness to pay for assurance rather than a direct measurement of demand for AI verification. We regard the bound as informative: a market that declines to purchase a non-disclosure agreement is unlikely to license a provenance checker.

Freelance work is unrepresentative. It is a market of strangers contracting once, which is the condition under which documentation is expected to substitute for social context, and therefore the most informative case rather than a peripheral one. Ulloa et al. [11] describe product managers weighing delegation at individual, team, and organisational levels; freelancers have the first of these and none of the others.

Cost may not be the only barrier. Clients may be unaware that these tools exist. Observational data cannot separate unawareness from unwillingness, and a field experiment offering provenance checking at no charge would distinguish them. Our account predicts low uptake even when the tool is free, provided it still requires an action.

A boundary case the mechanism predicts. The cleanest test of a cost-incidence account is a market where disclosure has a *payoff* rather than a cost. Translation is that market: machine-translation post-editing is a named contract type, priced below full human

translation per word, so whether a machine was involved changes what the client pays. Our mechanism predicts that disclosure should be routine there while provenance stays absent, because the answer is settled in the rate rather than by inspecting an artifact. We scoped this vertical but do not report it as a result: on the public boards we could collect, rate-differentiated post-editing did not appear in meaningful numbers. A corpus of priced post-editing briefs is the test that would confirm or refute the mechanism most directly.

6 Limitations

The study covers two platforms at one point in time. The Upwork sample is small and single-coded; seven of its 99 postings lack archived text. The Freelancer.com corpus is coded automatically with measured error, and the coder has low sensitivity on the more interpretive constructs, human preference in particular, so those quantities are reported with correction rather than as hand codes. The challenge–response construct was added after inspecting a screening miss, so its discovery was exploratory while its cross-platform prevalence was preregistered and is confirmatory; the classification of verification instruments is post hoc.

Snapshot selection and exposure. The census enumerates the *active* queue at one instant, which introduces two biases we test directly rather than assume away (Appendix A.4). First, postings that attract fewer bids linger in the open pool and are over-sampled, and AI-hostile postings are indeed older at collection than others, so the snapshot inflates the prevalence of AI hostility relative to the flow of postings. That works against a positive finding for the null constructs and cannot manufacture one—zero remains zero in every subsample, and the pool is if anything enriched for the AI-averse clients who never asked for provenance. Second, each posting’s bid count is observed at a different, incomplete point in its life, and controlling exposure dissolves the one bid-count contrast we had reported: AI-hostile postings receive fewer bids in raw counts (median 37 against 47, $p = 0.031$), but a quasi-Poisson model with $\log(\text{age})$ as an offset returns a rate ratio of 1.02 ($p = 0.89$). We therefore withdraw that comparison; the paid-assurance and challenge–response results, which do carry the argument, are unaffected by exposure. Re-polling the same projects 4 days later bounds the censoring directly: total bids rose by 5% (median 46 to 49), so the incidence estimate in Figure 1 stays conservative.

We could not split the corpus by client history: the public API anonymizes the owner identifier on every record, so first-time and returning clients cannot be told apart, and a coarse proxy based on the platform’s repost marker shows no difference in any construct (Appendix A.6). The corpus, codebook, coder, positive controls, and analysis code are released.

AI use statement

We used a large language model as a writing aid: to interrogate the papers this work responds to, to build the collection, coding, and analysis pipeline, and to draft and revise this text. The coding scheme, the reading of all Upwork postings, the interpretation, and the positions taken here are ours, and every reported quantity was computed from the released dataset. We note this in the spirit of the workshop’s own statement, and with awareness that it illustrates

¹The pattern is not confined to freelance markets. ICML’s 2026 enforcement action, which desk-rejected 497 papers, used neither output watermarking nor an AI-text detector, but hidden canary phrases placed in submission PDFs [5]—a challenge–response test, and free to deploy.

Table 2: Coder validation against the hand-coded Upwork gold set. Sensitivity for the two null constructs is unmeasurable there — no positives exist — and is established by positive control instead.

