What Guides Our Choices? Modeling Developers' Trust and Behavioral Intentions Towards GenAI

Rudrajit Choudhuri*, Bianca Trinkenreich*, Rahul Pandita†, Eirini Kalliamvakou†,
Igor Steinmacher‡, Marco Gerosa‡, Christopher Sanchez*, Anita Sarma*

*Oregon State University, United States, {choudhru, bianca.trinkenreich, christopher.sanchez, anita.sarma}@oregonstate.edu
†GitHub Inc., San Francisco, United States, {rahulpandita, ikaliam}@github.com
‡Northern Arizona University, United States, {igor.steinmacher, marco.gerosa}@nau.edu

Abstract*—Generative AI (genAI) tools, such as ChatGPT or Copilot, are advertised to improve developer productivity and are being integrated into software development. However, misaligned trust, skepticism, and usability concerns can impede the adoption of such tools. Research also indicates that AI can be exclusionary, failing to support diverse users adequately. One such aspect of diversity is cognitive diversity—variations in users' cognitive styles—that leads to divergence in perspectives and interaction styles. When an individual's cognitive style is unsupported, it creates barriers to technology adoption. Therefore, to understand how to effectively integrate genAI tools into software development, it is first important to model *what factors affect developers' trust and intentions to adopt genAI tools in practice?

We developed a theoretically grounded statistical model to (1) identify factors that influence developers' trust in genAI tools and (2) examine the relationship between developers' trust, cognitive styles, and their intentions to use these tools in their work. We surveyed software developers (N=238) at two major global tech organizations: GitHub Inc. and Microsoft; and employed Partial Least Squares-Structural Equation Modeling (PLS-SEM) to evaluate our model. Our findings reveal that genAI's *system/output quality*, *functional value*, and *goal maintenance* significantly influence developers' trust in these tools. Furthermore, developers' *trust* and *cognitive styles* influence their intentions to use these tools in their work. We offer practical suggestions for designing genAI tools for effective use and inclusive user experience.

***Index Terms*—Generative AI, LLM, Software Engineering, Trust, Cognitive Styles, Behavioral Intentions, PLS-SEM**

I. Introduction

Generative AI (genAI) tools (e.g., ChatGPT [1], Copilot [2]) are being increasingly used in software development [3]. These tools promise enhanced productivity [4] and are transforming how developers code and innovate [5]. However, this push for adoption [6] is marked with AI hype and skepticism [7], as well as interaction challenges [3, 8].

Trust has long been recognized as a critical design requirement of AI tools [9, 10, 11]. Miscalibrated levels of trust—over or under trust—can lead developers to overlook errors and risks introduced by AI [12] or deter them from using these tools [13]. Prior research has identified various factors that foster developers' trust in genAI tools [14, 15, 16]. For instance, interaction factors such as setting appropriate expectations and validating AI suggestions [14] along with community factors like shared experiences and community support [15] are relevant in building trust. Recently, Johnson et al. [16] introduced the *PICSE* framework through a qualitative investigation with software developers, outlining key components that influence the formation and evolution of trust in software tools (see Sec. II-A). What is missing, however, is an empirically grounded theoretical understanding of how the multitude of factors associate with developers' trust in genAI tools. Therefore, it becomes important to answer **(RQ1): *What factors predict developers' trust in genAI tools?*** Understanding the significance and strength of these associations is needed to inform the design and adoption of genAI tools in software development.

Another important concern in industry-wide integration of AI tools is that software design can be exclusionary in different ways [17, 18, 19], often failing to support diverse users [20]. While a substantial body of work exists on modeling users' technology acceptance [21, 22, 23, 24], these studies do not consider the inclusivity of the software design. One often overlooked aspect of inclusivity is supporting cognitive diversity—variations in individuals' cognitive styles—which fosters divergence in perspectives and thoughts (see Sec. II-B) [25]. Numerous studies have shown that when technology is misaligned with users' diverse cognitive styles [26, 18, 27], it creates additional barriers for those whose styles are unsupported, forcing them to exert additional cognitive effort [26]. Thus, it is essential to understand how developers' cognitive styles influence their intention to adopt genAI tools, and how trust contributes to this multi-faceted decision. Therefore, we investigate **(RQ2): *How are developers' trust and cognitive styles associated with their intentions to use genAI tools?***

We answer these research questions by establishing a theoretical model, grounded in prior literature, for trust and behavioral intentions toward genAI tools. We evaluated this model using Partial Least Squares-Structural Equation Modeling (PLS-SEM) with survey data from developers (N=238) at two global tech organizations: GitHub Inc. [28] and Microsoft[29].

Our theoretical model (Figure 1) empirically shows that genAI's *system/output quality* (presentation, adherence to safe and secure practices, performance, and output quality in relation to work style/practices), *functional value* (educational value and practical benefits), and *goal maintenance* (alignment between developers' immediate objectives and genAI's actions) are positively associated with developers' trust in these

tools. Furthermore, developers' trust and cognitive styles—intrinsic *motivations* behind using technology, *computer self-efficacy* within peer groups, and *attitudes towards risk*—are associated with their intentions to use these tools, which in turn, correlates with their reported genAI usage in work.

The main contributions of this paper are twofold: (1) an empirically grounded theoretical model for developers' trust and behavioral intentions towards genAI tools, extending our understanding of AI adoption dynamics in software development, and (2) a psychometrically validated instrument for capturing trust-related factors in the context of human-genAI interactions that can be leveraged in future work.

II. Background

A. Trust in AI

Trust in AI is commonly defined as *"the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"* [10, 30, 31, 14, 32]. Trust is subjective and thus a psychological construct that is not directly observable [33] and should be distinguished from observable measures such as reliance [34]. Trust involves users attributing intent and anthropomorphism to the AI [35], leading to feelings of betrayal when trust is violated. Despite AI systems being inanimate, users often anthropomorphize them [35], thereby shifting from reliance to trust in AI systems.

Unobservable psychological constructs are commonly measured through validated self-reported scales (instruments) [36] using questions designed to capture the construct of interest. In this paper, we measure developers' trust in genAI tools using the validated Trust in eXplainable AI (TXAI) instrument [32, 37]. TXAI has been derived from existing trust scales [38, 39, 37] and its psychometric quality has been validated [32]. Researchers frequently advocate using the TXAI instrument for measuring trust in AI [32, 40, 41].

Factors affecting trust: Prior research has extensively examined factors influencing human trust in automation [38, 39, 42, 43]. However, these preliminary insights do not necessarily transfer to human-AI interactions [14] because of the nuances in how users form trust in AI tools, alongside the inherent uncertainty [31] and variability [44] associated with these systems. Additionally, the context in which AI is applied (in our case, software development) influences how trust is developed and its contributing factors [45].

Relevant to our domain, Johnson et al. [16] interviewed software engineers to outline factors that engineers consider when establishing and (re)building trust in tools through the PICSE framework: (1) ***P**ersonal* (internal, external, and social factors), (2) ***I**nteraction* (aspects of engagement with a tool), (3) ***C**ontrol* (over the tool), (4) ***S**ystem* (properties of the tool), and (5) ***E**xpectations* (with the tool). Since PICSE is developed for software engineering (SE), we use it to design our survey instrument to identify factors influencing developers' trust in genAI tools. However, the PICSE framework was qualitatively developed and the psychometric quality—reliability and validity—of a survey based on it has not yet been assessed.

Our work builds upon PICSE to contribute (a) a validated instrument for capturing different factors that developers consider when forming trust in genAI tools (Sec. III-B) through a psychometric analysis of the PICSE framework and (b) assesses the significance and strength of these factors' association with trust in genAI tools (Sec. IV).

B. Users' Cognitive Styles

AI can be exclusionary in different ways often failing to support all users as it should [20, 17, 19]. E.g., Weisz et al. [46] found that some, but not all, participants could produce high-quality code with AI assistance, and the differences were linked to varying participant interactions with AI.

User experience in Human-AI interaction (HAI-UX) can be improved by supporting diverse cognitive styles [47], which refer to the *ways users perceive, process, and interact with information and technology, as well as their approach to problem-solving* [25]. While no particular style is inherently better or worse, if a tool insufficiently supports (or is misaligned with) users' cognitive styles; they pay an additional "cognitive tax" to use it, creating barriers to usability [27].

Here, we scope developers' diverse cognitive styles to the five cognitive styles in the GenderMag inclusive design method [26]. GenderMag's cognitive styles (facets) are users' diverse: *attitudes towards risk*, *computer self-efficacy* within their peer group, *motivations* to use the technology, *information processing style*, and *learning style* for new technology. Each facet represents a spectrum. For example, risk-averse individuals (one endpoint of the 'attitude towards risk' spectrum) hesitate to try new technology or features, whereas risk-tolerant ones (the other end) are inclined to try unproven technology that may require additional cognitive effort or time. GenderMag's cognitive styles are well-suited as they have been (a) repeatedly shown to align with users' interactions with technology both in the context of SE [27, 26, 48] and HAI interactions [47, 49], and (b) distilled from an extensive list of applicable cognitive style types [18, 26], intended for actionable use by practitioners. We used the validated GenderMag facet survey instrument [50] in our study.

