IMAGINATION APPLIED OPEN RESEARCH SERIES

Adoption Without Confidence?

Favorable Stance, Accuracy Trust, and Reported AI-Use Frequency in the 2025 Stack Overflow Survey

Josh Penzell | Imagination Applied

July 2026

A denominator-aware secondary analysis of the 2023-2025 Stack Overflow Developer Surveys

Abstract

Stack Overflow has already reported that AI use rose while trust weakened. This secondary analysis asks a narrower question: when a broad favorable stance toward using AI and trust in AI-output accuracy are measured separately, which carries more information about reported use frequency? The three-year context uses 203,812 CSV records from repeated cross-sections. The 2025 complete-case sample contains 33,231 respondents, of whom 25,698 (77.3%) identify as professional developers. To reduce the direct contrast between users and nonusers, the primary analysis is restricted to 26,102 current AI users and classifies daily versus weekly or less-frequent use. Daily use ranges from 18.8% among very unfavorable current users to 88.1% among very favorable current users, although the middle categories are not strictly ordered. A categorical stance-only model reaches five-fold cross-validated ROC AUC 0.758, compared with 0.669 for accuracy trust alone. Context plus trust reaches AUC 0.704; context plus stance reaches 0.790; adding trust raises that result to 0.793. Across 50 matched repeated splits, context plus stance exceeds context plus trust by 0.087 AUC on average, while trust adds 0.003 after stance. The full-sample gradient is larger (3.4% to 85.7%; stance AUC 0.822), consistent with construct proximity between favoring AI use and reporting AI use. These same-wave associations do not establish causal direction or future predictive validity.

CENTRAL FINDING Among people already using AI, favorable stance carries substantially more information about daily versus less-frequent use than accuracy trust does. That shows the two survey items are not interchangeable here; it does not show that changing stance will change behavior.

Key findings

- Among 26,102 current users, daily use is 18.8% for the very unfavorable and 88.1% for the very favorable; unsure (34.0%) is slightly above indifferent (31.5%), so the pattern is steep but not perfectly monotonic.
- Stance alone classifies daily versus less-frequent use with AUC 0.758, compared with 0.669 for accuracy trust alone.
- Context plus stance outperforms context plus trust in all 50 matched repeated splits; adding trust after stance produces a small but consistently positive improvement.
- Among current users who highly distrust AI accuracy, daily use ranges from 17.7% for the very unfavorable to 85.1% for the very favorable.
- The result remains similar among professional current users and in five-fold country-grouped validation, but these are internal robustness checks rather than external validation.
- The strongest counterinterpretation remains viable: the stance item is broader and closer to the use outcome than the narrower accuracy-trust item, and use may shape stance as much as stance shapes use.

1. What is new - and what is not

The broad adoption-with-low-trust story is not new. Stack Overflow has published the 2025 use, favorable-stance, and accuracy-trust distributions, related favorable stance to workflow integration, and later discussed the developer AI trust gap directly. Broader workplace research has also distinguished AI use, attitudes, and trust.

The contribution here is narrower: a reproducible large-sample comparison of how the Stack Overflow favorable-stance and accuracy-trust items classify reported use frequency, with the main test restricted to current users and supported by trust-stratified, professional-status, alternative-outcome, country-grouped, calibration, and repeated-split checks. The analysis does not claim that stance and trust are newly discovered constructs.

Research questions

- RQ1. What remains of the 2023-2025 use-trust pattern after harmonizing the trust denominator?
- RQ2. Among current AI users, how does reported daily use vary across favorable-stance responses?
- RQ3. How much classification information does favorable stance add beyond accuracy trust and respondent context?
- RQ4. Does the result remain visible among professional developers, within trust levels, and when countries form held-out groups?
- RQ5. Which managerial implications survive the causal and measurement limitations?