Construct	α	Sens.	Spec.	F_1
Prohibits AI in the work	0.74	0.60	1.00	0.75
Signals human-only, no ban	0.52	0.43	0.99	0.55
Mentions detection or verification	1.00	–	1.00	–
Expects or requires AI fluency	0.67	0.78	0.95	0.70
Requests evidence of competence	0.58	0.88	0.74	0.76
<i>Positive control (injected positives)</i>				
Requests disclosure of AI use	–	12/12	–	–
Mentions detection	–	11/12	–	–

the proposal’s point: the disclosure records that a tool was used and cannot indicate which framings were originally ours.

References

- [1] Coalition for Content Provenance and Authenticity. 2026. C2PA Technical Specification: Content Credentials. <https://c2pa.org>.
- [2] Jingyi Cui, Gabriel Dias, and Justin Ye. 2025. Signaling in the Age of AI: Evidence from Cover Letters. *arXiv preprint arXiv:2509.25054* (2025).
- [3] European Union. 2024. Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 50: Transparency Obligations. Article 50 applicable from 2 August 2026; Commission guidelines adopted 20 July 2026.
- [4] Angel Hsing-Chi Hwang, Senya Wong, Baixiao Chen, Jessica He, and Hyo Jin Do. 2026. “Better Ask for Forgiveness than Permission”: Practices and Policies of AI Disclosure in Freelance Work. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI ’26)*. doi:10.1145/3772318.3791920 arXiv:2603.07459.
- [5] ICML 2026 Program Chairs. 2026. On Violations of LLM Review Policies. ICML Blog. <https://blog.icml.cc/2026/03/18/on-violations-of-llm-review-policies/>.
- [6] Kashif Imteyaz, Mohammad Rashidujjaman Rifat, Divya Ramesh, Steven R. Rick, Simo Hosio, Hauke Sandhaus, Advait Sarkar, Christoph Riedl, and Saiph Savage. 2026. “Nobody Did This”: Contribution, Originality, and Accountability in Agent-Mediated Collaboration. arXiv:2607.26387 [cs.CY] doi:10.48550/arXiv.2607.26387 Workshop at the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW ’26).
- [7] Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. 2023. Let’s Verify Step by Step. *arXiv preprint arXiv:2305.20050* (2023).
- [8] Oliver Schilke and Martin Reimann. 2025. The Transparency Dilemma: How AI Disclosure Erodes Trust. *Organizational Behavior and Human Decision Processes* 188 (2025). doi:10.1016/j.obhdp.2025.104405
- [9] Michael Spence. 1978. Job Market Signaling. In *Uncertainty in Economics*. Academic Press, 281–306.
- [10] Lucy A. Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press.
- [11] Mara Ulloa, Jenna L. Butler, Sankeerti Haniyur, Courtney Miller, Barrett Amos, Advait Sarkar, and Margaret-Anne Storey. 2026. Product Manager Practices for Delegating Work to Generative AI: “Accountability must not be delegated to non-human actors”. In *Proceedings of the IEEE/ACM 48th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP ’26)*. doi:10.1145/3786583.3786849 arXiv:2510.02504.

A Supplementary material

All quantities below are produced by the released analysis scripts from the released corpus. They support claims made in the body and are not new results.

A.1 Coder validation

Table 2 reports agreement with the hand-coded Upwork gold set and the sensitivity established by positive control. The second coder, from a different model family, also produced unverifiable evidence spans for 15.1% of its positive codes, against 0.0% for the primary coder, which is why span verification is enforced rather than assumed.

Table 5: Client demand by pricing model on Freelancer.com. Hourly contracts are worked through the platform’s monitored Desktop App (time tracking, work-diary screenshots); fixed-price contracts give the client no such channel. Rates are model-coded prevalence; *adj. OR* is the odds ratio for the hourly indicator from a logistic model adjusting for job category, so the contrast is not an artifact of Development skewing hourly. *p* (Fisher) is Holm-adjusted across the eight constructs.

