Software Engineering and Artificial Intelligence Systems

Volume: 02 Issue: 01 (2026)

www.seais.org/

When Trust Becomes Substitution: Generative-AI Trust, Over-Reliance, and Learner Autonomy Among University Students

Daniel Foster and Olivia Young*
Marsal Family School of Education, University of Michigan, Ann Arbor, MI, USA
Corresponding author: olivia1.yng@umich.edu

Keyword: generative AI, trust calibration, over-reliance, learner autonomy, response surface analysis, latent class analysis, higher education

A B S T R A C T

Generative artificial intelligence (GenAI) has become a routine study companion for university students, yet the conditions under which trust in these tools turns into over-reliance that erodes learner autonomy remain poorly specified. This study models trust, over-reliance, and learner autonomy jointly with the 2025 Stack Overflow Developer Survey, analysing 1,308 university students from 102 countries and a contrast group of 9,860 professional developers. Over-reliance is operationalized through task-level delegation that separates substitution (core learning tasks done mostly by AI) from augmentation (tasks done partially with AI), and learner autonomy is measured through four facets: problem-solving confidence, comprehension ownership, epistemic agency, and self-directed resource diversity. Trust predicted substitution but not augmentation. Polynomial regression with response surface analysis showed that substitution was higher when trust exceeded students' own appraisal of AI capability than in the reverse configuration ($a3 = 0.146$, $p = .005$), establishing calibration rather than trust level alone as a driver of over-reliance. Trust and substitution both predicted lower epistemic agency, and causal mediation analysis confirmed a significant indirect path (ACME = −0.034, 95% CI [−0.061, −0.012]). Students and professionals reported identical trust, yet students reported problem-solving confidence loss at nearly twice the professional rate (43.8% vs. 23.0%), a gap unchanged after reweighting professionals to the students' reliance profile. Latent class analysis identified four reliance profiles, among which dependent delegators (15.9%) showed the weakest epistemic agency. The findings position trust calibration and reliance mode as actionable levers for autonomy-preserving GenAI integration in higher education.

INTRODUCTION

Generative artificial intelligence (GenAI) tools such as conversational assistants and code generators have moved from novelty to infrastructure in higher education within a few academic years. Students use them to explain concepts, draft and debug code, and plan study tasks, and the educational potential of these systems has been set out in detail [1]. Surveys of university students report broad enthusiasm tempered by concerns about accuracy, integrity, and the preservation of independent thinking [2], and computing educators regard capable code generators as a turning point for how programming is taught and assessed [3], [4].

Trust sits at the centre of this transition. In the human–automation literature, trust governs whether people rely on a system, and appropriate reliance requires that trust be calibrated to what the system can actually do [5], [6]. Miscalibration takes two forms. Distrust leads to the disuse of helpful tools, whereas excessive trust produces misuse, complacency, and automation bias, in which automated output displaces the user's own judgment [7], [8]. Recent experiments show that people follow AI advice even when it contradicts the available evidence [9], and confidence in GenAI is associated with less critical thinking among knowledge workers [10]. For learners the stakes are distinct. Learning depends on effortful processing, error detection, and self-regulation, all of which a fluent assistant can quietly absorb. Field and laboratory studies substantiate this concern. Unrestricted access to a GPT-4 tutor improved practice performance but lowered subsequent unaided performance in high-school mathematics [11]; ChatGPT support during a writing task was accompanied by fewer metacognitive processes [12]; and students who researched a topic with an LLM reported lower cognitive load but produced shallower reasoning than peers who used web search [13]. For STEM higher education, Mai et al. [14] provided an insightful mixed-methods account of AI tutoring, showing that moderate trust supports learning effectiveness and confidence whereas over-trust weakens critical thinking and independent problem solving. Their work offers a valuable foundation for examining how trust shapes the quality, and not only the quantity, of student reliance. Four gaps limit current understanding. First, trust is usually treated as a level, although calibration theory predicts that the fit between trust and perceived capability matters in its own right [5]. Second, reliance is commonly indexed by frequency of use, which conflates augmentation, where the learner remains the author of the work, with substitution, where the tool performs the task. Third, learner autonomy, classically defined as the capacity to take charge of one's own learning [15], is rarely measured across its cognitive, epistemic, and self-directive facets. Fourth, few studies contrast learners with experienced practitioners under a common instrument, which is required to establish whether novices are distinctively vulnerable. This study addresses these gaps with the public microdata of the 2025 Stack Overflow Developer Survey [16], which records trust in AI output, appraised AI capability, task-level delegation, and several autonomy-relevant experiences for tens of thousands of programmers, including a large group of university students. The paper makes four contributions. (i) It introduces a reliance-mode framework that separates the substitution of core learning tasks from their augmentation and links both to a four-facet measure of learner autonomy. (ii) It applies polynomial regression with response surface analysis to test trust calibration directly instead of inferring it from trust levels. (iii) It quantifies a learner-status gap in autonomy erosion that persists when professionals are reweighted to the students' exact trust and reliance profile. (iv) It identifies latent reliance profiles that separate autonomy-preserving from autonomy-eroding patterns of GenAI use.

II. Related Work and Hypotheses

A. Trust Calibration and Over-Reliance

Trust in automation is an attitude that guides reliance under uncertainty, and its appropriateness is judged by calibration, resolution, and specificity relative to system capability [5]. Reviews of trust in automation and in

AI identify dispositional, situational, and learned sources of trust and show that perceived performance is its strongest antecedent [17], [18]. When trust exceeds capability, users exhibit complacency and automation bias [7], [8]. Over-reliance on AI is accordingly defined as the acceptance of AI output that is incorrect or that users could have evaluated themselves [19]. Cognitive forcing functions [20] and explanations that lower verification costs [21] reduce over-reliance, and the appropriateness of reliance can be quantified as a two-dimensional construct [22]. Two implications follow for GenAI in education. The first concerns the behavioural expression of trust. A student who trusts GenAI output may use it to check an idea while retaining authorship (augmentation) or may hand the task over entirely (substitution). Calibration theory predicts that trust primarily licenses the second pattern, because only substitution requires accepting output without independent evaluation. The second concerns the reference point for calibration. Objective accuracy is rarely visible to students, but their own appraisal of what the tool can handle is. Trust that outruns this appraisal represents reliance beyond a perceived competence boundary, the configuration most likely to produce uncritical acceptance.