C. Behavioral Intention and Usage

Behavioral intention refers to *the extent to which a person has made conscious plans to undertake a specific future activity* [21]. Technology acceptance models, such as TAM [23] and UTAUT [21], identify behavioral intention as a key indicator of actual technology usage [22]. Understanding users' behavioral intentions is useful for predicting technology adoption and guiding future design strategies [21]. While there is an extensive body of work modeling users' behavioral intentions towards software tools [21, 22, 24], these studies primarily focus on socio-technical factors driving adoption.

Our work contributes to this line of research by examining the role of developers' trust and cognitive styles in shaping their intentions to use genAI tools (Sec. III-C and IV), thereby extending the understanding of AI adoption dynamics in SE. We used components of the UTAUT model [21] to capture developers' behavioral intentions and usage of genAI tools.

III. METHOD

To address our RQs, we surveyed software developers from two major global tech organizations, GitHub Inc. and Microsoft. We leveraged existing theoretical frameworks and instruments to design our data collection instrument (see Sec. II). While using existing theoretical frameworks is a first step in developing questionnaires, conducting a psychometric quality assessment is essential to ensure its subsequent reliability and validity [51]. As there was no validated instrument to measure the constructs of the PICSE framework [16]–our chosen trust framework–we performed its psychometric assessment [51] (Sec. III-B). This assessment helped us define a theoretical model of factors developers consider when forming trust in genAI tools, which we then evaluated using Partial Least Squares-Structural Equation Modeling (PLS-SEM) to answer RQ1. To answer RQ2, we assessed the relationships between developers' trust and cognitive styles with their intentions to use genAI tools. Next, we discuss each step.

A. Survey Design and Data Collection

*1) **Survey design**:* We defined the measurement model [52] based on the theoretical frameworks discussed in Sec. II to guide our survey design (Table I). Four researchers with experience in survey studies and GitHub's research team co-designed the survey over a four-month period (Oct 2023 to Jan 2024). We adapted existing (validated) instruments in designing the survey questions (Table I). The questions were contextualized for the target population and pragmatic decisions were made to limit the survey length. The complete questionnaire is available in the supplemental material [53].

TABLE I
MEASUREMENT MODEL CONSTRUCTS AND INSTRUMENTS

Construct	Instrument
Trust	TXAI instrument* [32, 37]
Factors affecting trust	PICSE framework** [16]
Users' cognitive styles	GenderMag facet survey [50]
Behavioral intention & usage	UTAUT model [21, 22]

*We used the 4-item TXAI scale [37] instead of the 6-item scale [32] to reduce participant fatigue. **PICSE does not have a validated questionnaire in [16].

After the IRB-approved informed consent, participants responded to closed questions about their familiarity with genAI technology and their attitudes and intentions towards using genAI tools in work. All closed questions utilized a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree") with a neutral option. These questions also included a 6^{th} option ("I'm not sure") for participants who either preferred not to or did not know how to respond to a question. This differs from being neutral–acknowledging the difference between ignorance and indifference [54].

Demographic questions covered gender, continent of residence, years of software engineering (SE) experience, and primary SE responsibilities at work. We did not collect data on country of residence or specific job roles/work contexts to maintain participant anonymity, as per GitHub and Microsoft's guidelines. An open-ended question for additional comments was included at the end of the survey.

TABLE II
RESPONDENT DEMOGRAPHICS (N=238)

Attribute	N	Percentage
Gender		
Man	186	78.2%
Woman	39	16.4%
Non-binary or gender diverse	6	2.5%
Prefer not to say	7	2.9%
Continent of Residence		
North America	129	54.2%
Europe	55	23.1%
Asia	33	13.8%
Africa	9	3.8%
South America	8	3.4%
Pacific/Oceania	4	1.7%
SE Experience		
1-5 years	57	23.9%
6-10 years	50	21.0%
11-15 years	52	21.9%
Over 15 years	79	33.2%
SE Responsibilities		
Coding/Programming	223	93.7%
Code Review	192	80.6%
System Design	148	62.1%
Documentation	110	46.2%
Maintenance & Updates	108	45.4%
Requirements Gathering & Analysis	108	45.4%
Performance Optimization	107	44.9%
Testing & Quality Assurance	98	41.2%
DevOps/(CI/CD)	90	37.8%
Project Management & Planning	53	22.3%
Security Review & Implementation	46	19.3%
Client/Stakeholder Communication	32	13.5%

The survey took between 7-10 minutes to complete. Attention checks were included to ensure the quality of the survey data. To reduce response bias, we randomized the order of questions within their respective blocks (each construct in Table I). We piloted the questionnaire with collaborators at GitHub to refine its clarity and phrasing.

*2) **Distribution**:* GitHub and Microsoft administered the online questionnaire using their internal survey tools. The survey was distributed to team leads, who were asked to cascade it to their team members. This approach was chosen over using mailing lists to ensure a broader reach [55]. The survey was available for one month (Feb-Mar, 2024), and while participation was optional, it was encouraged.

*3) **Responses**:* We received a total of 343 responses: 235 from Microsoft and 108 from GitHub. We removed patterned responses (n=20), outliers ($<$ 1 year SE experience, n=1), and those that failed attention checks (n=29). Further, we excluded respondents who discontinued the survey without answering all the close-ended questions (n=55). We considered "I'm not sure" responses as missing data points. As in prior work [55], we did not impute data points due to the unproven efficacy of imputation methods within SEM group contexts [56].

After filtration, we retained 238 valid responses (Microsoft: 154, GitHub: 84) from developers across six continents, representing a wide distribution of SE experience. Most respondents were from North America (54.2%) and Europe (23.1%), and most identified as men (78.2%), aligning with distributions reported in previous studies with software engineers [55, 24]. Table II summarizes the respondent demographics.

B. Psychometric Analysis of PICSE Framework

Psychometric quality [57, 58] refers to the objectivity, reliability, and validity of an instrument. We primarily used validated instruments in designing the survey. However, since PICSE was not validated, we conducted a psychometric analysis to empirically refine its factor groupings, which were then evaluated for their association with trust (Sec. IV). Table III presents the factors evaluated in our survey. We performed the analysis using the JASP tool [59], adhering to established psychometric procedures [58, 60, 32] as detailed below:

1) ***Confirmatory Factor Analysis (CFA) – Original grouping***: CFA is a statistical technique that examines intercorrelations between items and proposed factors to test whether a set of observed variables align with a pre-determined factor structure [61]. We assessed whether the PICSE items align with their original five-factor structure (Personal, Interaction, Control, System, and Expectations). The model fit was evaluated using multiple indices: Chi-square test, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI) [62]. Indications of a good model fit include $p > .05$ for χ^2 test, RMSEA $< .06$, SRMR $\leq .08$, and $0.95 \leq$ CFI, TLI ≤ 1 [62]. We employed robust maximum likelihood estimator (MLE),[1] since the data did not meet multivariate normality assumptions, confirmed using Mardia's test [63] (see supplemental [53]). As shown in Table V, results from the original five-factor structure did not indicate a good model fit based on RMSEA, SRMR, CFI, and TLI. This was not entirely unexpected given PICSE's conceptual nature [61]. Therefore, to identify a more appropriate model of factors, we proceeded with an exploratory factor analysis (EFA), uncovering alternative groupings that might better fit the data.

TABLE III
PICSE FRAMEWORK [16]

Category	***Items***
Personal	Community (P1), Source reputation (P2), Clear advantages (P3)
Interaction	Output validation support (I1), Feedback loop (I2), Educational value (I3)
Control	Control over output use (C1), Ease of workflow integration (C2)
System	Ease of use (S1), Polished presentation (S2), Safe and secure practices (S3), Consistent accuracy and appropriateness (S4), Performance (S5)
Expectations	Meeting expectations (E1), Transparent data practices (E2), Style matching (E3), Goal maintenance (E4)

We dropped C3 (tool ownership), as it pertained to AI engineers developing parts of genAI models.