	Fixed (<i>n</i> = 2, 155)	Hourly (<i>n</i> = 787)	adj. OR (hourly)	Holm <i>p</i>
Prohibits AI in the work	1.8%	0.9%	0.37	0.552
Signals human-only, no ban	2.6%	3.0%	1.00	1
Mentions detection or verification	0.5%	0.3%	0.45	1
Requests disclosure of AI use	0.1%	0.0%	–	1
Expects or requires AI fluency	2.8%	4.3%	1.73	0.305
Bans AI in the application only	0.3%	0.1%	–	1
Challenge–response human test	1.4%	1.8%	1.38	1
Requests evidence of competence	45.5%	49.9%	1.27	0.295
<i>Any verification-type ask</i>	47.6%	51.7%	1.24	0.0502

A.2 Prevalence, and the cross-platform comparison adjusted for composition

Table 3 gives the census rates with intervals. Table 4 re-estimates each construct with a Firth penalized logistic regression holding job category, budget type, posting language, and posting length fixed; it is the adjusted comparison cited in §2.2. Budget enters as type rather than amount because the Upwork corpus records no amount.

Table 3: What clients ask for on Freelancer.com (2,942 postings, July–August 2026). Model-coded; Wilson intervals, Clopper–Pearson at zero.

What the posting asks for	Count	Rate	95% CI
Prohibits AI in the work	46	1.6%	1.17–2.08
Signals human-only, no ban	80	2.7%	2.19–3.37
Mentions detection or verification	12	0.4%	0.23–0.71
Requests disclosure of AI use	2	0.1%	0.02–0.25
Expects or requires AI fluency	94	3.2%	2.62–3.89
Bans AI in the application only	7	0.2%	0.12–0.49
Challenge–response human test	44	1.5%	1.12–2.00
Requests evidence of competence	1374	46.7%	44.91–48.51

Table 4: Cross-platform comparison, adjusted for composition. Both columns are model-coded with the same codebook. Each row is a Firth penalized logistic regression on the pooled English-only postings (Upwork *n* = 92, Freelancer *n* = 2, 647) with the platform indicator, job category, budget type (hourly/fixed), and $\log(1 + \text{words})$; the odds ratio is for a posting being on Upwork rather than Freelancer, so > 1 means the construct is more common on Upwork once composition is held fixed. Budget is entered as type because the Upwork corpus records no amount. *p* is a penalized likelihood-ratio test, Holm-adjusted across the eight constructs.

Construct	Unadj. OR	Adj. OR	95% CI	Holm <i>p</i>
Prohibits AI in the work	2.46	3.02	0.95–9.59	0.345
Signals human-only, no ban	1.81	1.25	0.41–3.78	0.844
Mentions detection or verification	1.24	0.61	0.03–11.43	1
Requests disclosure of AI use	5.72	3.93	0.26–60.41	1
Expects or requires AI fluency	4.08	3.62	1.81–7.25	0.00449
Bans AI in the application only	9.73	15.33	3.53–66.48	0.0675
Challenge–response human test	5.89	4.44	1.86–10.61	0.0112
Requests evidence of competence	1.15	1.23	0.79–1.92	0.254

A.3 Demand by pricing model

Table 5 is the hourly-versus-fixed split behind §2.4: hourly work is performed through the platform’s monitored desktop application, fixed-price work is not, and the AI ban is the construct that the monitoring displaces.

A.4 Snapshot exposure

Figure 2 shows the two snapshot biases directly—the age distribution of the open pool, and bids against age for AI-hostile and other postings—and Table 6 gives the age-banded medians and the exposure-adjusted incidence-rate ratios summarised in §6.