H1*:* Trust in GenAI is positively associated with substitution-mode reliance but not with augmentation-mode reliance.

H2: At comparable levels, trust that exceeds appraised capability (over-trust) is associated with more substitution than appraised capability that exceeds trust (under-trust).

B. Generative AI, Cognitive Offloading, and Learning

Cognitive offloading, the use of external resources to reduce internal cognitive demand, improves immediate performance but can impair memory and skill formation when it replaces rather than supports processing [23]. Survey evidence links frequent AI tool use to weaker critical thinking through increased offloading [24], and a systematic review attributes reduced decision-making, analytical, and critical-thinking abilities to over-reliance on AI dialogue systems [25]. Among university students, AI dependency is associated with academic stress and performance expectations and is perceived to increase laziness and reduce independent thinking [26]. An experiment on student agency showed that learners supported by AI feedback kept relying on it instead of internalizing the underlying skill once support was withdrawn [27].

H3: Substitution-mode reliance is associated with lower learner autonomy.

H4: Trust in GenAI is associated with learner autonomy indirectly through substitution-mode reliance.

C. Learner Autonomy and Novice Vulnerability

Learner autonomy denotes the ability to take charge of one's own learning, including setting goals, selecting resources, monitoring progress, and evaluating outcomes [15], [28]. Later work treats autonomy as a multidimensional capacity with technical, psychological, and political aspects [29] that is closely related to self-regulated learning [30]. In a GenAI context four facets are directly at stake: confidence in one's own problem solving, which rests on mastery experiences [31]; ownership of comprehension, meaning that one understands the solutions one submits; epistemic agency, the inclination to seek understanding and compare alternatives instead of accepting a single answer; and self-directed resource management, the breadth of learning resources a student chooses beyond the AI tool. Novices may be more exposed on each facet. Studies of introductory programming show that AI code generators help novices complete tasks [32] but can widen the gap between students with strong and weak metacognitive skills, fostering an illusion of competence among the latter [33]. Experienced developers possess the schemas needed to verify output, whereas students are still building them.

H5: University students report more autonomy erosion than professional developers at comparable levels of trust and reliance.

A person-centred research question (RQ) complements the hypotheses: which latent reliance profiles exist among students, and how do these profiles differ in learner autonomy? Fig. 1 summarizes the framework.

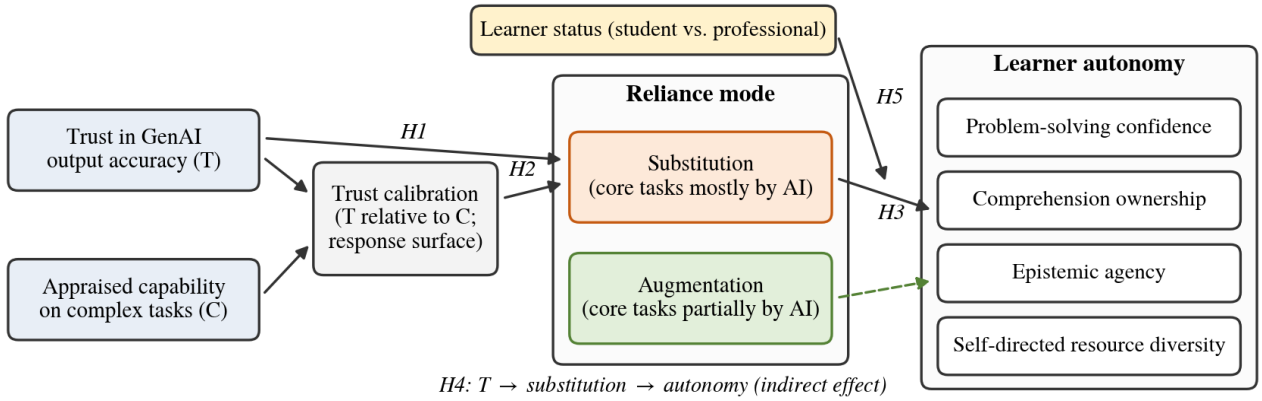


Figure 1. Conceptual framework linking trust calibration, reliance mode, and learner autonomy, with learner status as moderator. H1–H5 denote the hypotheses.

III. Method

A. Data Source and Samples

The study analyses the public microdata of the 2025 Stack Overflow Developer Survey [16], fielded online from 29 May to 23 June 2025 in 177 countries. The released file contains 49,193 unique respondent records and 172 variables, including a dedicated section on AI tools. University students were identified as respondents whose current employment status was student, who were 18 to 34 years old, and whose highest completed education was secondary school or higher, which places them in tertiary study (n = 4,001). The analytic sample retained students who used AI tools at least monthly and answered every item entering the models, yielding 1,308 students from 102 countries (86.8% aged 18–24; largest national groups: India, Germany, and the United States; median coding experience 5 years). A contrast group comprised employed professional developers who met the same criteria (n = 9,860 from 147 countries; median experience 14.5 years).

B. Measures

Table I summarizes the constructs. *Trust* was measured with the item on trust in the accuracy of AI output (1 = highly distrust to 5 = highly trust) and *appraised capability* with the item on how well AI tools handle complex tasks (1 = very poor to 5 = very well); respondents who did not use AI for complex tasks were outside the analytic sample. The two items correlated at $\rho = 0.62$ among students, high enough to reflect a shared evaluative core yet low enough to leave 54.4% of students with discrepant ratings, which is the variance a calibration analysis requires.