2) ***Exploratory Factor Analysis (EFA)***: Unlike CFA, which relies on an existing a priori expectation of factor structures, EFA identifies the suitable number of latent constructs (factors) and underlying factor structures without imposing a preconceived model [60]. Given the violation of multivariate normality, we used principal axis factoring (considering factors with eigenvalues> 1), as recommended for EFA [60]. We employed oblique rotation, anticipating correlations among the factors. The data met EFA assumptions: significant Bartlett's test for sphericity (χ^2 (136) = 1633.97, $p < .001$) and an adequate Kaiser-Meyer-Olkin (KMO) test (0.892, recommended $\geq .60$) [60]. We used both parallel analysis and a scree plot to determine the number of factors [60], suggesting an alternate five-factor model explaining 64.6% of the total variance. However, most of this variance (60.3%) was accounted for by factors 1, 2, 3, and 5 (22.7, 15.1, 10.6, and 11.9% respectively), while factor 4 explained only 4.3%. The factors showed low correlations (0.2-0.3), except for factor 4, which had high correlations with factors 2 (0.625) and 3 (0.524) (see supplemental). Table IV presents the factor loadings, which indicate the extent to which changes in the underlying factor are reflected in the corresponding indicator (item). All items loaded well onto their factors with primary loadings > 0.5 (items with loadings below 0.4 should be excluded) [64, 52]. For interpreting communality (i.e., the proportion of an item's variance explained by the common factors in factor analyses), values < 0.5 are considered problematic and not interpretable [64]. As shown in Table IV, the communality values for items I1, I2, E1, and E2 were below 0.5, and as they did not load onto any of the factors, they were excluded from the final model. Items in factor 4 (P1, P2, C1) also had low communality values and were likewise dismissed. Based on these results, we concluded that a four-factor solution was the most appropriate, dropping factor 4 due to its low variance explanation, high correlations with other factors, and low communality. The fit indices in Table V indicate a good model fit, showing that the EFA factor structure better fits the data than the original PICSE grouping in the above CFA analysis.

TABLE IV
EFA: FACTOR LOADINGS AND COMMUNALITIES ($h2$)

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	$h2$
S2	0.655					0.525
S3	0.729					0.609
S4	0.823					0.623
S5	0.638					0.657
E3	0.614					0.559
I3		0.941				0.779
P3		0.791				0.599
S1			0.739			0.668
C2			0.671			0.607
P1				0.613		0.418
P2				0.539		0.367
C1				0.517		0.312
E4					0.628	0.685
I1						0.398
I2						0.481
E1						0.405
E2						0.492

The applied oblique rotation method is promax. Communality values (h2)<0.5 are problematic [64] and are marked in RED. Items loaded well (>0.5) onto their primary factors without cross-loadings (>0.3) onto other factors [52]; hence their corresponding cells are kept blank.

3) ***CFA - Alternate grouping***: In the final step, as is best practice [58], we conducted CFA to validate the factor structure identified through EFA. The CFA fit indices in Table V confirm the EFA-derived four-factor model (RMSEA = 0.048; SRMR = 0.047, CFI = 0.982, TLI = 0.973). Table IV outlines the factor structure and corresponding item groupings. Factor 1, labeled ***System/Output quality***, includes items S2 through S5 and E3, which relate to the System group (in

[1]Robust maximum likelihood estimation adjusts for non-normality in data. In general, "robust" in factor analyses refer to methods and values resilient to deviations from ideal distributional assumptions.

PICSE) and the style matching of genAI's outputs. Factor 2, labeled ***Functional value***, encompasses items I3 and P3, reflecting the educational value and practical advantages of using genAI tools. Factor 3, labeled ***Ease of use***, comprises items S1 and C2, addressing the ease of using and integrating genAI in the workflow. Factor 5, labeled ***Goal maintenance***, includes a single item, E4, focusing on genAI's maintenance of human goals. The reliability and validity assessments further support the robustness of these constructs (see Sec. IV-A).

In summary, the psychometric analysis confirmed that a four-factor solution is most appropriate and provided a validated measurement instrument for capturing these factors.

TABLE V
MODEL FIT INDICES - PICSE PSYCHOMETRIC EVALUATION

Model	*RMSEA*	*SRMR*	*CFI*	*TLI*	χ^2	*p-val*
CFA-Original	0.104	0.084	0.925	0.927	147.3	<0.01
EFA	0.057	0.054	0.968	0.965	109.1	<0.01
CFA-Alternate	0.048	0.047	0.982	0.973	59.0	<0.01

χ^2 test results were not considered, as the test is affected by deviations from multivariate normality [65]. We still report the values for completeness.

C. Model Development

As discussed in the previous section, we refined the factor groupings within the PICSE framework. In this study, we are not proposing fundamentally different relationships to trust beyond those identified in the PICSE framework. Instead, we have constrained our focus to only those factors that were psychometrically validated. Next, we detail the hypotheses embedded in our theoretical model for each research question.

RQ1) *Factors associated with trust*

System/Output quality encompasses genAI tools' presentation, adherence to safe and secure practices (including privacy and security implications of using genAI), and its performance and output quality (consistency and correctness) in relation to the development style or work environment in which it is utilized (S2-S5, E3). Developers often place trust in AI based on its performance and output quality (accuracy and consistency), which serve as proxies for the system's perceived credibility [66, 14, 15, 67]. Prior work [14] evidenced that developers are often wary about the security and privacy implications of using AI tools in their work, which influences the level of trust they place in these tools. Drawing upon these insights, we hypothesize: ***(H1)*** *System/Output quality of genAI is positively associated with developers' trust in these tools.*

Functional value of a tool refers to the practical benefits and utility it offers users in their work [68]. In our context, genAI's functional value encompasses its educational value and clear advantages relative to work performance (I3, P3). Prior work highlights that developers' expectations of clear advantages from using AI tools (e.g., increased productivity, improved code quality) contribute to their trust in using these tools [16, 69]. Further, AI's ability to support learning fosters trust in these tools [14]. Based on these, we posit: ***(H2)*** *Functional value of genAI is positively associated with developers' trust in these tools.*

Ease of use associated with genAI tools includes the extent to which developers can easily use and integrate genAI into their current workflow (S1, C2). Prior research highlights that a tool's ease of use [70] and compatibility with existing workflows [10, 24] contribute to users' trust. Following this, we hypothesize: ***(H3)*** *GenAI's ease of use is positively associated with developers' trust in these tools.*

Goal maintenance is related to the degree to which genAI's actions and responses align with the developer's ongoing goals (E4). By its very nature, goals can vary depending on the task and context [16]. Therefore, aligning AI behavior with an individual's immediate goals is crucial in human-AI collaboration scenarios [34]. In terms of human cognition, this congruence is important for maintaining cognitive flow and reducing cognitive load [71], which, in turn, fosters trust in systems [72, 73]. Consequently, we propose: ***(H4)*** *Goal maintenance is positively associated with developers' trust in genAI tools.*

RQ2) *Factors associated with behavioral intentions*

Trust is a key factor in explaining resistance toward automated systems [34] and plays an important role in technology adoption [74, 75]. Multiple studies have correlated an individual's trust in technology with their intention to use it [76, 70, 77]. In our context, we thus posit: ***(H5)*** *Trust is positively associated with intentions to use genAI tools.*

In the context of GenderMag's cognitive styles:

Motivations behind why someone uses technology (technophilic or task-focused) not only influences their intention to use it but also affects how they engage with its features and functionalities [22, 78]. Naturally, individuals motivated by their interest and enjoyment in using and exploring the technology (opposite end of the spectrum from those motivated by task completion) are early adopters of new technology [26]. Based on this, we posit: ***(H6)*** *Motivation to use technology for its own sake is positively associated with intentions to use genAI tools.*

Computer self-efficacy refers to an individual's belief in their ability to engage with and use new technologies to succeed in tasks [79]. It shapes how individuals apply cognitive strategies and the effort and persistence they invest in using new technologies [80], thereby influencing their intention to use them [81, 82]. In line with this, we propose: ***(H7)*** *Computer self-efficacy is positively associated with intentions to use genAI tools.*

Attitude towards risk encompasses an individual's inclination to take risks in uncertain outcomes [83]. This cognitive facet influences decision-making processes, particularly in contexts involving new or unfamiliar technology [81]. Risk-tolerant individuals (one end of the spectrum) are more inclined to experiment with unproven technology than risk-averse ones (the other end) [26], and show higher intentions to use new tools [22, 82]. Thus, we posit: ***(H8)*** *Risk tolerance is positively associated with intentions to use genAI tools.*

Information processing style influences how individuals interact with technology when problem-solving: some gather information *comprehensively* to develop a detailed plan before acting; others gather information *selectively*, acting on initial promising pieces and acquiring more as needed [26]. GenAI

systems, by their very interaction paradigm, inherently support the latter by providing immediate responses to queries, allowing users to act quickly on the information received and gather additional details incrementally. Accordingly, we posit: ***(H9)*** *Selective information processing style is positively associated with intentions to use genAI tools.*

Learning style for technology (by process vs. by tinkering) refers to how an individual approaches problem-solving and how they structure their approach to a new technology [26]. Some prefer to learn through an organized, step-by-step process, while others prefer to tinker around—exploring and experimenting with new technology or its features [26]. Prior work indicates that software, more often than not, is designed to support and encourage tinkering [84], making individuals who prefer this approach more inclined to adopt and use new tools [21]. Thus, we propose: ***(H10)*** *Tinkering style is positively associated with intentions to use genAI tools.*

Behavioral intention. Successful technology adoption hinges on users' intention to use it, translating into future usage. Prior work has consistently shown these factors to be positively correlated [21, 22], suggesting that users who intend to use technology are more likely to do so. Accordingly, we hypothesize: ***(H11)*** *Behavioral intention to use genAI tools is positively associated with the usage of these tools.*

D. Data Analysis

We used Partial Least Squares-Structural Equation Modeling (PLS-SEM) to test our theoretical model. PLS-SEM is a second-generation multivariate data analysis technique that has gained traction in empirical SE studies investigating complex phenomena [24, 55, 85]. It allows for simultaneous analysis of relationships among constructs (measured by one or more indicators) and addresses multiple interconnected research queries in one comprehensive analysis. It is particularly suited for exploratory studies due to its flexibility in handling model complexity while accounting for measurement errors in latent variables [52]. Importantly, PLS-SEM does not require data to meet distributional assumptions. Instead, it uses a bootstrapping approach to determine the statistical significance of path coefficients (i.e., relationships between constructs). The PLS path model is estimated for a large number of random subsamples (usually 5000), generating a bootstrap distribution, which is then used to make statistical inferences [52].