2. Construct boundaries and closest prior work

Favorable stance is not a validated sentiment scale

The Stack Overflow item AISent asks how favorable the respondent's stance is toward using AI tools in the development workflow. Stack Overflow labels the item AI tool sentiment. This note uses 'favorable stance' for analytic claims because the item is a single broad evaluation, not a validated multi-item affect scale. It may combine usefulness, fit, identity, opportunity, prior experience, and willingness to use.

Accuracy trust is narrower

The AIAcc item asks how much the respondent trusts the accuracy of AI output. Trust research distinguishes attitude, reliance, and calibrated reliance. A person can find AI useful enough to use daily while distrusting individual answers enough to test them. Confidence in accuracy also does not guarantee that a tool is useful, available, permitted, or relevant.

Closest empirical comparison

Choudhuri and colleagues used validated constructs and a theoretically grounded model with 238 developers at GitHub and Microsoft. They found that system/output quality, functional value, and goal maintenance shaped trust, and that trust plus cognitive styles related to intentions to use GenAI. That work is theoretically and psychometrically stronger. The present analysis is larger, public, and focused on direct comparative classification of a reported frequency category. The approaches are complementary.

3. Data and methods

The public 2023, 2024, and 2025 Stack Overflow Developer Survey CSVs contain 203,812 rows: 89,184 in 2023, 65,437 in 2024, and 49,191 in the downloaded 2025 file. Stack Overflow's published 2025 totals differ slightly from the public CSV, consistent with versioning or post-release filtering. Full hashes and row counts are documented in the repository.

These are voluntary nonprobability samples. 'Stack Overflow survey respondents' is the population label used throughout. The sample includes professional developers, learners, hobbyists, former developers, and other people who code. Results are not population prevalence estimates.

Sample	n	Role in analysis
2025 total CSV rows	49,191	Survey file
Answered AI-use item	33,720	68.5% of CSV rows
Complete use, stance, and trust	33,231	Full-sample sensitivity
Professional developers	25,698	77.3% of complete cases
Current AI users	26,102	Primary analysis
Professional current users	20,760	Population-label sensitivity

Table 1. Analysis samples from the downloaded 2025 public CSV.

Measures and primary outcome

- Primary outcome: among current users, daily use versus weekly, monthly, or infrequent use.
- Full-sample sensitivity: daily use versus every other answered AI-use response, including nonuse.
- Favorable stance: the six AISent labels are modeled categorically; no equal spacing is assumed.
- Accuracy trust: the five AIAcc labels are modeled categorically.
- Context: work and coding experience, age, professional status, employment group, role, organization size, work mode, perceived AI threat, and country. Infrequent-country handling is learned inside each training fold.

DELIBERATE EXCLUSION AI-agent use, workflow integration, task-complexity evaluations, frustration items, and learning-route fields are excluded because they are conditional on, downstream from, or behaviorally too close to current use.

Analytic strategy

Six logistic specifications use common five-fold stratified splits: stance only; trust only; context; context plus trust; context plus stance; and context plus trust plus stance. Metrics include accuracy, balanced accuracy, ROC AUC, log loss, and Brier score. Preprocessing occurs within each fold.