Table 1CONSTRUCTS AND OPERATIONALIZATION

Construct	**Operationalization (survey item)**	**Coding**
Trust in GenAI (T)	Trust in the accuracy of AI output in the development workflow	1–5
Appraised capability (C)	How well AI tools handle complex tasks	1–5
Use frequency	Current use of AI tools (monthly, weekly, daily)	0–2
Substitution (SDI)	Core tasks done mostly with AI: writing code, debugging, learning concepts, learning a codebase	Count 0–4
Augmentation (AUI)	The same core tasks done partially with AI	Count 0–4
AI as learning resource	AI code-generation tools or AI-enabled apps used to learn coding in the past year	0/1
Confidence loss	Frustration: "I've become less confident in my own problem-solving"	0/1
Comprehension difficulty	Frustration: "It's hard to understand how or why the code works"	0/1
Epistemic agency	Would still ask a person to fully understand, to learn best practices, to compare solutions	Count 0–3
Resource diversity	Non-AI resources used to learn coding in the past year (11 options)	Count 0–11

Confidence loss and comprehension difficulty index autonomy erosion; epistemic agency and resource diversity index retained autonomy.

Reliance mode was derived from a task-by-mode matrix in which respondents indicated, for each workflow task, whether they currently performed it mostly with AI, partially with AI, or without AI. Four tasks that constitute core learning activity in programming were used: writing code, debugging or fixing code, learning new concepts or technologies, and learning about a codebase. The *substitution delegation index* (SDI) counts

tasks done mostly with AI (0–4), and the *augmentation index* (AUI) counts tasks done partially with AI (0–4). Use frequency and the use of AI tools as a learning resource served as additional reliance indicators.

Learner autonomy was represented by four facets. *Problem-solving confidence loss* is the endorsement of "I've become less confident in my own problem-solving" among AI-related frustrations, and *comprehension difficulty* is the endorsement of "It's hard to understand how or why the code works"; both index autonomy erosion. *Epistemic agency* (0–3) counts the situations in which a respondent would still consult a person even if AI could do most coding: to fully understand something, to learn best practices, and to compare different solutions. *Self-directed resource diversity* (0–11) counts the non-AI resources used to learn coding in the past year, such as documentation, books, courses, videos, and community forums. Covariates were age group, a bachelor-or-higher indicator, log-transformed years of coding, coding role, and country fixed effects for countries with at least 25 students.

C. Analytic Strategy

Group differences were tested with Welch t tests and χ^2 tests and expressed as Cohen's d or h. Within the student sample, SDI, AUI, and epistemic agency were modelled with ordinary least squares and HC3 robust errors, the binary facets with logistic regression, and resource diversity with Poisson regression; ordered logit models served as robustness checks for the ordinal outcomes.

Trust calibration was tested with polynomial regression and response surface analysis [34], [35]. With trust T and appraised capability C centred at the scale midpoint, each outcome Y was modelled as

$$Y = b_0 + b_1T + b_2C + b_3T^2 + b_4TC + b_5C^2 + \gamma'x + \varepsilon \quad (1)$$

where x contains use frequency and the covariates. The surface is summarized by four parameters,

$$a_1 = b_1 + b_2, \quad a_2 = b_3 + b_4 + b_5, \quad (2)$$

$$a_3 = b_1 - b_2, \quad a_4 = b_3 - b_4 + b_5. \quad (3)$$

Here a_1 and a_2 give the slope and curvature along the congruence line ($T = C$), and a_3 and a_4 give them along the incongruence line ($T = -C$). A positive a_3 indicates higher outcomes when trust exceeds appraised capability than in the mirror configuration, which is the test of H2. Standard errors were computed from the full coefficient covariance matrix. Indirect effects (H4) were estimated in the causal mediation framework of Imai et al. [36] by quasi-Bayesian Monte Carlo simulation with 5,000 draws. The exposure contrast compared "somewhat distrust" with "somewhat trust" ($T = 2$ vs. 4), SDI served as mediator, and each autonomy facet as outcome, with capability, frequency, AUI, and covariates included in both equations. The average causal mediation effect (ACME) and average direct effect (ADE) are reported with 95% intervals.

Learner-status differences (H5) were estimated in pooled logistic models ($n = 11{,}168$) with student × reliance interactions. To compare students with professionals who hold identical trust and reliance profiles, professionals were reweighted with propensity-based odds weights [37] estimated from trust, capability, frequency, SDI, AUI, and AI-based learning, and balance was checked with standardized mean differences (SMD). Average marginal effects (AME) of student status were obtained from counterfactual predictions with 200 bootstrap resamples, and a robustness model restricted professionals to the students' age band of 18–34 years. Latent class analysis [38-40] was fitted to six categorical indicators (frequency, trust, capability, SDI, AUI, and AI-based learning) for $K = 1$ to 6 classes with eight random starts each. The Bayesian information criterion (BIC) guided model selection, subject to a minimum class share of 5%.

IV. Results

A. Students Versus Professional Developers

Table II contrasts the two groups. Students and professionals reported virtually identical trust in AI output (2.90 vs. 2.88, $d = 0.02$) and appraised capability (2.72 vs. 2.71, $d = 0.01$). Students used AI less often (daily

use 53.7% vs. 63.9%) and relied less on it as a learning resource (48.0% vs. 54.2%), yet they delegated more core tasks in substitution mode (SDI 0.51 vs. 0.36, $d = 0.17$). The largest differences concerned autonomy erosion: 43.8% of students reported problem-solving confidence loss compared with 23.0% of professionals ($h = 0.45$), and 32.4% reported comprehension difficulty compared with 17.0% ($h = 0.36$). Students also showed higher epistemic agency (1.73 vs. 1.63) and broader self-directed learning (4.41 vs. 4.00 resources). Fig. 2 shows that the erosion gap holds at every trust level and that debugging is the task students substitute most distinctively (14.1% vs. 7.6%).

