Empirical Software Engineering (2026) 31:182
https://doi.org/10.1007/s10664-026-10902-7

From values to adoption: on the role of individual cultural values on fairness toolkit adoption in software development

Stefano Lambiase[1] · Gianmario Voria[2] · Maria Concetta Schiavone[2] · Gemma Catolino[2] · Fabio Palomba[2]

Received: 6 October 2025 / Accepted: 3 June 2026

Abstract

Fairness is a critical concern in the integration of machine learning and AI—e.g., Generative AI, LLMs, and Agents—into decision-making, yet the adoption of fairness toolkits by software practitioners remains limited. This gap hinders efforts to operationalize fairness, especially when cultural and ethical values are overlooked. Given that fairness is socially constructed, individual cultural values may significantly influence how practitioners perceive and adopt fairness tools. The objective of this study is, therefore, to investigate whether and how individual cultural values influence software practitioners' intention to adopt and actual use of fairness toolkits. Specifically, we integrate the UTAUT2 model with Hofstede's cultural dimensions to examine both direct and moderating cultural effects within the adoption process. A survey of 181 software professionals was conducted, and data were analyzed using Partial Least Squares Structural Equation Modeling. Findings show that cultural values—specifically Power Distance, Collectivism, and Long-Term Orientation—not only directly affect adoption intention and behavior but also moderate key relationships in the adoption process. For example, collectivist individuals were more likely to act on their intention to use fairness tools, highlighting the importance of shared team goals. Overall, our findings indicate that fairness toolkit adoption is shaped not only by technology-related perceptions but also by culturally grounded value orientations. These results provide actionable insights for promoting fairness tool adoption through culturally aware strategies in software development environments.

Keywords Software engineering · Socio-technical aspects · Technology adoption · Fairness · Culture · AI · ML

Communicated by: Daniel Méndez.

Extended author information available on the last page of the article

Published online: 08 July 2026

1 Introduction

Fairness, in the context of contemporary software engineering (SE) research, refers to the principle that machine learning (ML) models should make impartial decisions and avoid bias or discrimination against specific social groups (Mehrabi et al. 2021). As ML-enabled systems become increasingly embedded in critical decision-making processes, fairness is emerging as a mandatory quality requirement for any software including AI components. Unfairness typically arises when models reproduce or amplify the biases present in their training data (Pagano et al. 2023; Pessach and Shmueli 2022), leading to decisions that undermine trust and raise significant ethical and legal concerns (Miller 2019). Numerous high-profile incidents have demonstrated the consequences of such biases, including a Facebook vision model labeling Black individuals as "primates," or Amazon assigning lower sales rankings to books with LGBTQIA+ content (Brun and Meliou 2018; Mac 2021; Johnson and Pidd 2009; Wei and Zhou 2022). These examples highlight the urgent need for the development and integration of fairness-aware ML systems.

Recognizing the central importance of fairness, the SE for artificial intelligence research community has made significant advances in developing bias mitigation techniques (Hort et al. 2024), positioning fairness as a key non-functional requirement in the development of ML-enabled software systems. To support the practical integration of these mitigation approaches into development workflows, both researchers and industry organizations have released a variety of **fairness toolkits**, such as AIF360 (Bellamy et al. 2019) and FairLearn (Microsoft, contributors 2019). These toolkits provide ready-to-use metrics and bias mitigation algorithms that assist software professionals in incorporating fairness considerations into their development pipelines (Deng et al. 2022).

Despite the availability of such toolkits, their effective integration into the software development process remains challenging due to the socio-technical nature of SE, where technical decisions are intertwined with human and organizational factors (Hoda 2021; Storey et al. 2020). Recent work has further emphasized that fairness considerations intensify this socio-technical character. For instance, de Souza Santos et al. (2024) conceptualize *software fairness debt* as accumulated bias in SE practices that leads to discriminatory outcomes, explicitly framing fairness as emerging from the interaction between technical artefacts and social dynamics. Later, Sotolani et al. (2026) further explored this concept in the gray literature, underscoring the importance of embedding fairness as a socio-technical aspect of AI development. Similarly, Ferrara et al. (2024b) highlight that implementing fairness practices requires not only technical expertise but also sociological awareness and process-level support across the ML lifecycle. Dolata et al. (2021) likewise argue that fairness is inherently social, as outcomes depend on the mutual shaping of technical systems and organizational structures. These studies position fairness as a socio-technical phenomenon that cannot be addressed through purely technical interventions, but rather with the interplay of social and technical considerations.