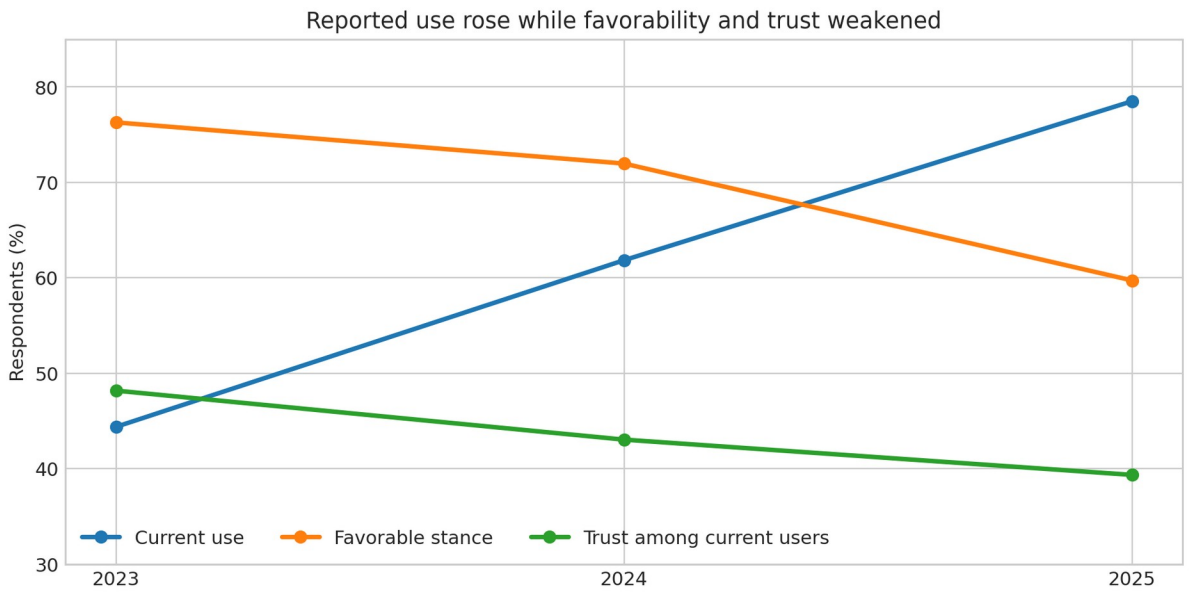
Robustness checks include 50 matched repeated splits, a separate 80/20 held-out calibration and permutation check, country-grouped folds, professional-current-user restriction, alternative outcome thresholds, and the full complete-case sample. None identifies causal direction.

4. Denominator-aware three-year context

Among respondents who answered the AI-use item, current use increased from 44.4% in 2023 to 61.8% in 2024 and 78.5% in 2025, while favorable stance fell from 76.3% to 59.7%. The 2025 use item had a lower response rate and introduced frequency categories. These are repeated cross-sections, not a panel or a clean population trend.

Indicator	2023	2024	2025	Interpretation
Current use among AI-use respondents	44.4%	61.8%	78.5%	Wording and response-rate caveat
Favorable stance	76.3%	72.0%	59.7%	Repeated cross-sections
Trust among current users	48.2%	43.0%	39.3%	Harmonized denominator
Distrust among current users	24.2%	30.4%	37.3%	Harmonized denominator

Table 2. Repeated cross-sectional context. Trust is restricted to current users for comparability.



Repeated cross-sections; use wording changed. Trust uses the current-user denominator in every year.

Figure 1. Reported use rose while favorability and harmonized current-user trust weakened.

The trust routing change matters. In 2023 and 2024, the trust item was routed to current users. In 2025, nonusers also answered it. Restricting 2025 to current users yields 39.3% trust and 37.3% distrust, compared with 43.0% and 30.4% in 2024.

5. Primary result: frequency among current users

The primary restriction removes nonusers and people who only plan to use AI. Within the 26,102 current users, daily use is 18.8% for a very unfavorable stance and 88.1% for a very favorable stance. The middle categories are not strictly monotonic: unsure is 34.0% and indifferent is 31.5%. The defensible description is steep separation between unfavorable and favorable positions, not a perfectly even ordinal ladder.

Favorable stance	n	Daily use	95% Wilson interval
Very unfavorable	585	18.8%	15.8-22.2%
Unfavorable	1,928	20.1%	18.3-21.9%
Unsure	288	34.0%	28.8-39.7%
Indifferent	4,353	31.5%	30.2-32.9%
Favorable	11,540	62.0%	61.1-62.9%
Very favorable	7,408	88.1%	87.4-88.8%

Table 3. Daily versus less-frequent use among current AI users, n=26,102.