Table 2STUDENTS VERSUS PROFESSIONAL DEVELOPERS

Indicator	**Students (n = 1,308)**	**Professionals (n = 9,860)**	**ES**
Trust in AI output (1–5)	2.90 (1.13)	2.88 (1.10)	0.02
Appraised capability (1–5)	2.72 (1.25)	2.71 (1.21)	0.01
Daily AI use	53.7%	63.9%	−0.21
Substitution, SDI (0–4)	0.51 (0.95)	0.36 (0.80)	0.17
Augmentation, AUI (0–4)	1.60 (1.28)	1.69 (1.33)	−0.07
AI as learning resource	48.0%	54.2%	−0.12
Confidence loss	43.8%	23.0%	0.45
Comprehension difficulty	32.4%	17.0%	0.36
Epistemic agency (0–3)	1.73 (1.02)	1.63 (1.04)	0.10
Resource diversity (0–11)	4.41 (1.92)	4.00 (1.87)	0.21

Means (SD) or percentages. ES = Cohen's d for means and h for proportions (students minus professionals). All differences $p < .001$ except trust and capability (n.s.) and augmentation ($p = .013$).

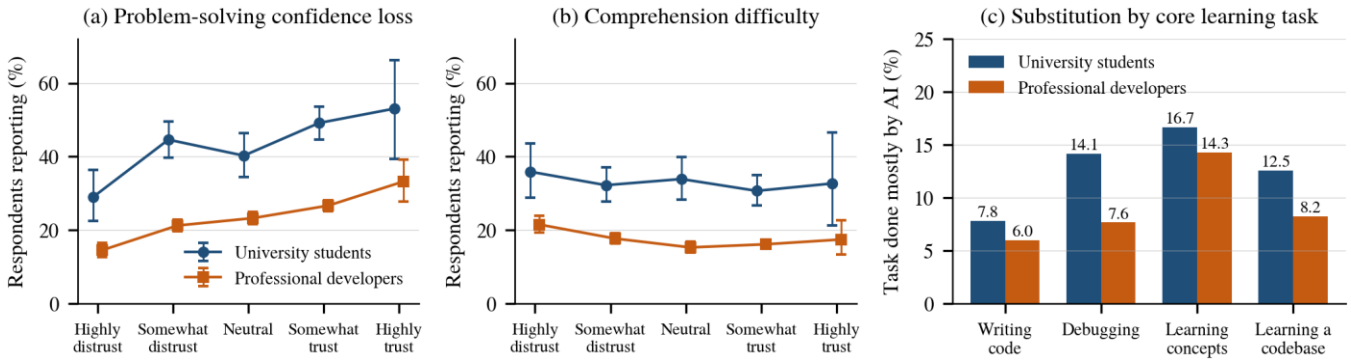


Figure 2Students versus professional developers: (a) problem-solving confidence loss and (b) comprehension difficulty by trust level, with 95% Wilson intervals; (c) share of respondents performing each core learning task mostly with AI.

B. Trust, Reliance Mode, and Autonomy Among Students

Table III reports the student models. In line with H1, trust predicted substitution ($b = 0.163$ per scale point, $p < .001$) but not augmentation ($b = -0.021$, $p = .598$). Use frequency predicted both modes, and coding experience predicted augmentation positively and substitution negatively, indicating that experience channels AI use toward collaborative rather than delegative patterns. The ordered logit specification reproduced the trust–substitution association ($b = 0.538$, $p < .001$).

H3 was supported for epistemic agency. Each additional core task delegated in substitution mode was associated with 0.105 fewer understanding-oriented reasons for consulting people ($p = .005$), trust showed a further direct association of −0.110 ($p = .001$), and augmentation was unrelated ($b = 0.014$, $p = .555$). Problem-solving confidence loss was associated with use frequency (OR = 1.18, $p = .049$) and fewer years of coding (OR = 0.70 per log-year, $p = .011$) rather than with substitution (OR = 1.01, $p = .920$). Relative to students who highly distrusted AI, those who somewhat distrusted it (OR = 1.76, $p = .007$) or somewhat trusted it (OR = 1.63, $p = .037$) more often reported confidence loss, so confidence erosion accompanies substantive engagement of any kind beyond outright distrust. Comprehension difficulty was lower at higher trust (OR = 0.85, $p = .032$), and resource diversity was modestly lower at higher appraised capability (IRR = 0.975, $p = .039$).

Table 3STUDENT MODELS OF RELIANCE MODE AND LEARNER AUTONOMY

Predictor	**SDI (b)**	**AUI (b)**	**Confidence loss (OR)**	**Comprehension difficulty (OR)**	**Epistemic agency (b)**	**Resource diversity (IRR)**
Trust (T)	0.163***	−0.021	1.08	0.85*	−0.110***	0.982
Appraised capability (C)	0.062*	0.029	1.07	1.01	−0.050	0.975*
Use frequency	0.301***	0.366***	1.18*	1.00	−0.078	1.009
Substitution (SDI)	—	—	1.01	1.12	−0.105**	0.977
Augmentation (AUI)	—	—	1.10	1.06	0.014	1.010
Age 25–34	−0.024	−0.128	0.84	0.81	−0.028	0.941
Bachelor or higher	0.031	−0.052	1.04	1.03	−0.030	1.003
Log years of coding	−0.133*	0.393***	0.70*	0.64**	−0.071	1.039
Role: learning to code	−0.063	0.043	1.03	1.23	0.046	1.121***
Role: professional developer	−0.053	0.068	0.99	1.23	0.062	1.103***
R^2 / pseudo-R^2	0.208	0.084	0.025	0.021	0.077	—
AUC	—	—	0.607	0.598	—	—

n = 1,308 students. OLS with HC3 errors (SDI, AUI, epistemic agency), logistic regression (binary facets), and Poisson regression (resource diversity), all with country fixed effects. $p < .05$, $p < .01$, ** $p < .001$.