We used the SmartPLS (v4.1.0) software [86] for PLS-SEM analyses, which comprised two main steps, each involving specific tests and procedures. First, we evaluated the measurement model, empirically assessing the relationships between the latent constructs and their indicators (Sec. IV-A). Next, we evaluated the theoretical (or structural) model (Sec. IV-B), representing the hypotheses presented in Section III-C.

The appropriate sample size was determined by conducting power analysis using the G*Power tool [87]. We performed an *F*-test with multiple linear regression, setting a medium effect size (0.25), a significance level of 0.05, and a power of 0.95. The maximum number of predictors in our model is seven (six theoretical constructs and one control variable to Behavioral Intention) (see Fig. 1). The calculation indicated a minimum sample size of 95; our final sample size of 238 exceeded it considerably.

IV. Results

In this section, we report the evaluation of the measurement model (Sec. IV-A), followed by the evaluation of the structural model (Sec. IV-B). We adhered to the evaluation protocols outlined in prior studies [85, 52]. The analysis was performed using the survey data, which met the assumptions for factor analysis [52]: significant Bartlett's test of sphericity on all constructs ($\chi^2(496)=4474.58$, $p < .001$) and adequate KMO measure of sampling adequacy (0.901), well above the recommended threshold (0.60) [60].

A. Measurement Model Evaluation

Our model evaluates several theoretical constructs that are not directly observable (e.g., Trust, Behavioral Intention). These constructs are modeled as latent variables, each measured by a set of indicators or manifest variables (see Fig. 1). The first step in evaluating a structural equation model is to ensure the soundness of the measurement of these latent variables, a process referred to as evaluating the 'measurement model' [52]. We performed a series of tests to validate the measurement model [85], detailed as follows:

TABLE VI
Internal Consistency Reliability and Convergent Validity

	Cronbach's α	*CR*(ρ_a)	*CR*(ρ_c)	*AVE*
System/Output quality	0.816	0.834	0.874	0.781
Functional value	0.816	0.895	0.914	0.842
Ease of use	0.780	0.782	0.902	0.822
Trust	0.856	0.889	0.906	0.710
Motivations	0.713	0.722	0.835	0.718
Risk tolerance	0.715	0.754	0.795	0.667
Computer self-efficacy	0.802	0.809	0.847	0.736
Selective information processing	0.711	0.714	0.849	0.741
Learning by tinkering	0.721	0.722	0.817	0.697
Behavioral intention	0.827	0.831	0.920	0.851

Cronbach's α tends to underestimate reliability, whereas composite reliability (CR: ρ_c) tends to overestimate it. The true reliability typically lies between these two estimates and is effectively captured by CR(ρ_a) [85].

1) ***Convergent validity*** examines how a measure correlates with alternate measures of the same construct, focusing on the correlations between indicators (questions) and their corresponding construct. This evaluation assesses whether respondents interpret the questions as intended by the question designers [88]. Our theoretical model comprises latent constructs that are reflectively measured, meaning the changes in the construct should be reflected in changes in the indicators [85]. Consequently, these indicators should exhibit a significant proportion of shared variance by converging on their respective constructs [52]. We assessed convergent validity using Average Variance Extracted (AVE) and indicator reliability through outer loadings [52].

AVE represents a construct's communality, indicating the shared variance among its indicators, and should exceed 0.5 [52]. AVE values for all latent constructs in our model surpassed this threshold (see Table VI). Regarding outer loadings, values above 0.708 are considered sufficient, while values

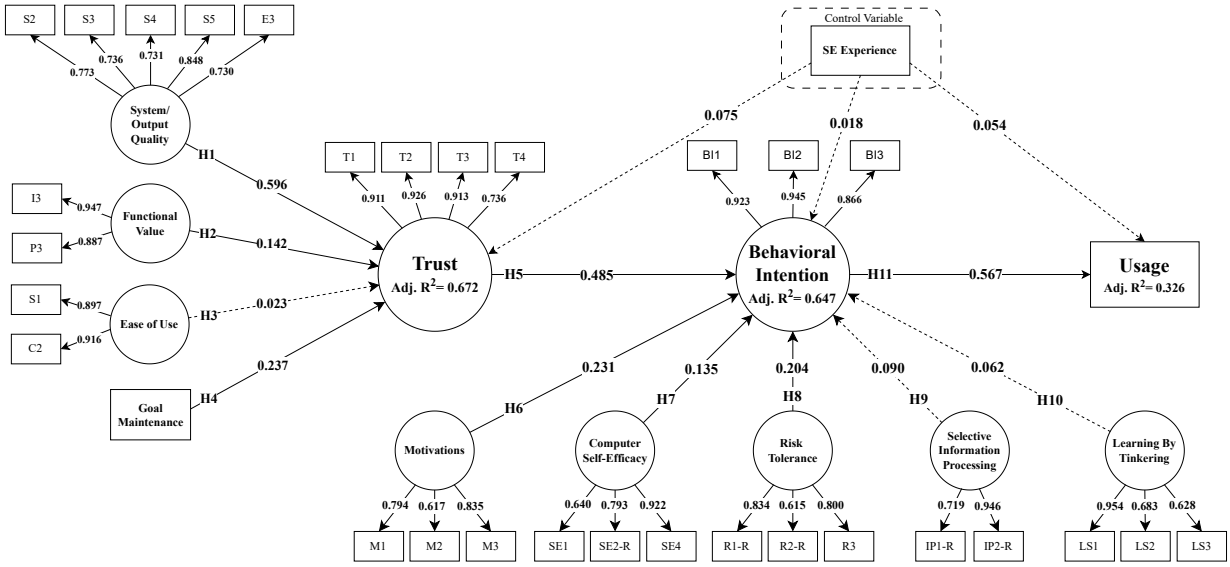

Fig. 1. PLS-SEM Model: Solid lines indicate item loadings and path coefficients ($p < 0.05$); dashed lines represent non-significant paths. Reverse-coded items are suffixed with '-R' (e.g., SE2-R). Latent constructs are depicted as circles and adjusted R^2 (Adj. R^2) values are reported for endogenous constructs.

above 0.60 are sufficient for exploratory studies [52]. We removed variables that did not sufficiently reflect changes in the latent construct (SE3 from computer self-efficacy and IP3 from selective information processing).[2] Thus, all indicators in our model exceeded the threshold, ranging between 0.615 and 0.954 (see Fig. 1).

2) ***Internal consistency reliability*** seeks to confirm that the indicators are consistent with one another and that they consistently and reliably measure the same construct. To assess this, we performed both Cronbach's α and Composite Reliability (CR: ρ_a, ρ_c) tests [85]. The desirable range for these values is between 0.7 and 0.9 [52]. As presented in Table VI, all values corresponding to our model constructs fall within the acceptable range, confirming that the constructs and their indicators meet the reliability criteria.

3) ***Discriminant validity*** assesses the distinctiveness of each construct in relation to the others. Our model includes 10 latent variables (Table VI). A primary method for assessing discriminant validity is the Heterotrait-Monotrait (HTMT) ratio of correlations [89]. Discriminant validity may be considered problematic if the HTMT ratio > 0.9, with a more conservative cut-off at 0.85 [52]. In our case, the HTMT ratios between the latent constructs ranged from 0.064 to 0.791, all below the threshold. We report the HTMT ratios in the supplemental [53], along with the cross-loadings of the indicators, and the Fornell-Larcker criterion values for the sake of completeness.

4) ***Collinearity assessment*** is conducted to evaluate the correlation between predictor variables, ensuring they are independent to avoid potential bias in the model path estimations. We assessed collinearity using the Variance Inflation Factor (VIF). In our model, all VIF values are below 2.1, well below the accepted cut-off value of 5 [52].

[2]After removing SE3 and IP3, the AVE values for computer self-efficacy (now with 3 indicators) and selective information processing (now with 2 indicators) increased from 0.627 to 0.736 and 0.609 to 0.741, respectively.

B. Structural Model Evaluation

After confirming the constructs' reliability and validity, we assess the structural model (graphically represented in Fig. 1). This evaluation involves validating the research hypotheses and assessing the model's predictive power.