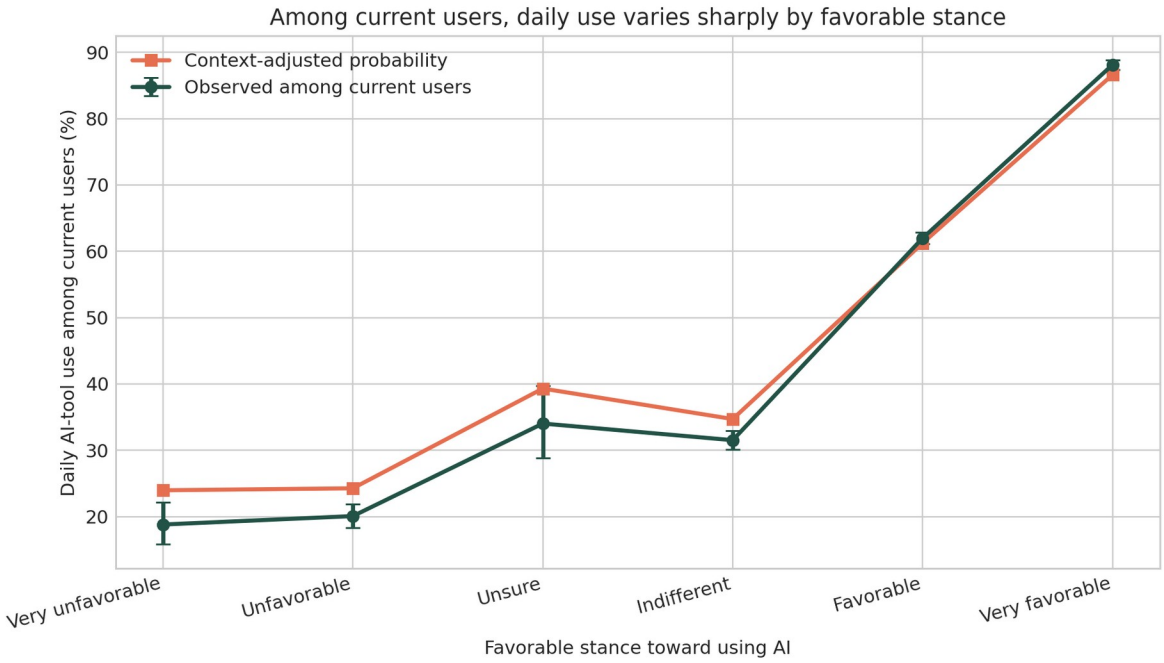


Figure 2. Observed and context-adjusted daily-use probabilities among current AI users.

6. Comparative classification

Among current users, stance alone reaches AUC 0.758; trust alone reaches 0.669. Context plus trust reaches 0.704, while context plus stance reaches 0.790. The complete model reaches 0.793. Accuracy trust is not irrelevant, but it adds little frequency-classification information after stance is known.

Model	Accuracy	ROC AUC	Log loss	Brier
Majority baseline	0.599	0.500	0.673	0.240
Stance only	0.723	0.758	0.556	0.188
Trust only	0.654	0.669	0.627	0.219
Context	0.624	0.627	0.649	0.229
Context + trust	0.666	0.704	0.610	0.211
Context + stance	0.729	0.790	0.539	0.181
Context + trust + stance	0.731	0.793	0.536	0.180

Table 4. Five-fold cross-validation among current users. Higher AUC and accuracy are better; lower log loss and Brier are better.