C. Calibration Surfaces

Trust and appraised capability were frequently incongruent: 33.6% of students trusted AI output more than they rated its capability (over-trust), and 20.9% showed the reverse pattern. Table IV and Fig. 3 present the response surfaces. For substitution, the slope along the congruence line was positive ($a_1 = 0.285$, $p < .001$) and convex ($a_2 = 0.152$, $p < .001$), so substitution accelerated as trust and appraised capability rose together. Supporting H2, the slope along the incongruence line was positive ($a_3 = 0.146$, $p = .005$): at equal distance from congruence, over-trusting students substituted more than under-trusting students. The positive curvature along the incongruence line ($a_4 = 0.117$, $p = .0497$) further shows that substitution grows as the discrepancy widens, with the steeper side in the over-trust region. Epistemic agency and resource diversity declined along the congruence line ($a_1 = −0.189$ and $−0.222$, both $p < .001$) without incongruence effects, so these facets respond to the joint level of trust and appraised capability rather than to their mismatch. No surface parameter reached significance for confidence loss.

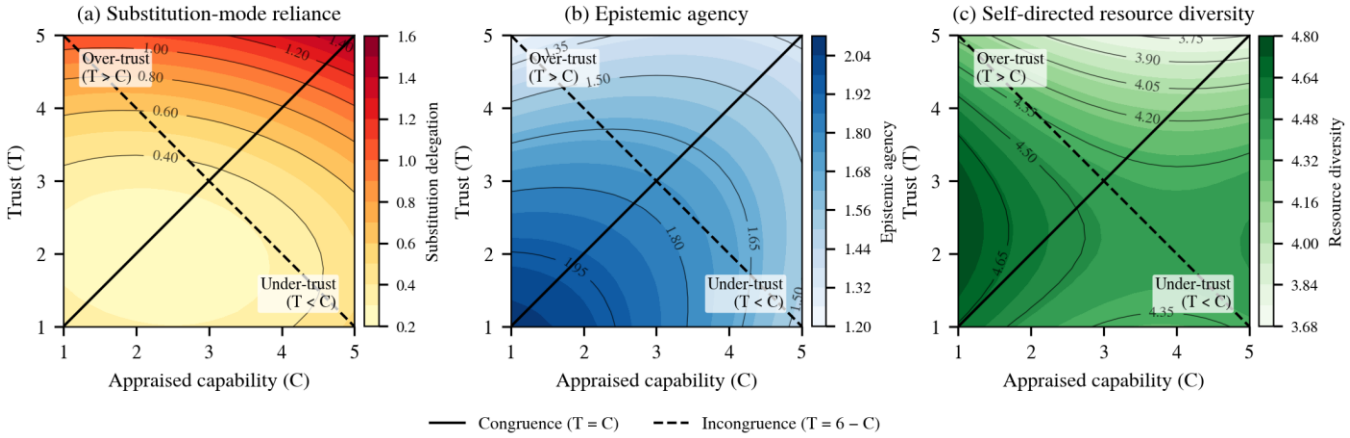


Figure 3Response surfaces of average adjusted predictions over trust (T) and appraised capability (C) for (a) substitution delegation, (b) epistemic agency, and (c) self-directed resource diversity among students. Solid and dashed lines mark the congruence and i

Table 4RESPONSE SURFACE PARAMETERS (STUDENTS)

Parameter	**SDI**	**Epistemic agency**	**Resource diversity**	**Confidence loss**
a_1	0.285*** (0.032)	−0.189*** (0.032)	−0.222*** (0.066)	0.102 (0.066)
a_2	0.152*** (0.027)	−0.011 (0.027)	−0.058 (0.057)	−0.090 (0.058)
a_3	0.146** (0.052)	−0.059 (0.060)	−0.053 (0.113)	0.009 (0.117)
a_4	0.117* (0.060)	−0.102 (0.069)	−0.039 (0.130)	−0.024 (0.136)
R^2 / pseudo-R^2	0.234	0.069	0.047	0.025

Estimates with standard errors in parentheses; confidence loss on the logit scale. Models include use frequency and covariates. $p < .05$, $p < .01$, ** $p < .001$.

D. Indirect Effects

Table V summarizes the mediation results. Moving from "somewhat distrust" to "somewhat trust" raised SDI by 0.33 tasks ($p < .001$). For epistemic agency, the indirect effect through substitution was significant (ACME = −0.034, 95% CI [−0.061, −0.012]) alongside a direct effect (ADE = −0.220, [−0.349, −0.089]), so substitution carried 13.4% of the total association (95% CI [4.2%, 33.4%]), which supports H4 for this facet. For comprehension difficulty the direct effect of trust was negative (ADE = −0.066, [−0.125, −0.009]) with a non-significant indirect component, and neither confidence loss nor resource diversity showed a significant indirect path.

Table 5CAUSAL MEDIATION THROUGH SUBSTITUTION

Outcome	**ACME [95% CI]**	**ADE [95% CI]**	**Total**
Epistemic agency	−0.034* [−0.061, −0.012]	−0.220* [−0.349, −0.089]	−0.255
Comprehension difficulty	0.008 [−0.002, 0.019]	−0.066* [−0.125, −0.009]	−0.058
Confidence loss	0.001 [−0.010, 0.011]	0.035 [−0.027, 0.097]	0.036
Resource diversity	−0.033 [−0.086, 0.015]	−0.155 [−0.430, 0.111]	−0.188

Effects of moving from T = 2 to T = 4 with SDI as mediator (1,308 students; 5,000 quasi-Bayesian draws); probability scale for binary outcomes. * 95% interval excludes zero. Proportion mediated for epistemic agency: 13.4%.

E. Learner Status

In the pooled sample (Table VI), student status predicted confidence loss (OR = 1.26, 95% CI [1.09, 1.47], p = .002) and comprehension difficulty (OR = 1.48, [1.26, 1.72], $p < .001$) after adjustment for trust, capability, frequency, reliance mode, education, coding experience, and country. The corresponding average marginal effects were +4.3 percentage points (95% CI [2.0, 6.5]) and +6.3 points ([3.8, 9.0]). Estimates were nearly identical when professionals were restricted to ages 18–34 (OR = 1.29 and 1.47; n = 5,135). Reweighting professionals to the students' reliance profile reduced every SMD on the reweighting variables below 0.01, yet the reweighted professional rates stayed at 23.0% for confidence loss and 17.2% for comprehension difficulty, against 43.8% and 32.4% among students. H5 is supported.