1) ***Path coefficients and significance****:* Table VII presents the results of the hypotheses testing, including the mean of the bootstrap distribution (B), the standard deviation (SD), the 95% confidence interval (CI), and the p-values. The path coefficients in Fig. 1 and Table VII are interpreted as standard regression coefficients, indicating the direct effects of one variable on another. Each hypothesis is represented by an arrow between constructs in Fig. 1. For instance, the arrow from "Functional Value" to "Trust" corresponds to H2. Given its positive path coefficient (*B*=0.142), genAI's *functional value* is positively associated with developers' *trust* in these tools. The coefficient of 0.142 indicates that when the score for *functional value* increases by one standard deviation unit, the score for *trust* increases by 0.142 standard deviation units. The analysis results (Table VII) show that most of our hypotheses are supported, except for H3 (p=0.58), H9 (p=0.06), and H10 (p=0.33). Next, we detail the factors associated with trust and behavioral intentions for the supported hypotheses with some exemplary quotes from responses to the open-ended question to illustrate our findings.

Factors associated with trust (RQ1): Our analysis supported Hypotheses H1 (p=0.00), H2 (p=0.03), and H4 (p=0.00) (Table VII). First, the support for *system/output quality* in fostering trust (H1) can be explained by how developers prefer tools that deliver accurate, reliable outputs matching their work style and practices [14, 67]. Next, the *functional value* of genAI, encompassing educational benefits and practical advantages, promotes trust (H2) since developers prioritize tools that offer tangible utility in their work [16, 69]. For instance, one survey respondent mentioned the practical utility of genAI

tools, stating, *"I find value in these models for creative endeavors, gaining different perspectives, or coming up with ideas I wouldn't have otherwise"*. Finally, *goal maintenance* is relevant for cultivating trust (H4). The alignment between a developer's goals and genAI's actions supports using genAI tools to achieve these goals. This eliminates the need for developers to constantly verify the relevance of genAI's outputs, thereby reducing cognitive load. This congruence ultimately enhances genAI's credibility as a cognitive collaborator [34] rather than as an independent and potentially untrustworthy tool, thus bolstering trust in these tools.

Factors associated with behavioral intentions (RQ2): Our analysis supported Hypotheses H5 (p=0.00), H6 (p=0.01), H7 (p=0.01), and H8 (p=0.00), indicating that developers' trust (H5) and their cognitive styles—motivations (H6), computer self-efficacy (H7), and risk tolerance (H8)—are significantly associated with their behavioral intentions to use genAI tools.

Trust (H5) is pivotal in shaping adoption decisions as it reduces resistance to new technologies [74, 75]. When developers trust genAI tools, they perceive them as credible partners, enhancing their willingness to use these tools. Moreover, developers' cognitive styles significantly shape their intentions to adopt genAI tools. Developers motivated by the intrinsic enjoyment of technology (H6) have higher intentions to adopt genAI tools. In contrast, those with a task-oriented approach tend to be more cautious and hesitant about the cognitive effort they are willing to invest in these tools [26]. One respondent echoed this, stating, *"I am slow to adopt new workflows; I put off actively exploring new tools unless it is related to what I need to do"*. Higher computer self-efficacy within peer groups is also significantly associated with increased intentions to use genAI tools (H7). Despite generally high self-efficacy, some developers face interaction challenges with genAI that may impact their confidence and adoption rates. One respondent shared, *"I see my colleagues getting good responses while I fiddle around to get the answers I need. Also, it does not always show the right document relevant to me, so I prefer traditional ways"*. Furthermore, we found that developers with higher risk tolerance are significantly more inclined to use these tools than risk-averse individuals (H8). The context (and involved stakes) in which these tools are used further play a role, as highlighted by another respondent: *"I don't use it yet to write code that I can put my name behind in production; I just use it for side projects or little scripts to speed up my job, but not in actual production code"*.

Finally, our analysis supported Hypothesis H11 (p=0.00), highlighting a significant positive association between developers' behavioral intention to use genAI tools and its usage in their work. This corroborates with prior technology acceptance models [21, 22], emphasizing the pivotal role of behavioral intentions in predicting use behavior.

Control variables: Although experience is often relevant for technology adoption [21], our analysis found no significant associations between SE experience and trust, behavioral intentions, or usage of genAI tools. This is likely since genAI introduces a new interaction paradigm [44], which diverges from traditional SE tools and requires different skills and interactions not necessarily linked to SE experience. Familiarity with genAI, while potentially influential, was excluded as a control variable due to a highly skewed distribution of responses, with most participants reporting high familiarity. Including such skewed variables could lead to unreliable estimates and compromise the model's validity [52, 91]. Similarly, the gender variable was excluded due to its skewed distribution. The analysis of *unobserved heterogeneity* (see supplemental [53]) confirms the absence of any group differences in the model (e.g., organizational heterogeneity) caused by unmeasured criteria.

TABLE VII
STANDARIZED PATH COEFFICIENTS (B), STANDARD DEVIATIONS (SD), CONFIDENCE INTERVALS (CI), P VALUES, AND EFFECT SIZES (f^2)

	B	SD	95% CI	*p*	f^2
H1 System/Output quality→Trust	0.60	0.60	(.45, .77)	**0.000**	0.46
H2 Functional value→Trust	0.14	0.07	(.01, .26)	**0.029**	0.17
H3 Ease of use→Trust	0.02	0.06	(-.08, .16)	0.588	0.01
H4 Goal maintenance →Trust	0.24	0.07	(.07, .36)	**0.002**	0.19
H5 Trust→BI	0.49	0.05	(.38, .58)	**0.000**	0.54
H6 Motivations→BI	0.23	0.08	(.07, .40)	**0.005**	0.26
H7 Computer self-efficacy→BI	0.14	0.05	(.03, .24)	**0.012**	0.14
H8 Risk tolerance→BI	0.20	0.06	(.09, .33)	**0.001**	0.18
H9 Selective information processing →BI	0.09	0.05	(-.02, .14)	0.065	0.03
H10 Learning by tinkering→BI	0.06	0.06	(-.06, .18)	0.331	0.04
H11 BI→Usage	0.57	0.05	(.46, .67)	**0.000**	0.45
SE Experience→Trust	0.08	0.05	(-.01, 0.2)	0.125	0.02
SE Experience→BI	0.02	0.04	(-.03, .11)	0.332	0.00
SE Experience→Usage	0.05	0.05	(-.06, .15)	0.275	0.01

BI: Behavioral Intention. We consider $f^2 < 0.02$ to be no effect, $f^2 \in [0.02, 0.15)$ to be small, $f^2 \in [0.15, 0.35)$ to be medium, and $f^2 > 0.35$ to be large [90].

*2) **Model evaluation**:* We assessed the relationship between constructs and the predictive capabilities of the theoretical model by evaluating the model's explanatory power (R^2, Adjusted (Adj.) R^2), model fit (SRMR), effect sizes (f^2), and predictive relevance (Q^2) [85].

Explanatory power: The coefficient of determination (R^2 and Adj. R^2 values) indicate the proportion of variance in the endogenous variables explained by the predictors. Ranging from 0 to 1, higher R^2 values signify greater explanatory power, with 0.25, 0.5, and 0.75 representing weak, moderate, and substantial levels, respectively [52]. As shown in Table VIII, the R^2 values in our model are 0.68 for Trust, 0.66 for Behavioral intention, and 0.33 for Usage, demonstrating moderate to substantial explanatory power, well above the accepted threshold of 0.19 [92]. Further, Table VII presents the effect sizes (f^2), which measure the impact of each predictor on the endogenous variables. The effect sizes indicate that the predictors exhibit medium to large effects on their respective endogenous variables for all supported hypotheses in our model, with values ranging from 0.14 to 0.54 [90], further

corroborating the model's explanatory power.[3]

Model fit: We analyzed the overall model fit using the standardized root mean square residual (SRMR), a recommended fit measure for detecting misspecification in PLS-SEM models [85]. Our results suggest a good fit of the data in the theoretical model, with SRMR = 0.077, which is below the suggested thresholds of 0.08 (conservative) and 0.10 (lenient) [93].

Predictive relevance: Finally, we evaluated the model's predictive relevance using Stone-Geisser's Q^2 [94], a measure of external validity [52] obtainable via the PLS-predict algorithm [95] in SmartPLS. PLS-predict is a holdout sample-based procedure: it divides the data into k subgroups (folds) of roughly equal size, using $(k-1)$ folds as a training sample to estimate the model, while the remaining fold serves as a holdout to assess out-of-sample predictive power. Q^2_{predict} values are calculated for endogenous variables; values greater than 0 indicate predictive relevance, while negative values suggest the model does not outperform a simple average of the endogenous variable. Our sample was segmented into k=10 parts, and 10 repetitions were used to derive the Q^2_{predict} statistic [52], all of which were greater than 0 (Table VIII), confirming our model's adequacy in terms of predictive relevance.