In this setting, the adoption of fairness toolkits is not solely a technical choice but a discretionary decision shaped by developers' perceptions of fairness, relevance, and alignment with team priorities. Adopting fairness-enhancing tools therefore depends not only on technical feasibility but also on developers' values and interpretations of fairness (Balayn et al. 2023). Because fairness is inherently value-laden, these interpretations may reflect deeper individual value orientations, including culturally grounded beliefs.

Among the human factors associated with ethics and fairness, **culture** has emerged as a relevant construct. Culture shapes individual beliefs and behavioral orientations through shared value systems (Hofstede 1980; House et al. 2004; Yoo et al. 2011). Prior research shows that cultural dimensions influence how fairness and justice are perceived and prioritized across contexts (Kim and Leung 2007; Lund et al. 2013; Gelfand et al. 2002). This suggests that cultural value orientations may also shape concrete decisions related to fairness-supporting practices, such as the adoption of fairness toolkits.

To date, however, no research has systematically examined whether individual cultural values influence fairness toolkit adoption. A prior study by our research group (Voria et al. 2025a) investigated cognitive and behavioral drivers of adoption using UTAUT2 and found that performance expectancy and habit were the primary predictors. However, that study did not consider whether underlying cultural value orientations shape or condition these adoption mechanisms. The role of cultural values in this process therefore remains underexplored.

This gap presents an opportunity to inform both academic inquiry and managerial practice. Understanding whether individual cultural values influence the adoption of fairness tools could guide organizations in forming development teams strategically—for example, by placing individuals more predisposed to fairness in teams allocated to fairness-related tasks, thereby fostering behavioral change through cultural influence.

To address this gap, we aimed to extend our previous contribution (Voria et al. 2025a) by: (1) conducting a new data collection process to measure cultural characteristics alongside the adoption of the Fairness Tool, and (2) adding a series of new empirical evaluations to deeply explore the role of cultural values in their adoption. Specifically, the present study aims to answer the following research question:

> ◎ **Research Question.** *What role do individual cultural values play in shaping software development practitioners' adoption of fairness toolkits?*

The study began with a review of relevant literature to construct a theoretical model that integrates both technology adoption factors and individual cultural values, with the goal of characterizing software practitioners' intention to adopt and actual adoption of fairness toolkits. For technology-related factors, the widely adopted UTAUT2 framework (Venkatesh et al. 2012) was selected, which had already proven effective in our previous work (Voria et al. 2025a). To account for cultural influences, we employed Hofstede's framework of cultural dimensions (Hofstede 1980, 2011), examining their role in the adoption process both as antecedents (i.e., independent variables) and as moderators (i.e., variables able to strengthening or weakening other relationships). This dual-level analysis was designed to offer a comprehensive and nuanced understanding of the interplay between individual culture and adoption behavior. A survey instrument was then developed using validated measurement scales (Venkatesh et al. 2012; Yoo et al. 2011), and data were analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM) (Hair Junior et al. 2014) to ensure a rigorous and robust examination of the research question.

The results offer valuable insights. As hypothesized, individual cultural values play an active and significant role in the adoption of fairness toolkits, confirming their relevance within the software development context. The primary implication is that organizations and managers seeking to promote fairness awareness within development teams should consider assessing individual cultural values and incorporating this information into team composition and management strategies. Furthermore, from a research perspective, these findings open new avenues for investigation. Cultural factors are increasingly recognized as important within SE (Lambiase et al. 2024a), and communities focused on social and human aspects can build upon this work to expand the body of knowledge—both horizontally, by exploring related phenomena, and vertically, by deepening our understanding of cultural influences across different stages of the software development lifecycle.

Structure of the Paper Section 2 presents the related work. Section 3 reports the theoretical constructs that lead to the definition of hypotheses that are foundational for the tested theory. In Section 4, we present the research method and steps adopted to answer our research question. Section 5 reports the results of the statistical analysis, while Section 6 presents the implications of our study. Section 7 closes our research paper with some conclusions and our future research agenda.

2 Background & Related Work

In the following subsections, we summarize the most relevant literature regarding fairness and fairness toolkits.

2.1 Cultural Values in Software Engineering

Global Software Engineering (GSE) research emerged to investigate challenges associated with distributed software development. Among the factors identified as particularly influential, culture has received significant attention, with researchers examining its impact across the entire software development lifecycle (Borchers 2003; Yasin et al. 2024; Javed et al. 2023).