Table 6LEARNER-STATUS ESTIMATES FOR AUTONOMY EROSION

Estimate	Confidence loss	Comprehension difficulty
Adjusted OR [95% CI]	1.26 [1.09, 1.47]	1.48 [1.26, 1.72]
Adjusted OR, professionals 18–34	1.29 [1.11, 1.51]	1.47 [1.25, 1.73]
AME, percentage points [95% CI]	4.3 [2.0, 6.5]	6.3 [3.8, 9.0]
Students (observed)	43.8%	32.4%
Professionals (reweighted)	23.0%	17.2%
Professionals (unweighted)	23.0%	17.0%

Pooled models (n = 11,168) adjust for trust, capability, frequency, SDI, AUI, education, coding experience, and country. Odds weights balance trust, capability, frequency, SDI, AUI, and AI-based learning (maximum |SMD| = 0.008); reweighted means for epistemic agency and resource diversity are 1.61 and 3.96 versus 1.73 and 4.41 for students. The interaction model revealed a qualitative difference in how erosion arises. Among professionals, confidence loss rose with each substituted task (OR = 1.12, $p < .001$), whereas the student slope was flatter (student × SDI OR = 0.87, $p = .062$). Fig. 4 illustrates the consequence: professionals approach the students' level of erosion only at high substitution, while students report elevated erosion even when they substitute nothing. For learners, autonomy erosion is a pervasive correlate of GenAI use rather than a consequence confined to heavy delegation.

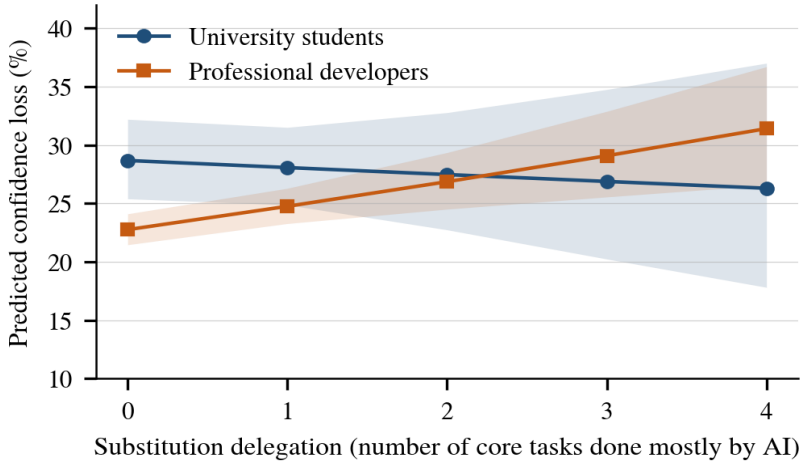


Figure 4Predicted probability of problem-solving confidence loss by substitution delegation for students and professionals (pooled logit with interactions; other covariates at pooled means; 95% confidence bands).

F. Latent Reliance Profiles

BIC reached its minimum at four classes (Table VII; entropy = 0.75; smallest class 15.9%). Fig. 5 displays the profiles. *Skeptical minimalists* (24.7%) combined low trust with infrequent use and virtually no substitution. *Critical augmenters* (30.0%) also held low trust but used AI frequently in augmentation mode, with 92% augmenting at least one core task. *Trusting collaborators* (29.4%) paired high trust and appraised capability with predominantly augmentative use. *Dependent delegators* (15.9%) used AI almost daily (96%), held high trust, and substituted two or more core tasks in 83% of cases.

Table 7LATENT CLASS MODEL FIT (STUDENTS)

K	LL	BIC	SABIC	Entropy	Smallest class
1	−7327.7	14734.3	14699.3	—	100.0%
2	−6868.6	13902.2	13829.1	0.718	49.3%
3	−6751.9	13755.0	13643.8	0.774	16.1%
4	**−6663.1**	**13663.5**	**13514.2**	**0.747**	**15.9%**
5	−6635.9	13695.3	13507.9	0.735	11.4%
6	−6616.8	13743.1	13517.6	0.726	9.9%

The selected 4-class solution (bold) has the lowest BIC among solutions whose smallest class exceeds 5%. The profiles differed in confidence loss ($\chi^2(3) = 25.5$, $p < .001$), epistemic agency ($H = 57.4$, $p < .001$), and resource diversity ($H = 29.5$, $p < .001$) but not in comprehension difficulty ($p = .692$). All three engaged profiles reported more confidence loss than skeptical minimalists (adjusted OR = 1.82 to 1.96, all $p < .001$). On the remaining facets the profiles separated clearly: dependent delegators showed the lowest epistemic agency (1.41, against 2.02 for minimalists), whereas critical augmenters retained high epistemic agency (1.83) and the broadest self-directed resource use (4.75). Intensive use therefore coincides with markedly different autonomy profiles depending on trust and reliance mode.