TABLE VIII
COEFFICIENT OF DETERMINATION AND PREDICTIVE RELEVANCE

Construct	R^2	*Adj.* R^2	$Q^2_{predict}$
Trust	0.679	0.672	0.679
Behavioral Intention	0.658	0.647	0.648
Usage	0.331	0.326	0.234

*3) **Common method bias**:* We collected data via a single survey instrument, which might raise concerns about Common Method Bias/Variance (CMB/CMV) [85]. To test for CMB, we applied Harman's single factor test [96] on the latent variables. No single factor explained more than 23% variance. An unrotated exploratory factor analysis with a forced single-factor solution was conducted, which explained 30.3% of the variance, well below the 50% threshold. Additionally, we used Kock's collinearity approach [97]. The VIFs for the latent variables ranged from 1.01 to 2.45, all under the cut-off of 3.3. These indicate that CMB was not a concern in our study.

V. DISCUSSION

A. Implications for practice

Design to maintain developers' goals. Our findings suggest that developers' trust in genAI tools manifests when these tools align with their goals (H4). This is likely because goal maintenance reduces the need for developers to constantly verify the relevance of genAI's contributions, easing cognitive load and allowing them to focus on task-related activities.

To achieve goal maintenance, the AI should consistently account for the developer's **(1)** *current state*; **(2)** *immediate goals*, and the *expected success* (outcomes from AI) [34]; as well as **(3)** *preferences* for transitioning from their current state (1) to achieving their immediate goals (2). These preferences may involve process methodologies, interaction styles, output specifications, alignment with the work style, or safety/security considerations [16, 26]. Given that an individual's goal(s) may evolve in response to changes in their task state, toolsmiths must also prioritize *adaptability* in genAI tool design. For instance, adaptability can be incorporated by iteratively adjusting AI's actions based on user input on (1), (2), and (3) to ensure its ongoing alignment with their shifting goals.

Allowing developers the flexibility to *explicitly steer AI's actions* as needed is also important for goal maintenance. This control can be essential if the genAI tool deviates from the expected trajectory, enabling developers to (re)calibrate it to support their goals. Further, this can also support developers' metacognitive flexibility [98], i.e., they may adapt their cognitive strategies based on new information or task-state changes. The ability to steer the AI to align with these adapted strategies helps accommodate their evolving goals.

Design for contextual transparency. Developers often face decisions about incorporating genAI tool support for their tasks. Any mismatch between their expectations and genAI's true capabilities can lead to over/under-estimating the tool's functionality, increasing the risk of errors, lost productivity, or potential adverse outcomes in critical tasks [12]. Such outcomes are detrimental to fostering subsequent trust. Thus, helping developers to calibrate their expectations to match the true tool capabilities is essential. Creating this alignment involves designing interfaces that *explicitly communicate the tool's capabilities and limitations*, consistent with HAI interaction guidelines [99, 100]. This clarity allows individuals to form accurate expectations about system quality (H1) within their task contexts and assess its functional value (H2), thereby fostering appropriate trust [35]. To that end, we suggest:
(a) *Communicating genAI's reasoning process and derivation of its outputs in the context of the current task*. Such transparency about the source and how an AI arrived at its output can enable developers to assess correctness more accurately and evaluate any safety or security issues related to using these outputs in their work. This can enhance the assessment of system/output quality, thus cultivating appropriate trust (H1).
(b) *Communicating genAI's limitations for the current task and scoping assistance under conditions of uncertainty to incite warranted trust* [35]. Doing so allows developers to consider the tool's practical utility for their task context, clarifying its appropriate functional value (H2) in their work.

Inclusive tool design for HAI-UX fairness. AI fairness has gained substantial traction over the years [101]. While much of the research and discussion has focused on data or algorithmic fairness, fairness should also include user experiences in human-AI interactions (HAI-UX). We advocate promoting fairness in HAI-UX through inclusive genAI tool design, specifically by supporting developers' diverse cognitive styles.

Our findings indicate that developers who are motivated to use genAI tools for their own sake (H6), have higher computer self-efficacy (H7), and have greater tolerance for risk (H8) are more inclined to adopt these tools compared to their peers.

[3]Large R^2 and f^2 can occasionally indicate overfitting. We thoroughly evaluated this issue by analyzing residuals and conducting cross-validation, finding no evidence of model overfitting (see [53]).

To achieve more equitable acceptance of these tools across the cognitive spectrum of developers, future designs must prioritize *adaptability* based not only on developers' goals (discussed above) but also on their cognitive styles. This can be achieved by capturing these styles (e.g., using the survey questions from this study) and designing genAI tools that dynamically adapt to align with these styles using various strategies. For example, to support developers motivated by task completion, genAI tools can solicit their immediate goals and expected outcomes and deliver contextually appropriate information (and explanations) consistent with their preferred styles. This would simplify developers' ongoing tasks without overwhelming them with unrelated features or extraneous information, helping them complete their tasks effectively. As another example, genAI tools can support individuals with lower self-efficacy by offering explicit cues that differentiate between errors arising from prompting issues and those due to system limitations. For instance, it can caution users about tasks where it typically underperforms, based on prior feedback, or highlight specific parts of the prompt that influenced the generated output. This distinction can help prevent individuals from doubting their ability to effectively use these tools in their work. Further, for risk-averse developers, transparency in genAI behavior and outputs should be emphasized. GenAI should also provide explicit indicators of any potential flaws or uncertainties in its outputs (e.g., confidence scores [14]). Such transparency can help developers make informed decisions about incorporating these outputs into their work (e.g., using AI-generated code), accommodating the caution and deliberation associated with risk aversion.

B. Implications for research

Our study establishes an understanding of developers' trust and intention-related factors during the early stage of genAI adoption. Furthermore, our study offers a validated instrument for capturing relevant factors in the context of human-genAI interaction. Researchers can utilize this instrument to operationalize theoretical expectations or hypotheses—such as capturing the dynamics of trust and intentions in finer contexts, refining genAI tools with design improvements, and comparing user experiences before and after design changes; thus advancing the understanding of AI adoption.

Non-significant associations: Our analysis did not find support for Hypotheses H3 (p=0.59), H9 (p=0.06), and H10 (p=0.33). These findings are surprising, as ease of use, information processing, and tinkering learning style are relevant when considering traditional software tools [26, 21]. However, in genAI contexts, these constructs may manifest differently due to the altered dynamics of user engagement compared to more traditional software. The intuitive nature of genAI interfaces might diminish the traditional impact of these factors. For example, ease of use might not show a relation as using these interfaces is inherently easy; instead, the appropriateness of the queries is what matters. Similarly, developers' information processing style (H9) did not significantly influence their intentions to use genAI tools, likely because how individuals articulate their needs—a single comprehensive prompt or sequence of queries—often aligns with their preferences for consuming information (comprehensive or selective). The lack of a relationship for tinkering style (H10), as well, could be attributed to genAI's interaction paradigm, which is primarily centered around (re)formulating and following up with queries rather than "tinkering" with the software's features. If these speculations hold, how certain validated constructs were framed in the current study [53] might have indeed limited our understanding of these dynamics. Future research should explore these constructs more deeply within the context of human-genAI interactions. For instance, instead of focusing on 'ease of use' or 'tinkering with software features', studies could examine 'ease of prompting' or 'tinkering with prompt strategies' and how preferences (and proficiency) in these areas influence developers' trust and behavioral intentions. Understanding these dynamics can inform the future design and adoption strategies of genAI tools, aligning them more closely with user interaction patterns and cognitive styles.

C. Threats to validity and limitations

Construct validity: We captured constructs through self-reported measures, asking participants to express their agreement with literature-derived indicators. This approach assumes that participants' responses accurately reflect their beliefs and experiences, which might not always be the case due to various biases or misinterpretations. To reduce this threat, we used validated instruments, evaluated the psychometric validity of the PICSE questions, involved practitioners in designing the questions, ran pilot studies, incorporated attention checks, randomized the questionnaire within blocks, and screened the responses. Further, our analysis confirmed that the constructs were internally consistent, reliable, and met convergent and discriminant validity criteria. We did not directly ask participants about their genAI experience; instead, we used their familiarity with genAI tools and its frequency of use at work as proxies to capture their experience with these tools.

Internal validity: Our hypotheses propose associations between constructs, rather than causal relationships, given the cross-sectional nature of the study [102]. We acknowledge the limitation of self-selection bias, as respondents interested in (or skeptical about) genAI tools might be more willing to complete the questionnaire. Further, a theoretical model like ours cannot capture an exhaustive list of factors. Other factors can play a role, thus positioning our results as a reference for future studies. Future work should also consider using longitudinal data and control for potential confounding factors, such as familiarity with genAI and demographic variables. Additionally, trust is a situation-dependent construct [35]. Although we focused on software development, trust in genAI tools may vary based on finer work contexts (e.g., software design vs. software testing tasks). Therefore, our results should be interpreted as a theoretical starting point, guiding future studies to explore these contextual influences.

External validity: Our survey was conducted within GitHub and Microsoft. While the sample includes engineers from

around the globe, it may not fully represent the broader software developer community. However, the sample distribution aligns with previous empirical studies involving software engineers [24, 55], providing a suitable starting point to understand the associations presented in our model. The responses were sufficiently consistent to find full or partial empirical support for the hypotheses. Nevertheless, theory development is an iterative process [103]. Thus, our results should be interpreted as a starting point, aiming for theoretical rather than statistical generalizability. Future studies should replicate, validate, and extend our theoretical model in various contexts.