An early contribution in this area is the work by Borchers (2003), which investigated the influence of cultural factors on SE practices such as code review. The study focused on developers from three countries—Japan, India, and the United States—and operationalized the Hofstede cultural framework (Hofstede 1980). The results showed that cultural backgrounds influence how developers approach SE activities. For example, Japanese developers exhibited lower tolerance for risk, leading to more deliberate decision-making processes. Moreover, Borchers highlighted that cultural differences within software teams can also influence architectural decisions, suggesting the need for further research on the role of culture in SE practices.

Another contribution is the work by Yasin et al. (2024), who conducted an empirical investigation combining experiments and surveys to explore how group activities can mitigate socio-cultural challenges in GSE. Through structured discussion sessions and survey-based evaluation, their study showed that collaborative activities can help identify and address critical challenges related to teamwork in culturally diverse environments, even within simulated distributed development scenarios.

Similarly, Javed et al. (2023) investigated mitigation strategies for socio-cultural distance issues in GSE. Their study combined a literature review with survey-based validation, resulting in a framework comprising twenty-eight mitigation strategies addressing six major socio-cultural challenges in distributed development contexts.

More recently, Lambiase (2024); Lambiase et al. (2022b, 2022c, 2024b, 2024a, 2022a) introduced the concept of *cultural dispersion* in SE. This concept describes situations in which members of a development community exhibit behaviors that diverge from those typically associated with their cultural background according to established cultural frameworks (e.g., Hofstede or GLOBE). Their studies in open-source communities show that unmanaged cultural differences can correlate with socio-technical problems—such as community smells (Lambiase et al. 2022b), as well as with reduced productivity (Lambiase et al. 2022a, 2024b). To address these challenges, the authors proposed an empirically derived framework, *Dealing with Cultural Dispersion*, which helps software teams better understand cultural influences and adopt strategies for effective collaboration. This framework aims to bridge the gap between research and practice by providing actionable guidance for managing cultural diversity in software development teams (Lambiase et al. 2024a).

2.2 Machine Learning Fairness

Fairness refers to the absence of bias or favoritism based on inherent or acquired characteristics (Pessach and Shmueli 2022; Mehrabi et al. 2021; Starke et al. 2021). Various metrics and strategies evaluate fairness in ML, focusing on data similarities, decision probabilities, and cause-effect relationships (Verma and Rubin 2018). Majumder et al. (2023) categorized fairness metrics into seven groups, although not all nuances are captured.

From an intervention perspective, bias in ML systems can be mitigated through three main categories of techniques, depending on the stage of the development pipeline in which they operate. *Pre-processing* techniques modify the training data before model learning in order to reduce discriminatory patterns, with Sharma et al. (2020) and Calmon et al. (2017) using probabilistic methods, and Chakraborty et al. (2021) introducing Fair-SMOTE. *In-processing* techniques intervene during model training, typically by adapting the learning objective or incorporating fairness constraints directly into the optimization process; Zhang et al. (2018) used adversarial methods, and Kamishima et al. (2012) applied regularization. Reweighting methods, like those from Kamiran and Calders (2012) and Chakraborty et al. (2020), adjust instance weights. *Post-processing* techniques instead adjust the model outputs after training to correct for unfair outcomes, with Galhotra et al. (2017) introducing Themis and Udeshi et al. (2018) developing Aequitas.

2.3 Software Engineering for ML Fairness

Fairness in ML has gained traction in the SE community, with various studies addressing it from multiple perspectives (Pessach and Shmueli 2022; Mehrabi et al. 2021; Starke et al. 2021). Ferrara et al. (2024a) stressed the importance of context-aware fairness requirements. Additionally, Ferrara et al. (2024b) advocated for fairness integration across the development lifecycle. Discrimination often stems from biased training datasets (Vasudevan and Kenthapadi 2020; Voria et al. 2025b). Zhang and Harman (2021) argued that increasing dataset features does not inherently reduce discrimination, while Chakraborty et al. (2021) emphasized the role of feature selection. Sesari et al. (2024) highlighted the need to examine fairness across the entire lifecycle, and Voria et al. (2025c, 2025d); Parziale et al. (2025) cataloged and evaluated fairness-aware practices, surveying domain experts. Hort et al. performed a comprehensive survey of methods to address bias in ML classifiers, collecting the most used datasets, metrics, and solutions from the SE literature to achieve fairer AI (Hort et al. 2024). Finally, recent studies in the SE field have benchmarked commonly used bias mitigation methods, understanding their efficiency over multiple protected attributes (Chen et al. 2024) and the trade-offs with other critical non-functional requirements, such as accuracy and energy consumption (De Martino et al. 2025).