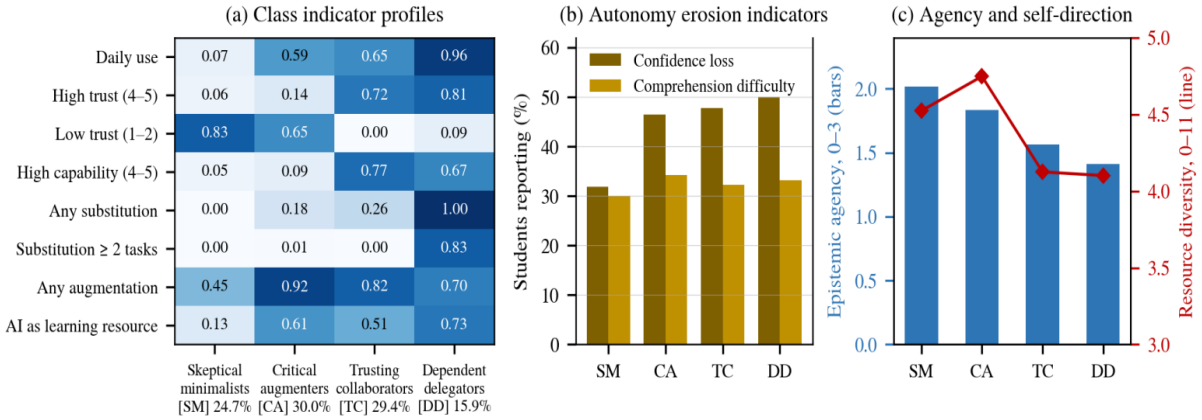


Figure 5. Latent reliance profiles among students: (a) class-conditional indicator proportions, (b) autonomy erosion indicators, and (c) epistemic agency and self-directed resource diversity by class (SM = skeptical minimalists, CA = critical augmenters, TC = tr

V. Discussion

A. Theoretical Implications

The results refine the account of how trust turns into over-reliance in learning contexts. Trust did not raise reliance uniformly; it selectively raised substitution, the mode that removes the learner from the task. This mode specificity offers an explanation for the mixed findings obtained with frequency-based reliance measures, since frequency aggregates augmentative use, which is benign for autonomy, with substitutive use, which is not. The response surfaces add that calibration matters beyond level. Students whose trust outran their own appraisal of AI capability substituted more than students with the mirror-image discrepancy, which translates the calibration principle of Lee and See [5] into a measurable, learner-referenced form. The autonomy results show that erosion is facet-specific. Epistemic agency was the facet most tightly bound to trust and substitution, consistent with evidence that confidence in GenAI reduces critical engagement [10] and that AI support can displace metacognitive processing [12]. Confidence loss followed a different logic, tracking frequency, inexperience, and learner status more closely than delegation depth. Together with the finding that students report twice the confidence loss of professionals with identical reliance profiles, this pattern points to a mastery-experience mechanism [31]: novices who obtain solutions from a tool forgo the successful unaided performances from which self-efficacy grows, whether or not they formally hand the task over. The pattern accords with the widening-gap account of novice programmers [33] and with the moderate-trust advantage reported by Mai et al. [14], whose AI-tutor findings are extended here to generative tools, a multinational sample, and a calibration-based formulation. The latent profiles identify an autonomy-preserving pattern of intensive use. Critical augmenters used AI frequently yet kept the broadest resource use and near-top epistemic agency. Their combination of low trust and high augmentation corresponds to the posture that calibration theory recommends, namely active use with retained verification. Dependent delegators represent the opposite configuration and form a clearly identifiable target group of about one student in six.

B. Implications for Design and Teaching

Three implications follow. First, reliance mode deserves more attention than usage volume. Course policies and learning analytics can distinguish the substitution of core learning tasks from their augmentation, for instance by asking students to explain AI-assisted code or to document their verification steps. Debugging, which students substituted at nearly twice the professional rate, merits protection as a site of productive struggle. Second, interfaces can support calibration by asking learners to state their own expectation before an answer is revealed, a form of cognitive forcing [20], and by making verification inexpensive through explanations that expose reasoning [21]. Third, confidence building needs explicit attention for learners: short unaided practice cycles, a graded withdrawal of AI support, and reflective self-assessment restore the mastery experiences that routine assistance displaces [11], [27].

C. Future Directions

The scale and international reach of the survey make these results a strong basis for designs that follow the same students over time. Panel data will trace how calibration develops across a degree programme, and linking survey measures with interaction logs will allow reliance mode to be observed directly. Extending the reliance-mode and autonomy framework to writing-intensive and laboratory disciplines is a natural next step, as are classroom interventions that target over-trust and substitution.

VI. Conclusion

This paper examined how trust in GenAI, over-reliance, and learner autonomy relate among university students, using large-scale survey data with a professional contrast group. Trust

translated into substitution rather than augmentation, over-trust relative to appraised capability intensified substitution, and substitution carried part of the negative association between trust and epistemic agency. Students reported nearly twice the problem-solving confidence loss of professionals with identical trust and reliance profiles, and four latent profiles showed that intensive GenAI use can coexist with strong autonomy when it takes an augmentative, verification-oriented form. Supporting learner autonomy in the GenAI era therefore depends less on limiting use than on calibrating trust and on keeping students as the authors of their core learning tasks.

REFERENCES

E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individ. Differ.*, vol. 103, Art. no. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," *Int. J. Educ. Technol. High. Educ.*, vol. 20, Art. no. 43, 2023.

J. Prather et al., "The robots are here: Navigating the generative AI revolution in computing education," in *Proc. Working Group Rep. Innov. Technol. Comput. Sci. Educ. (ITiCSE-WGR '23)*, 2023, pp. 108–159, doi: 10.1145/3623762.3633499.

B. A. Becker, P. Denny, J. Finnie-Ansley, A. Luxton-Reilly, J. Prather, and E. A. Santos, "Programming is hard—or at least it used to be: Educational opportunities and challenges of AI code generation," in *Proc. 54th ACM Tech. Symp. Comput. Sci. Educ. (SIGCSE '23)*, 2023, pp. 500–506.

J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.

R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Hum. Factors*, vol. 39, no. 2, pp. 230–253, 1997.

Wang, Y., & Xing, S. (2025). AI-driven CPU resource management in cloud operating systems. Journal of Computer and Communications, 13(06), 135-149.

Xing, S., & Wang, Y. (2025). Cross-modal attention networks for multi-modal anomaly detection in system software. IEEE Open Journal of the Computer Society.

R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," *Hum. Factors*, vol. 52, no. 3, pp. 381–410, 2010.

A. Klingbeil, C. Grützner, and P. Schreck, "Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI," *Comput. Hum. Behav.*, vol. 160, Art. no. 108352, 2024, doi: 10.1016/j.chb.2024.108352.