VI. Conclusion

Our findings highlight that genAI's system/output quality, functional value, and goal maintenance significantly influence developers' trust. Furthermore, trust and cognitive styles (motivations, computer self-efficacy, and attitude towards risk) influence intentions to use these tools, which, in turn, drives usage. We also contribute a validated instrument to capture trust-related factors in human-genAI interaction contexts.

Beyond theoretical contributions, our study offers practical implications, including pointers for design mechanisms that foster appropriate trust and promote equitable user experiences (Sec. V). While our work enhances the understanding of developers' trust and behavioral intentions towards genAI tools, long-term longitudinal studies are essential for refining the knowledge of AI adoption dynamics in software engineering.

Acknowledgments

We thank the GitHub Next team, Tom Zimmermann, and Christian Bird for providing valuable feedback on the survey contents and Brian Houck for facilitating the survey distribution. We also thank all the survey respondents for their time and insights. This work was partially supported by the National Science Foundation under Grant Numbers 2235601, 2236198, 2247929, 2303042, and 2303043. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] OpenAI, "Gpt-4," https://openai.com/product/gpt-4, 2024.
[2] Microsoft, "Copilot," https://copilot.microsoft.com, 2024.
[3] A. Fan, B. Gokkaya, M. Harman, M. Lyubarskiy, S. Sengupta, S. Yoo, and J. M. Zhang, "Large language models for software engineering: Survey and open problems," *arXiv preprint arXiv:2310.03533*, 2023.
[4] E. Kalliamvakou, "A developer's second brain: Reducing complexity through partnership with AI," 2024.
[5] S. Peng, E. Kalliamvakou, P. Cihon, and M. Demirer, "The impact of AI on developer productivity: Evidence from GitHub copilot," *arXiv preprint arXiv:2302.06590*, 2023.
[6] M. N. Center, "Siemens and microsoft partner to drive cross-industry AI adoption," 2023.
[7] McKinsey, "the state of AI in early 2024: genAI adoption spikes and starts to generate value," 2024.
[8] J. T. Liang, C. Yang, and B. A. Myers, "A large-scale survey on the usability of AI programming assistants: Successes and challenges," in *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*, 2024, pp. 1–13.
[9] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human factors*, vol. 57, no. 3, pp. 407–434, 2015.
[10] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human factors*, vol. 46, no. 1, pp. 50–80, 2004.
[11] A. Sellen and E. Horvitz, "The rise of the ai co-pilot: Lessons for design from aviation and beyond," *Communications of the ACM*, 2023.
[12] H. Pearce, B. Ahmad, B. Tan, B. Dolan-Gavitt, and R. Karri, "Asleep at the keyboard? assessing the security of GitHub copilot's code contributions," in *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022, pp. 754–768.
[13] J. G. Boubin, C. F. Rusnock, and J. M. Bindewald, "Quantifying compliance and reliance trust behaviors to influence trust in human-automation teams," *Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 750–754, 2017.
[14] R. Wang, R. Cheng, D. Ford, and T. Zimmermann, "Investigating and designing for trust in AI-powered code generation tools," *arXiv preprint arXiv:2305.11248*, 2023.
[15] R. Cheng, R. Wang, T. Zimmermann, and D. Ford, ""it would work for me too": How online communities shape software developers' trust in AI-powered code generation tools," *ACM Transactions on Interactive Intelligent Systems*, 2023.
[16] B. Johnson, C. Bird, D. Ford, N. Forsgren, and T. Zimmermann, "Make your tools sparkle with trust: The PICSE framework for trust in software tools," in *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 2023, pp. 409–419.
[17] A. Adib-Moghaddam, *Is Artificial Intelligence Racist?: The Ethics of AI and the Future of Humanity*. Bloomsbury Publishing, 2023.
[18] L. Beckwith and M. Burnett, "Gender: An important factor in end-user programming environments?" in *2004 IEEE symposium on visual languages-human centric computing*. IEEE, 2004, pp. 107–114.
[19] "AI mislabeled essays by non-native english speakers as bots," *Business Insider*, Jul 2023.
[20] V. Eubanks, *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.
[21] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS quarterly*, pp. 425–478, 2003.
[22] V. Venkatesh, J. Y. Thong, and X. Xu, "Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology," *MIS quarterly*, pp. 157–178, 2012.
[23] P. Y. Chau, "An empirical assessment of a modified technology acceptance model," *Journal of management information systems*, vol. 13, no. 2, pp. 185–204, 1996.
[24] D. Russo, "Navigating the complexity of generative AI adoption in software engineering," *ACM Transactions on Software Engineering and Methodology*, 2024.
[25] R. J. Sternberg and E. L. Grigorenko, "Are cognitive styles still in style?" *American psychologist*, vol. 52, no. 7, p. 700, 1997.
[26] M. Burnett, S. Stumpf, J. Macbeth, S. Makri, L. Beckwith, I. Kwan, A. Peters, and W. Jernigan, "Gendermag: A method for evaluating software's gender inclusiveness," *Interacting with Computers*, vol. 28, no. 6, pp. 760–787, 2016.
[27] E. Murphy-Hill, A. Elizondo, A. Murillo, M. Harbach, B. Vasilescu, D. Carlson, and F. Dessloch, "Gendermag improves discoverability in the field, especially for women," in *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE)*. IEEE Computer Society, 2024, pp. 2333–2344.
[28] "Github inc." https://github.com/, 2024.
[29] "Microsoft," https://microsoft.com/, 2024.
[30] Q. V. Liao and S. S. Sundar, "Designing for responsible trust in AI systems: A communication perspective," *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1257–1268, 2022.
[31] O. Vereschak, G. Bailly, and B. Caramiaux, "How to evaluate trust in AI-assisted decision making? a survey of empirical methodologies," *Human-Computer Interaction*, vol. 5, no. CSCW2, pp. 1–39, 2021.
[32] S. A. Perrig, N. Scharowski, and F. Brühlmann, "Trust issues with trust scales: examining the psychometric quality of trust measures in the context of AI," in *Extended abstracts of the 2023 CHI Conference on human factors in computing systems*, 2023, pp. 1–7.
[33] K. D. Hopkins, *Educational and psychological measurement and evaluation*. ERIC, 1998.
[34] M. Wischnewski, N. Krämer, and E. Müller, "Measuring and understanding trust calibrations for automated systems: a survey of the state-of-the-art and future directions," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–16.

[35] A. Jacovi, A. Marasović, T. Miller, and Y. Goldberg, "Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI," in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 2021, pp. 624–635.

[36] R. F. DeVellis and C. T. Thorpe, *Scale development: Theory and applications*. Sage publications, 2021.

[37] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, "Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance," *Frontiers in Computer Science*, vol. 5, p. 1096257, 2023.

[38] M. Madsen and S. Gregor, "Measuring human-computer trust," in *11th australasian conference on information systems*, vol. 53. Citeseer, 2000, pp. 6–8.

[39] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *International journal of cognitive ergonomics*, vol. 4, no. 1, pp. 53–71, 2000.

[40] N. Scharowski, S. A. Perrig, L. F. Aeschbach, N. von Felten, K. Opwis, P. Wintersberger, and F. Brühlmann, "To trust or distrust trust measures: Validating questionnaires for trust in AI," *arXiv preprint arXiv:2403.00582*, 2024.

[41] G. Makridis, G. Fatouros, A. Kiourtis, D. Kotios, V. Koukos, D. Kyriazis, and J. Soldatos, "Towards a unified multidimensional explainability metric: Evaluating trustworthiness in ai models," in *2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*. IEEE, 2023, pp. 504–511.

[42] D. H. McKnight, V. Choudhury, and C. Kacmar, "Developing and validating trust measures for e-commerce: An integrative typology," *Information systems research*, vol. 13, no. 3, pp. 334–359, 2002.

[43] S. M. Merritt, "Affective processes in human–automation interactions," *Human Factors*, vol. 53, no. 4, pp. 356–370, 2011.

[44] J. D. Weisz, M. Muller, J. He, and S. Houde, "Toward general design principles for generative AI applications," *arXiv preprint arXiv:2301.05578*, 2023.

[45] N. Omrani, G. Rivieccio, U. Fiore, F. Schiavone, and S. G. Agreda, "To trust or not to trust? an assessment of trust in AI-based systems: Concerns, ethics and contexts," *Technological Forecasting and Social Change*, vol. 181, p. 121763, 2022.

[46] J. D. Weisz, M. Muller, S. I. Ross, F. Martinez, S. Houde, M. Agarwal, K. Talamadupula, and J. T. Richards, "Better together? an evaluation of AI-supported code translation," in *27th International conference on intelligent user interfaces*, 2022, pp. 369–391.

[47] A. Anderson, J. Noa Guevara, F. Moussaoui, T. Li, M. Vorvoreanu, and M. Burnett, "Measuring user experience inclusivity in human-AI interaction via five user problem-solving styles," *ACM Transactions on Interactive Intelligent Systems*, 2022.