2.4 Fairness Toolkits in Practice

Various open-source fairness toolkits, i.e., frameworks or libraries that comprise ready-to-use bias measures or mitigation methods, help researchers and developers create fairer ML models (Balayn et al. 2023). AIF360 (Bellamy et al. 2019), developed by IBM, offers a comprehensive suite of fairness metrics and bias mitigation techniques. Fairlearn (Microsoft, contributors 2019) is a Python-based library focused on fairness assessment and mitigation using in-processing techniques. Google's What-If Tool (Wexler et al. 2019) emphasizes fairness and explainability through interactive analysis, while Scikit-fairness (Warmerdam 2020) extends scikit-learn with bias analysis tools.

Lee and Singh (2021) examined the misalignment between current open-source fairness toolkits and practitioners' needs through focus groups, interviews, and surveys, identifying gaps that necessitate improved support for implementing fairness. Similarly, Holstein et al. (2019) documented challenges in developing fair ML systems within commercial teams based on interviews and surveys. Deng et al. (2022) conducted an empirical investigation into industry practitioners' engagement with fairness toolkits, identifying usability and effectiveness improvements through think-aloud interviews and surveys. Cannavale et al. (2025) explored the use of fairness toolkits in open source communities, underscoring a trend of limited adoption.

Beyond technical limitations, broader structural issues impact fairness tool adoption. Rakova et al. (2021) explored fairness challenges from an organizational perspective, analyzing how internal culture and structure shape responsible AI efforts. Relatedly, recent work (Ayling and Chapman 2022) has observed a shift in AI ethics assessment focus from data concerns to models and algorithms, indicating evolving priorities in ethical AI governance. However, most impact assessment tools remain internal-facing, with external oversight being rare (Ayling and Chapman 2022). While fairness toolkits primarily support developers and product teams, other stakeholders, such as end-users, are often excluded from assessment processes (Ayling and Chapman 2022), mirroring the practitioner-toolkit misalignment identified by Lee and Singh (2021) and Deng et al. (2022).

Additionally, while prior work has focused on fairness in AI development, there is growing recognition of ethical concerns in procurement. A subset of impact assessment tools now addresses AI system acquisition, emphasizing the role of organizations deploying third-party AI solutions (Ayling and Chapman 2022). This broadening scope aligns with calls for a more comprehensive approach to fairness that extends beyond developers to decision-makers and end-users (Rakova et al. 2021).

Research Gap.

In our previous work (Voria et al. 2025a), we investigated the factors influencing engineers' decisions to adopt fairness toolkits. The findings highlighted performance expectancy and habit as the primary drivers of adoption. These results suggested that emphasizing the effectiveness of fairness toolkits in mitigating bias, together with fostering their habitual use, can promote wider adoption within organizations. Building on these insights, the present study extends our prior work by incorporating a new set of variables that capture individuals' cultural values. Since culture shapes everyday choices and is closely tied to perceptions of fairness, it offers a compelling lens for further exploration of this phenomenon. Specifically, this study: (1) extends the analysis by adding six independent variables representing developers' cultural backgrounds, (2) relies on a new round of data collection designed for this purpose, and (3) applies three additional statistical tests (detailed in the following section) to provide a more comprehensive understanding of culture's role in fairness toolkit adoption.

2.5 Partial Least Squares Structural Equation Modeling (PLS-SEM)

Structural Equation Modeling (SEM) is a powerful multivariate statistical technique well-suited for examining complex models involving multiple interrelated variables. A key strength of SEM lies in its ability to handle both *non-latent* variables—those that are directly observable and measurable—and *latent* variables, which represent underlying constructs that cannot be observed directly but are instead estimated through a set of measurable indicators. This makes SEM particularly valuable in situations where classical linear regression is insufficient, such as when variables are connected through a network of paths (also referred to as *hypotheses*) and the model involves multiple dependent and independent variables that must be examined simultaneously (Hair Junior et al. 2014).

Central to SEM is the concept of latent variable measurement. Since latent variables cannot be directly observed, they are operationalized through observable indicators, commonly referred to as *indexes*. For example, in marketing research, "Customer Satisfaction" is a latent construct that, while not directly measurable, can be approximated using indicators such as Overall Satisfaction Rating, Recommendation Likelihood, or Service Quality Perception. In practice, researchers typically rely on questionnaires with validated items and scales to serve as these indexes, ensuring reliable and accurate representation of the underlying construct (Hair Junior et al. 2014).