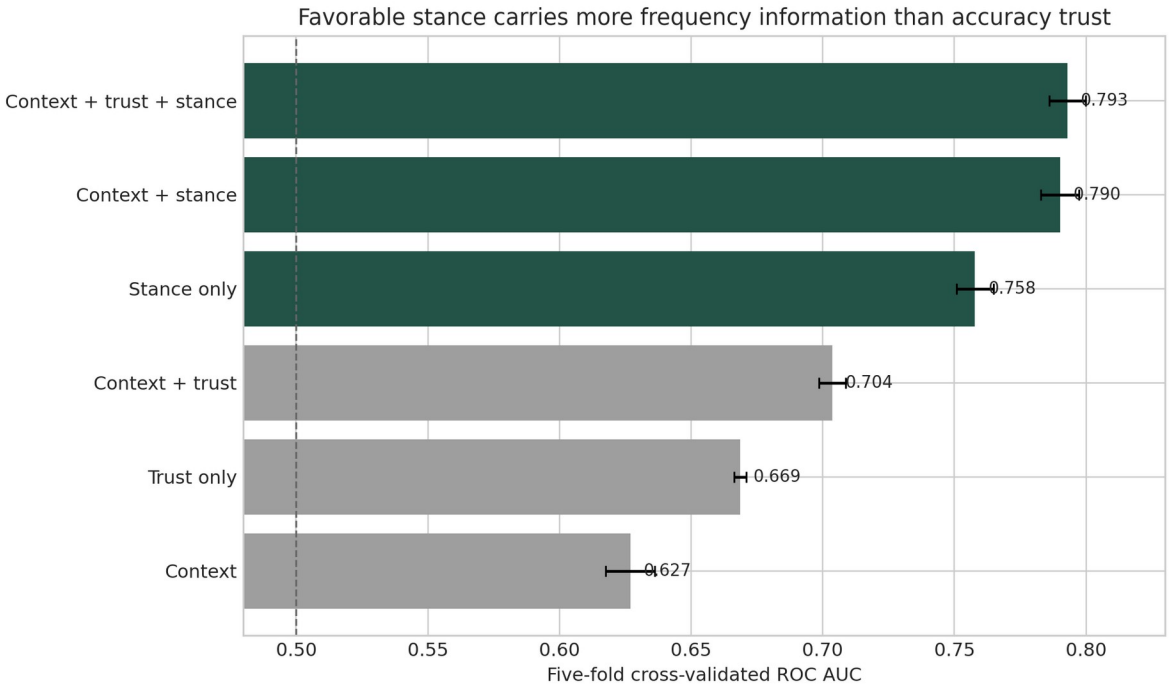


Figure 3. Cross-validated AUC for daily versus less-frequent use among current users.

Across 50 matched repeated splits, context plus stance exceeds context plus trust by 0.087 AUC on average (2.5th to 97.5th empirical split percentiles: 0.076 to 0.098) and wins in all 50 splits. Adding trust after stance improves AUC by 0.003 on average and is positive in all 50 splits. These overlapping resamples are stability summaries, not independent confidence intervals.

On a separate 20% holdout, the complete model reaches AUC 0.789, calibration slope 0.96, and mean absolute observed-versus-predicted decile gap 0.016. This supports internal calibration, not external transportability.

7. Robustness and full-sample amplification

In the full 33,231-person complete-case sample, daily use runs from 3.4% to 85.7% across the stance extremes, and stance AUC is 0.822 versus 0.728 for trust. This is valid descriptively, but the larger separation partly reflects mixing users and nonusers while the stance item asks about favoring AI use.

Check	n	Stance AUC	Trust AUC	Interpretation
Current users only	26,102	0.758	0.669	Primary frequency comparison
Professional current users	20,760	0.763	0.675	Population-label sensitivity
Country-grouped current users	26,102	0.756	0.665	173 countries
Full complete-case sample	33,231	0.822	0.728	Includes nonusers; amplified

Table 5. Robustness checks. Country-grouped folds are not external organizational validation.

Stance classification is stronger for broader thresholds such as any current use or current use plus plans. That consistency is also a warning: the broader the target becomes adoption versus nonadoption, the closer it is to the wording of a favorable stance toward using AI.

8. Favorable stance is not accuracy trust

The within-trust comparison is now restricted to current users. Among current users who highly distrust AI accuracy, daily use is 17.7% for the very unfavorable (n=389) and 85.1% for the very favorable (n=241). Broad favorability can coexist with skepticism about individual outputs.