[48] M. Guizani, I. Steinmacher, J. Emard, A. Fallatah, M. Burnett, and A. Sarma, "How to debug inclusivity bugs? a debugging process with information architecture," *International Conference on Software Engineering: Software Engineering in Society*, pp. 90–101, 2022.

[49] M. M. Hamid, F. Moussaoui, J. N. Guevara, A. Anderson, and M. Burnett, "Improving user mental models of XAI systems with inclusive design approaches," *arXiv preprint arXiv:2404.13217*, 2024.

[50] M. M. Hamid, A. Chatterjee, M. Guizani, A. Anderson, F. Moussaoui, S. Yang, I. Escobar, A. Sarma, and M. Burnett, "How to measure diversity actionably in technology," *Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights*, 2023.

[51] M. Furr, "Scale construction and psychometrics for social and personality psychology," *Scale Construction and Psychometrics for Social and Personality Psychology*, pp. 1–160, 2011.

[52] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," *Eur. Bus. Rev.*, 2019.

[53] "Supplemental package," https://zenodo.org/record/12798074.

[54] W. L. Grichting, "The meaning of "i don't know" in opinion surveys: Indifference versus ignorance," *Aust Psychol*, vol. 29, no. 1, 1994.

[55] B. Trinkenreich, K.-J. Stol, A. Sarma, D. M. German, M. A. Gerosa, and I. Steinmacher, "Do i belong? modeling sense of virtual community among linux kernel contributors," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2023, pp. 319–331.

[56] M. Sarstedt, C. M. Ringle, and J. F. Hair, "Treating unobserved heterogeneity in PLS-SEM: A multi-method approach," in *Partial least squares path modeling*. Springer, 2017, pp. 197–217.

[57] F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. IAP, 2008.

[58] T. Raykov and G. A. Marcoulides, *Introduction to psychometric theory*. Routledge, 2011.

[59] JASP Team, "Jasp (version 0.16.4) [computer software]," 2024, accessed: 2024-07-07. [Online]. Available: https://www.jasp-stats.org

[60] M. C. Howard, "A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve?" *International journal of human-computer interaction*, vol. 32, no. 1, pp. 51–62, 2016.

[61] D. Harrington, *Confirmatory factor analysis*. Oxford university press, 2009.

[62] L.-t. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural equation modeling: a multidisciplinary journal*, vol. 6, no. 1, pp. 1–55, 1999.

[63] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[64] J. F. Hair, "Multivariate data analysis," 2009.

[65] R. E. Schumacker and R. G. Lomax, *A beginner's guide to structural equation modeling*. psychology press, 2004.

[66] B. J. Fogg and H. Tseng, "The elements of computer credibility," in *Human Factors in Computing Systems*, 1999, pp. 80–87.

[67] K. Yu, S. Berkovsky, R. Taib, J. Zhou, and F. Chen, "Do i trust my machine teammate? an investigation from perception to decision," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 460–468.

[68] J. N. Sheth, B. I. Newman, and B. L. Gross, "Why we buy what we buy: A theory of consumption values," *Journal of business research*, vol. 22, no. 2, pp. 159–170, 1991.

[69] A. Ziegler, E. Kalliamvakou, X. A. Li, A. Rice, D. Rifkin, S. Simister, G. Sittampalam, and E. Aftandilian, "Measuring GitHub copilot's impact on productivity," *Communications of the ACM*, 2024.

[70] D. Gefen, E. Karahanna, and D. W. Straub, "Trust and TAM in online shopping: An integrated model," *MIS quarterly*, pp. 51–90, 2003.

[71] N. Unsworth, T. S. Redick, G. J. Spillers, and G. A. Brewer, "Variation in working memory capacity and cognitive control: Goal maintenance and microadjustments of control," *Quarterly Journal of Experimental Psychology*, vol. 65, no. 2, pp. 326–355, 2012.

[72] L. van der Werff, A. Legood, F. Buckley, A. Weibel, and D. de Cremer, "Trust motivation: The self-regulatory processes underlying trust decisions," *Organizational Psychology Review*, vol. 9, no. 2-3, pp. 99–123, 2019.

[73] W. S. Chow and L. S. Chan, "Social network, social trust and shared goals in organizational knowledge sharing," *Information & management*, vol. 45, no. 7, pp. 458–465, 2008.

[74] S. Xiao, J. Witschey, and E. Murphy-Hill, "Social influences on secure development tool adoption: why security tools spread," in *17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 1095–1106.

[75] J. Witschey, O. Zielinska, A. Welk, E. Murphy-Hill, C. Mayhorn, and T. Zimmermann, "Quantifying developers' adoption of security tools," in *10th Joint Meeting on Foundations of Software Engineering*, 2015, pp. 260–271.

[76] J. W. Kim, H. I. Jo, and B. G. Lee, "The study on the factors influencing on the behavioral intention of chatbot service for the financial sector: Focusing on the UTAUT model," *Journal of Digital Contents Society*, vol. 20, no. 1, pp. 41–50, 2019.

[77] T. H. Baek and M. Kim, "Is chatgpt scary good? how user motivations affect creepiness and trust in generative artificial intelligence," *Telematics and Informatics*, vol. 83, p. 102030, 2023.

[78] H. L. O'Brien, "The influence of hedonic and utilitarian motivations on user engagement: The case of online shopping experiences," *Interacting with computers*, vol. 22, no. 5, pp. 344–352, 2010.

[79] A. Bandura, "Self-efficacy the exercise of control. new york: H," *Freeman & Co. Student Success*, vol. 333, p. 48461, 1997.

[80] D. R. Compeau and C. A. Higgins, "Computer self-efficacy: Development of a measure and initial test," *MIS quarterly*, pp. 189–211, 1995.

[81] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Management science*, vol. 46, no. 2, pp. 186–204, 2000.

[82] X. Li, J. Zhang, and J. Yang, "The effect of computer self-efficacy on the behavioral intention to use translation technologies among college students: Mediating role of learning motivation and cognitive engagement," *Acta Psychologica*, vol. 246, p. 104259, 2024.

[83] J. P. Byrnes, D. C. Miller, and W. D. Schafer, "Gender differences in risk taking: A meta-analysis." *Psychological bulletin*, vol. 125, no. 3, p. 367, 1999.

[84] J. M. Carroll and M. B. Rosson, "Design rationale as theory," *HCI models, theories and frameworks: Toward a multidisciplinary science*, pp. 431–461, 2003.

[85] D. Russo and K.-J. Stol, "PLS-SEM for software engineering research: An introduction and survey," *ACM Comput Surv*, vol. 54, no. 4, 2021.

[86] SmartPLS Team, "Smartpls (version 4.1.0) [computer software]," 2024, accessed: 2024-07-07. [Online]. Available: https://smartpls.com/

[87] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses," *Behav Res Methods*, vol. 41, no. 4, pp. 1149–1160, 2009.

[88] N. Kock, "Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM," *International Journal of e-Collaboration*, vol. 10, no. 1, 2014.

[89] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *J. Acad. Mark. Sci.*, vol. 43, no. 1, pp. 115–135, 2015.

[90] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[91] M. Sarstedt and J.-H. Cheah, "Partial least squares structural equation modeling using smartPLS: a software review," 2019.

[92] W. W. Chin *et al.*, "The partial least squares approach to structural equation modeling," *Modern methods for business research*, vol. 295, no. 2, pp. 295–336, 1998.

[93] J. Henseler, G. Hubona, and P. A. Ray, "Using pls path modeling in new technology research: updated guidelines," *Industrial management & data systems*, vol. 116, no. 1, pp. 2–20, 2016.

[94] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J R Stat Soc Series B Stat Methodol*, vol. 36, 1974.

[95] G. Shmueli, S. Ray, J. M. V. Estrada, and S. B. Chatla, "The elephant in the room: Predictive performance of pls models," *Journal of business Research*, vol. 69, no. 10, pp. 4552–4564, 2016.

[96] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: a critical review of the literature and recommended remedies." *Journal of applied psychology*, vol. 88, no. 5, p. 879, 2003.

[97] N. Kock, "Common method bias in pls-sem: A full collinearity assessment approach," *International Journal of e-Collaboration (ijec)*, vol. 11, no. 4, pp. 1–10, 2015.

[98] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, and S. Rintel, "The metacognitive demands and opportunities of generative AI," in *CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–24.

[99] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen *et al.*, "Guidelines for human-AI interaction," in *2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[100] Google, "People+AI guidebook," https://pair.withgoogle.com/guidebook/, 2023.

[101] B. Green and Y. Chen, "Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments," in *Conference on fairness, accountability, and transparency*, 2019, pp. 90–99.

[102] K.-J. Stol and B. Fitzgerald, "The ABC of software engineering research," *ACM TOSEM*, vol. 27, no. 3, 2018.

[103] F. Shull, J. Singer, and D. I. Sjøberg, "Guide to advanced empirical software engineering," 2007.