Latent variables can be measured through two types of approaches: *reflective* and *formative*. In reflective measurement, the latent variable is assumed to cause its indicators, meaning each indicator is a manifestation of the underlying construct. Consequently, all indicators are expected to co-vary, and a change in the latent variable should be reflected across all of them. Because of this, reflective indicators are considered interchangeable. In contrast, formative measurement assumes that the indicators collectively define the latent variable, with each indicator capturing a distinct facet of the construct. As a result, formative indicators need not correlate with one another and are not interchangeable. In practice, reflective measurement is more commonly adopted due to its interchangeability property, which simplifies both model specification and validation (Hair Junior et al. 2014).

Among the various SEM variants, *Partial Least Squares Structural Equation Modeling (PLS-SEM)* is particularly well-regarded for its suitability in predictive and exploratory research. PLS-SEM is especially prevalent in fields such as marketing, information systems, and the social sciences, where theory is still developing and strict distributional assumptions may be difficult to meet. Unlike covariance-based SEM approaches, PLS-SEM imposes minimal requirements on data distribution and sample size, offering considerable flexibility in research design. The method operates by maximizing the explained variance in the dependent variables, thereby optimizing the model's predictive accuracy. Through its iterative estimation algorithm, PLS-SEM produces latent variable scores that are subsequently used to estimate path coefficients and evaluate both the measurement and structural components of the model (Hair Junior et al. 2014).

PLS-SEM analysis is organized around two core components: the *measurement model* and the *structural model*. The measurement model (also referred to as the outer model) defines the relationships between each latent variable and its corresponding observed indicators, assessing how well those indicators represent the underlying construct. The structural model (or inner model), on the other hand, specifies the hypothesized directional relationships among the latent variables themselves, forming the theoretical backbone of the analysis (Hair Junior et al. 2014).

Beyond the evaluation of the measurement and structural models, PLS-SEM also supports the investigation of more nuanced relationships among constructs through *mediation* and *moderation* analysis. Mediation analysis examines whether the relationship between two constructs is transmitted—either partially or fully—through a third, intermediary variable, known as the *mediator*. In this case, the independent variable influences the mediator, which in turn affects the dependent variable, thereby providing a more fine-grained account of the underlying mechanism driving the relationship. Moderation analysis, on the other hand, investigates whether the strength or direction of a relationship between two constructs changes depending on the values of a third variable, referred to as the *moderator*. The moderating variable interacts with the independent variable to influence the outcome, revealing boundary conditions under which a given relationship holds or breaks down. Both analyses enrich the explanatory power of the model, allowing researchers to move beyond simple direct effects and uncover the conditions and mechanisms that shape the relationships among constructs (Hair Junior et al. 2014).

Conducting a PLS-SEM study involves a sequence of well-defined steps, which can be organized into three macro-steps:

<u>Hypothesis Development.</u> The process begins with the formulation of theoretically grounded hypotheses that define the expected relationships among constructs. While this step is not formally part of the SEM methodology itself, it constitutes an essential prerequisite: the hypotheses, derived from an extensive review of the existing literature, determine the structure of the model and guide all subsequent analytical decisions.[1]

<u>Data Collection.</u> Once the hypotheses are established, empirical data must be gathered. This is typically accomplished through a questionnaire-based approach, in which validated instruments from the literature are administered to a sample of participants. The minimum required sample size is determined by means of an a-priori power analysis using G*Power (Faul et al. 2009).[2]

[1] In the context of our study, such hypotheses are reported in Section 3.

[2] In the context of our study, data collection is described in detail in Section 4.

Data Analysis. The analytical phase proceeds in two sequential stages. The first is the evaluation of the *measurement model*, which assesses the reliability and validity of the indicators used to operationalize each latent construct. The second is the evaluation of the *structural model*, which examines the strength, direction, and significance of the hypothesized paths between constructs. Together, these two stages allow researchers to draw conclusions about both the quality of their measures and the plausibility of their theoretical propositions.[3]

PLS-SEM has gained increasing traction in SE research as a method for examining complex relationships involving latent constructs. It is now regarded as a state-of-the-art technique for both theory development and predictive modeling in the field, although consensus on best practices and reporting standards is still evolving (Russo and Stol 2021). Its ability to accommodate complex models, relatively small sample sizes, and non-normal data distributions makes it particularly well-suited for empirical SE studies, enabling researchers to move beyond traditional productivity metrics—such as lines of code—and incorporate organizational and behavioral dimensions into their analyses. Recent applications span a variety of SE contexts, including educational settings (Che et al. 2021), software management and DevOps (Khattak et al. 2023), and the adoption of software technologies (Russo 2024; Lambiase et al. 2024a; Choudhuri et al. 2024; Lambiase et al. 2025; Lambiase and De Lucia 2026).