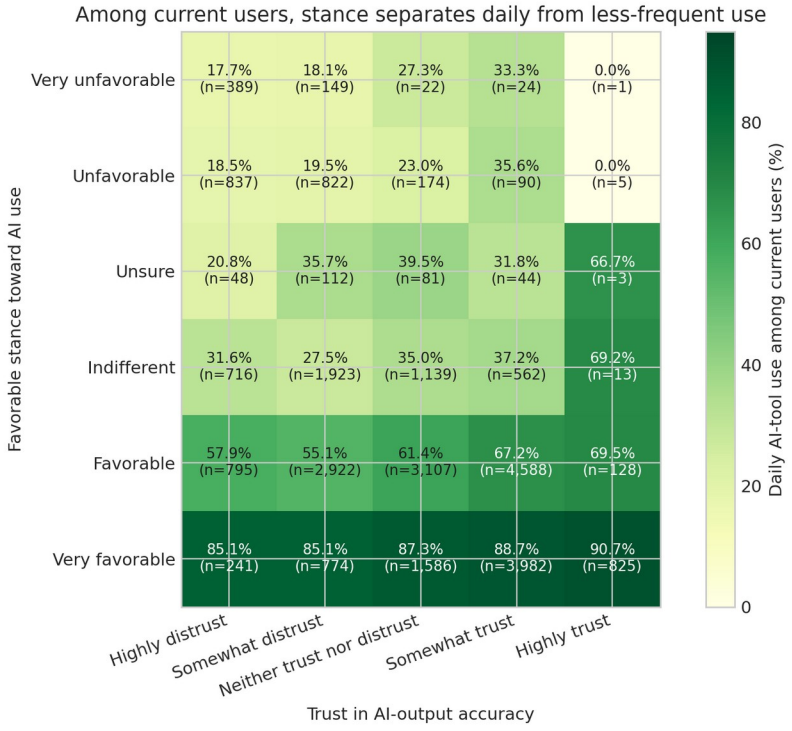


Figure 4. Daily-use rates across favorable stance and accuracy trust among current AI users.

INTERPRET CAREFULLY Restricting to current users removes the nonuser contrast but conditions on adoption. It is a tougher descriptive sensitivity check, not a causal fix.

9. The strongest counterarguments

The stance item is broader and closer to the outcome

A broad question about favoring AI use should align more closely with self-reported AI use than a narrow question about accuracy. The evidence supports 'accuracy trust alone is an incomplete summary of use orientation,' not 'stance is psychologically more important.'

Use may shape stance

Daily use may make a respondent more favorable through experienced value or less favorable through repeated failures. Temporal order and reciprocal effects cannot be recovered from one survey wave.

Same-wave self-report can inflate coherence

Stance, trust, and use are reported in the same session. Consistency motives, shared interpretation, and omitted variables can strengthen alignment. Cross-validation and calibration do not remove common-method bias.

Classification importance is not managerial importance

Accuracy trust adds little AUC after stance, but it remains essential for verification, over-reliance, and safety. A variable can matter greatly to outcomes while adding little information about use frequency.

10. What organizations can responsibly do

The results justify separate measurement and longitudinal testing, not persuasion campaigns. The management question is not how to maximize enthusiasm or trust. It is how to create useful, appropriately skeptical, well-verified use.

Dimension	Example measures	What it answers
Behavior	Frequency, task, workflow depth	What people do
Favorable stance	Broad evaluation of using AI	Perceived value, fit, or willingness
Accuracy trust	Confidence in output correctness	Belief about claims and answers
Verification burden	Testing, review, rework, defects	Cost of appropriate skepticism
Outcomes	Quality, speed, learning, risk	Whether use creates value safely

Table 6. An analytically distinct adoption-measurement dashboard.