A Mixed observation ages

B No effect once adjusted

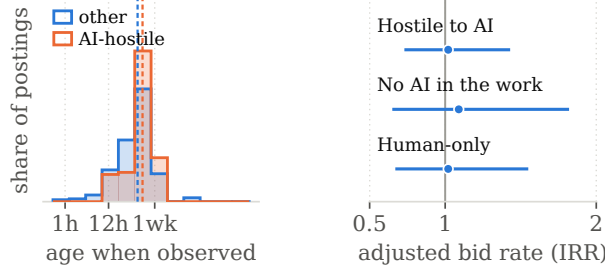


Figure 2: Snapshot selection in the active-pool census. (A) Postings are observed at widely differing ages, so every bid count is censored at a different point in the posting’s life; dashed lines are group medians, and AI-hostile postings sit later (84 against 63 hours, $p = 8.1e - 05$). (B) Exposure-adjusted incidence-rate ratios for bids, with 95% intervals. All three cover one, so the raw median bid gap between AI-hostile and other postings is a difference in when they were observed rather than in demand.

Table 6: Exposure control for the bid comparison. The active-pool snapshot observes each posting’s bid count at a different age, so the raw gap between AI-hostile and other postings confounds demand with exposure. Median bids are shown within age bands (a); a quasi-Poisson GLM with $\log(\text{age})$ as an offset and covariates for category, budget band, project type, and description length gives the exposure-adjusted incidence-rate ratio (b).

(a) Median bids within age bands, AI-hostile vs other			
Age band	AI-hostile	Other	Mann-Whitney p
<24h	48	40	0.82
24–72h	34	54	0.11
72–168h	36	46	0.086
>168h	–	32	–
(b) Exposure-adjusted incidence-rate ratio (neg. binomial)			
Construct	IRR	95% CI	p
Hostile to AI use	1.02	[0.73, 1.43]	0.89
Prohibits AI in the work	1.09	[0.65, 1.82]	0.74
Signals human-only	1.02	[0.67, 1.55]	0.93

A.5 Worker-side corpus

Table 7 gives the worker-side rates discussed in §3. It is a self-selected convenience corpus of forum posts, not a census, and over-represents grievance. Where a post reports an actual penalty (105 of them), the penalty attaches to being caught: 65 of those posts turn on a detector or a suspicion of AI, and 3 on a disclosed use.

Table 7: Worker-side codes over 1,472 freelancer-forum posts (Reddit, model-coded against a preregistered worker-side codebook). This is a self-selected convenience corpus, not a census; rates describe what the forums discuss. Wilson intervals.

What the post does	Rate	95% CI
Uses / considers AI for client work	26.1%	23.9–28.4
Hides AI use, or advises hiding	3.9%	3.0–5.0
Discloses to clients, or urges it	4.1%	3.2–5.2
Concern about AI detection	16.1%	14.3–18.1
Reports client/platform anti-AI stance	9.4%	8.1–11.0
Reports a penalty tied to AI	7.1%	5.9–8.6

A.6 Client history proxy

Table 8 splits the census by the platform’s repost marker, the only available proxy for a client having listed before, and finds no difference in any construct.

Table 8: Demand on first-listing versus reposted projects on Freelancer.com. A reposted project is one whose title carries the platform’s repost suffix (“– 2”, “– 3”), so its client has listed before. This is a weak, one-directional proxy: it cannot identify the same client across projects or count repeat hiring, and the intended split by `owner_id` is impossible because the API withholds it. p from Fisher’s exact test, Holm-adjusted.

	First listing ($n = 2,817$)	Reposted ($n = 125$)	Holm p
Prohibits AI in the work	1.6%	1.6%	1
Signals human-only, no ban	2.7%	3.2%	1
Mentions detection or verification	0.4%	0.0%	1
Requests disclosure of AI use	0.1%	0.0%	1
Expects or requires AI fluency	3.1%	4.8%	1
Bans AI in the application only	0.2%	0.0%	1
Challenge-response human test	1.5%	0.8%	1
Requests evidence of competence	46.6%	48.0%	1
Any verification-type ask	48.7%	48.8%	1