3 Objective, Research Model, and Hypothesis

This study investigated the role of software practitioners' individual cultural values in shaping both the intention to adopt and the actual adoption of fairness toolkits. To provide a comprehensive understanding of the adoption phenomenon, the Extended Unified Theory of Acceptance and Use of Technology (UTAUT2) (Venkatesh et al. 2012) was employed, while Hofstede's cultural dimensions (Hofstede 1980, 2011) were used to operationalize individual-level cultural values.

Specifically, constructs—i.e., variables that cannot be directly observed and should be measured by other related indicators—from the two theoretical frameworks were integrated to build a structural model in which relationships between constructs were formalized through hypotheses. To investigate the role of culture in technology adoption, individual-espoused cultural values were incorporated into the model as both antecedents and moderators of the UTAUT2 constructs. An **antecedent** is a construct that exerts a direct influence on another construct, typically positioned earlier in the causal chain, thus contributing to the explanation of the variance of the endogenous (dependent) variable. In contrast, a **moderator** is a construct that affects the strength or direction of the relationship between two other constructs, weakening or strengthening it. By modeling cultural values in both roles, the analysis accounts for their potential to shape practitioners' perceptions directly (as antecedents) and to alter the impact of technology acceptance factors on intention and usage (as moderators).

The remainder of this section elaborates on each hypothesis. Section 3.1 first introduces the technology-acceptance hypotheses (H1–H7) derived from UTAUT2, summarised in Table 1. Section 3.2 then introduces the hypotheses involving cultural values as antecedents and moderators (H8–H11), summarised in Table 2.

[3] In the context of our study, the results of the statistical analysis are reported in Section 5.

Table 1 Summary of the UTAUT2 technology-acceptance hypotheses (H1–H7) tested in the research model

ID	Relation	Description
H1	PE → BI	Performance Expectancy positively influences Behavioral Intention
H2	EE → BI	Effort Expectancy positively influences Behavioral Intention
H3	SI → BI	Social Influence positively influences Behavioral Intention
H4	HM → BI	Hedonic Motivation positively influences Behavioral Intention
H5a	FC → BI	Facilitating Conditions positively influence Behavioral Intention
H5b	FC → UB	Facilitating Conditions positively influence Use Behavior
H6a	HB → BI	Habit positively influences Behavioral Intention
H6b	HB → UB	Habit positively influences Use Behavior
H7	BI → UB	Behavioral Intention positively influences Use Behavior

Table 2 Summary of the cultural-values hypotheses (H8–H11) tested in the research model. The upper block lists the hypotheses in which cultural values act as antecedents of UTAUT2 constructs; the lower block lists those in which cultural values act as moderators of UTAUT2 relationships

ID	Relation	Description
H8a	PD → SI	Power Distance positively influences perceived social influence
H8b	PD → BI	Power Distance positively influences intention to adopt fairness toolkits
H8c	PD → UB	Power Distance positively influences actual use of fairness toolkits
H9a	UA → BI	Uncertainty Avoidance negatively influences intention to adopt fairness toolkits
H9b	UA → UB	Uncertainty Avoidance negatively influences actual use of fairness toolkits
H10a	LT → PE	Long-Term Orientation positively influences performance expectancy
H10b	LT → HB	Long-Term Orientation positively influences Habit
H10c	LT → BI	Long-Term Orientation positively influences intention to adopt fairness toolkits
H11a	CO → EE	Collectivism positively influences effort expectancy
H11b	CO → SI	Collectivism positively influences perceived social influence
H11c	CO → FC	Collectivism positively influences Facilitating Conditions
H8d	PD × SI → BI	Power Distance positively moderates the effect of Social Influence on Behavioral Intention
H8e	PD × BI → UB	Power Distance positively moderates the effect of Behavioral Intention on Use Behavior
H9c	UA × SI → BI	Uncertainty Avoidance positively moderates the effect of Social Influence on Behavioral Intention
H9d	UA × BI → UB	Uncertainty Avoidance negatively moderates the effect of Behavioral Intention on Use Behavior
H10d	LT × PE → BI	Long-Term positively moderates the effect of Performance Expectancy on Behavioral Intention
H10e	LT × BI → UB	Long-Term positively moderates the effect of Behavioral Intention on Use Behavior
H11d	CO × PE → BI	Collectivism positively moderates the effect of Performance Expectancy on Behavioral Intention
H11e	CO × EE → BI	Collectivism positively moderates the effect of Effort Expectancy on Behavioral Intention
H11f	CO × SI → BI	Collectivism positively moderates the effect of Social Influence on Behavioral Intention
H11g	CO × BI → UB	Collectivism positively moderates the effect of Behavioral Intention on Use Behavior

3.1 Adoption of Fairness Toolkit

The adoption of fairness toolkits was modeled using the constructs of the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) (Venkatesh et al. 2012), a well-established framework to explain individual technology acceptance behaviors.