Test interventions longitudinally

- Measure stance, accuracy trust, expected value, and verification practices before a change.
- Use behavioral telemetry and quality outcomes alongside self-report.
- Randomize or stagger workflow support, task examples, or verification scaffolds where feasible.
- Re-measure after real work rather than immediately after a demonstration.
- Look for calibrated reliance rather than maximum trust.

11. Limitations

- Voluntary nonprobability sample; prevalence does not represent all developers.

- Only 77.3% of complete cases identify as professional developers.
- Repeated cross-sections; no individual change is observed across years.
- Routing and wording changes; the 2025 AI-module response rate is lower.
- Single-item constructs; favorable stance and accuracy trust are not validated scales.
- Same-wave self-report; temporal order, common-method bias, and objective behavior are unresolved.
- Criterion proximity; favoring AI use is semantically closer to use frequency than accuracy trust is.
- Conditioning; the current-user restriction reduces one artifact but is not a causal design.
- Model transportability; internal and country-grouped resampling do not prove performance in a new organization or future year.
- Unmeasured confounding; availability, mandates, task mix, tool quality, usefulness, and prior experience may explain the association.
- No objective calibration measure; the survey cannot determine whether trust is warranted.

12. Conclusion

The familiar story is adoption rising while trust falls. The more precise finding is that a broad favorable stance and accuracy-specific trust carry different information about use frequency. Even among people already using AI, favorable stance separates daily from less-frequent use substantially better than accuracy trust does. The result survives professional-current-user and country-grouped checks.

The same evidence blocks the tempting causal story. Favorable stance is broad, criterion-proximal, and potentially shaped by use. It should be treated as an analytically distinct survey item, not a lever proven to drive adoption. Organizations should measure stance, trust, behavior, verification, and outcomes separately, then test what changes.

> **BOTTOM LINE** Access is not adoption. Favoring use is not trusting output. Frequent use is not calibrated reliance. Measure each one - then test what changes.

Technical appendix

A.1 Reproducibility

The repository includes source-data hashes, recodes, denominator notes, machine-readable results, analysis and publication scripts, and figures. Raw survey files are excluded because of size and licensing; official download paths and integrity checks are documented.

A.2 Claim ladder

Evidence level	Claim
Supported descriptively	Daily-use rates differ sharply across stance categories among current users.
Supported within this survey	Stance adds substantially more cross-validated frequency-classification information than accuracy trust.
Plausible but unresolved	Stance influences use, use influences stance, or both.
Not supported	Changing employee stance will cause greater or better adoption.
Not supported	Trust is unimportant because it adds little AUC after stance.

Table A1. Boundaries for publication and public communication.

References

Choudhuri, R., Trinkenreich, B., Pandita, R., Kalliamvakou, E., Steinmacher, I., Gerosa, M., Sanchez, C. A., & Sarma, A. (2025). What guides our choices? Modeling developers' trust and behavioral intentions towards GenAI. Proceedings of ICSE 2025, 1691-1703. https://doi.org/10.1109/ICSE55347.2025.00087

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340. https://doi.org/10.2307/249008

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407-434. https://doi.org/10.1177/0018720814547570

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). Trust, attitudes and use of artificial intelligence: A global study 2025. University of Melbourne & KPMG. https://doi.org/10.26188/28822919

Stack Overflow. (2023). 2023 Developer Survey. https://survey.stackoverflow.co/2023

Stack Overflow. (2024). 2024 Developer Survey: AI. https://survey.stackoverflow.co/2024/ai

Stack Overflow. (2025). 2025 Developer Survey: AI. https://survey.stackoverflow.co/2025/ai

Stack Overflow. (2026, February 18). Mind the gap: Closing the AI trust gap for developers. https://stackoverflow.blog/2026/02/18/closing-the-developer-ai-trust-gap/

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478. https://doi.org/10.2307/30036540

Wang, R., Cheng, R., Ford, D., & Zimmermann, T. (2024). Investigating and designing for trust in AI-powered code generation tools. Proceedings of FAccT 2024. https://doi.org/10.1145/3630106.3658984