H.-P. Lee, A. Sarkar, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, and N. Wilson, "The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI '25)*, 2025, Art. no. 1121, doi: 10.1145/3706598.3713778.

H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI without guardrails can harm learning: Evidence from high school mathematics," *Proc. Natl. Acad. Sci. USA*, vol. 122, no. 26, Art. no. e2422633122, 2025, doi: 10.1073/pnas.2422633122.

M. Stadler, M. Bannert, and M. Sailer, "Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry," *Comput. Hum. Behav.*, vol. 160, Art. no. 108386, 2024, doi: 10.1016/j.chb.2024.108386.

N. T. Mai, Q. Fang, and W. Cao, "Measuring student trust and over-reliance on AI tutors: Implications for STEM learning outcomes," *Int. J. Soc. Sci. Engl. Lit.*, vol. 9, no. 12, pp. 11–17, 2025, doi: 10.55220/2576-683x.v9.799.

H. Holec, *Autonomy and Foreign Language Learning*. Oxford, U.K.: Pergamon, 1981.

Stack Overflow, "2025 Stack Overflow Developer Survey," 2025. [Online]. Available: https://survey.stackoverflow.co/2025

K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Hum. Factors*, vol. 57, no. 3, pp. 407–434, 2015.

E. Glikson and A. W. Woolley, "Human trust in artificial intelligence: Review of empirical research," *Acad. Manag. Ann.*, vol. 14, no. 2, pp. 627–660, 2020.

S. Passi and M. Vorvoreanu, "Overreliance on AI: Literature review," Microsoft Corp., Tech. Rep. MSR-TR-2022-12, 2022.

Z. Buçinca, M. B. Malaya, and K. Z. Gajos, "To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, Art. no. 188, 2021, doi: 10.1145/3449287.

H. Vasconcelos, M. Jörke, M. Grunde-McLaughlin, T. Gerstenberg, M. S. Bernstein, and R. Krishna, "Explanations can reduce overreliance on AI systems during decision-making," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW1, Art. no. 129, 2023, doi: 10.1145/3579605.

M. Schemmer, N. Kühl, C. Benz, A. Bartos, and G. Satzger, "Appropriate reliance on AI advice: Conceptualization and the effect of explanations," in *Proc. 28th Int. Conf. Intell. User Interfaces (IUI '23)*, 2023, pp. 410–422, doi: 10.1145/3581641.3584066.

E. F. Risko and S. J. Gilbert, "Cognitive offloading," *Trends Cogn. Sci.*, vol. 20, no. 9, pp. 676–688, 2016.

M. Gerlich, "AI tools in society: Impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, Art. no. 6, 2025, doi: 10.3390/soc15010006.

A. Darvishi, H. Khosravi, S. Sadiq, D. Gašević, and G. Siemens, "Impact of AI assistance on student agency," Comput. Educ., vol. 210, Art. no. 104967, 2024, doi: 10.1016/j.compedu.2023.104967.

D. Little, Learner Autonomy 1: Definitions, Issues and Problems. Dublin, Ireland: Authentik, 1991.

Fang, Q., & Liu, W. (2026). Neuro-symbolic pedagogical alignment (NSPA) for long-horizon classroom discourse analysis: Mitigating dialect bias via counterfactual preference optimization. Computers and Education: Artificial Intelligence, 100664.

Xing, S., Wang, Y., & Liu, W. (2025). Self-adapting CPU scheduling for mixed database workloads via hierarchical deep reinforcement learning. Symmetry, 17(7), 1109.

Ji, E., Wang, Y., Xing, S., & Jin, J. (2025). Hierarchical reinforcement learning for energy-efficient API traffic optimization in large-scale advertising systems. IEEE Access.

Chen, T., Shang, W., & Guo, Z. (2024). Xsla: An SLA-aware heterogeneous XPU scheduling framework for edge AI inference. World Journal of Information Technology, 2(2), 48–59.

Zou, J., Qin, Y., & Rhee, M. (2024). Beyond model scale: A task-aware reliability benchmark for large language models in clinical decision support. Innovation and Technology Studies, 1(1), 39–47.

Fang, Q. (2024). Bias-aware large language models for equitable educational feedback and assessment. Trends in Social Sciences and Humanities Research, 2(1), 41–50.

Qin, Y., Zou, J., & Rhee, M. (2022). An AI-driven intelligent transaction execution framework for internet finance platforms: Cost-sensitive real-time integrity interdiction. Eurasia Journal of Science and Technology, 4(2), 36–44.

Wang, Y., & Zhang, C. (2023). AERIS: Adaptive executor scoped resource management for privacy preserving data processing in Apache Spark. Eurasia Journal of Science and Technology, 5(2), 25-33.

Zhao, W., & Wang, B. (2022). CAPA: A prediction-driven autoscaling framework for SLO-aware cloud-native machine learning inference. Journal of Computer Science and Electrical Engineering, 4(1), 8-16.

Guo, Z., Chen, T., & Shang, W. (2022). Class-conditional intermediate-domain adversarial adaptation for cross-device acoustic scene classification. Journal of Computer Science and Electrical Engineering, 4(1), 26-35

Shen, A. (2023). Compute-efficient scaling of large sequential recommendation models under dynamic user behavior. World Journal of Information Technology, 1(1), 62–71.

Fang, Q. (2024). True-EWS: Trustworthy early warning with equalized conformal guarantees for equitable student support in U.S. education. Educational Research and Human Development, 1(2), 44–54.

Shen, A., & Huang, S. (2022). A systems-aware framework for intelligent large-scale recommendation models. Journal of Computer Science and Electrical Engineering, 4(1), 36–44.

Lin, Y., Wang, Y., & Bian, M. (2024). XOR EW: An explainable and cost aware artificial intelligence framework for operational risk early warning in financial institutions. AI and Data Science Journal, 1(1), 94-102.