3.1.1 The Extended Unified Theory of Acceptance and Use of Technology (UTAUT2)

Originally introduced by Venkatesh et al. (2003), the UTAUT framework consolidated eight previous theories of technology adoption and identified four core constructs as predictors of *Behavioral Intention* (BI) and *Usage Behavior* (UB): (1) *Performance Expectancy* (PE), the degree to which individuals believe that using a technology will enhance their job performance; (2) *Effort Expectancy* (EE), the perceived ease associated with technology use; (3) *Social Influence* (SI), the extent to which individuals perceive that important others believe they should use the technology; and (4) *Facilitating Conditions* (FC), the perceived availability of technical and organizational support for using the system.

To better capture adoption in consumer-oriented and voluntary contexts, UTAUT was later extended to UTAUT2 (Venkatesh et al. 2012). The extended model introduced three additional constructs: *Hedonic Motivation* (HM), reflecting the enjoyment or pleasure derived from technology use; *Price Value* (PV), which evaluates the trade-off between the perceived benefits and monetary costs; and *Habit* (HB), representing the extent to which Usage Behavior becomes automatic due to learning.

The decision to adopt UTAUT2 over alternative adoption models was made for several reasons. UTAUT2 has been employed in several empirical studies in SE (Russo 2024; Lambiase et al. 2024a, 2025; Lambiase and De Lucia 2026), including prior work conducted by the authors (Voria et al. 2025a), where it exhibited robust measurement properties. Compared to other prominent models such as Rogers' Diffusion of Innovation Theory (Rogers 2003) or the Technology Acceptance Model (TAM) by Davis et al. (1989), UTAUT2 provides a more comprehensive perspective on technology use, as it emphasizes actual usage and behavioral routines rather than focusing solely on attitudes toward innovation or prior experience with related tools. For this reason, given that the model had shown excellent metrics in the previous analysis (Voria et al. 2025a), and given that the focus of this work is not on technology-related factors but cultural ones (and it was therefore necessary to have a solid basis for the former in order to better analyze the latter), we decided to adopt UTAUT2.[4]

3.1.2 Hypothesis from the UTAUT2

In this study, eight UTAUT2 constructs were considered relevant to the context of fairness toolkit adoption: PE, EE, SI, HM, FC, HB, BI, and UB. The construct PV was excluded, as fairness toolkits are generally freely available and not associated with direct financial costs.

[4] Although explicitly clarified in the text, readers might still be misled into assuming that the authors claim UTAUT2 to be superior to other models. This is not the case. Each model emphasizes different aspects—for example, DoI focuses more on comparisons with existing technologies performing similar tasks—and is thus more suitable for specific stages of technology development. In this work, we chose to rely on UTAUT2, but this does not preclude the analysis of fairness toolkits through other models. On the contrary, employing multiple perspectives is recommended to achieve a more comprehensive understanding of the technology under study.

These constructs have been widely demonstrated to positively influence technology adoption by capturing multiple facets of user perception (Venkatesh et al. 2012; Russo 2024; Lambiase et al. 2024a). Behavioral Intention serves as a mediating factor between these perceptions and actual use (Venkatesh et al. 2012), while Habit and Facilitating Conditions are also known to exert a direct influence on Usage Behavior, independent of intention (Venkatesh et al. 2012; Lambiase et al. 2024a). Thus, we formulated the following hypotheses.

- H1—Performance Expectancy (PE) positively influences software practitioners' intention to adopt (BI) fairness toolkits.
- H2—Effort Expectancy (EE) positively influences software practitioners' intention to adopt (BI) fairness toolkits.
- H3—Social Influence (SI) positively influences software practitioners' intention to adopt (BI) fairness toolkits.
- H4—Hedonic Motivation (HM) positively influences software practitioners' intention to adopt (BI) fairness toolkits.
- H5a—Facilitating Conditions (FC) positively influence software practitioners' intention to adopt (BI) fairness toolkits.
- H5b—Facilitating Conditions (FC) positively influence the actual use behavior (UB) of fairness toolkits.
- H6a—Habit (HB) positively influences software practitioners' intention to adopt (BI) fairness toolkits.
- H6b—Habit (HB) positively influences the actual use behavior (UB) of fairness toolkits.
- H7—Behavioral Intention (BI) to use fairness toolkits positively influences their actual use behavior (UB).

3.2 Influence of Individual Cultural Values

Culture has been defined by Geert Hofstede as "the programming of the human mind by which one group of people distinguishes itself from another group" (Hofstede 1980). An individual's cultural background plays a significant role in shaping a wide range of everyday decisions and behaviors. However, the concept of culture remains complex and challenging to operationalize, often presenting substantial difficulties for researchers aiming to empirically study it. To address this challenge, scholars in the field of cross-cultural research have conducted empirical studies to identify a set of underlying values—referred to as cultural dimensions—that can systematically represent and differentiate cultural orientations across individuals and groups. These cultural dimension frameworks provide structured constructs that capture shared beliefs, norms, and behaviors, allowing for the comparative analysis of cultural differences (Hofstede 1980, 2011).

3.2.1 The Hofstede Cultural Framework

In this study, we employed Hofstede's cultural dimensions to conceptualize individual-level cultural values. The framework decomposes cultural into six dimensions.

Power Distance (PD). Captures the extent to which individuals accept and expect unequal power distribution within a society. High Power Distance is associated with strong deference to authority and adherence to hierarchical norms.

Individualism vs. Collectivism (CO). Reflects the degree to which individuals are integrated into social groups. High individualism indicates a preference for autonomy and self-reliance, while Collectivism emphasizes group cohesion and loyalty in exchange for protection.

Motivation towards Achievement (MA). Contrasts values such as achievement, assertiveness, and material success (high MA) with values such as cooperation, modesty, and quality of life (low MA).[5]

Uncertainty Avoidance (UA). Measures the extent to which individuals feel threatened by uncertainty and ambiguity. High UA is associated with risk aversion, a preference for structured environments, and resistance to novel or unconventional ideas.

Long-Term vs. Short-Term Orientation (LT). Assesses the value placed on long-term planning and perseverance versus short-term traditions and normative commitments. High LT indicates a forward-looking, pragmatic approach, while low LT emphasizes tradition and immediate outcomes.

Indulgence vs. Restraint (IR). Reflects the degree to which societies allow relatively free gratification of basic human desires. Indulgent cultures promote enjoyment and self-expression, while restrained cultures emphasize the regulation of desires through social norms.[6]

Although the original framework comprises six dimensions (Hofstede 1980, 2011), only four were included in this investigation. The dimension Indulgence vs. Restraint was excluded due to the absence of validated instruments for measuring it at the individual level—as existing scales are primarily designed for national-level analysis—and the lack of empirical evidence supporting its relevance for hypothesis formulation. The dimension Motivation towards Achievement was also excluded, as prior studies have consistently reported its limited or non-significant influence in the context of technology adoption.[7]

The decision to adopt Hofstede's cultural dimensions was deliberate and grounded in both methodological and empirical considerations. Among the various cultural frameworks available, Hofstede's model is one of the most widely adopted in SE research (Lambiase et al. 2024a; Borchers 2003; Abufardeh and Magel 2010; Casey 2011; Lambiase et al.

[5] The dimension *Motivation towards Achievement and Success* was previously known as *Masculinity vs. Femininity* and has been recently redefined.

[6] This dimension was excluded from the present study for reasons outlined in the text.

[7] The authors acknowledge that Hofstede's framework is primarily known for its function of assigning numerical scores to characterize countries along various cultural dimensions. This may lead readers to assume that the framework was used in this study for that purpose—i.e., by applying national-level scores to characterize individuals. The authors find it important to clarify that this is not the case. The use of national scores at the individual level is a well-documented methodological fallacy and should be avoided, as it leads to misleading conclusions. Instead, in this study, Hofstede's framework is employed solely for its conceptual value in decomposing cultural background into distinct dimensions. As will be detailed later, the measurement of individual cultural values was carried out using instruments that are well-established and validated in the literature for assessing individual-level cultural traits.

2022c, b, 2024b, 2022a; Trinkenreich et al. 2023, 2024). In addition, Hofstede's cultural dimensions have been successfully integrated with the UTAUT2 framework in prior studies to examine the influence of cultural values on technology adoption (Tamilmani et al. 2021; Baptista and Oliveira 2015; Lai et al. 2016; Nistor et al. 2014). Furthermore—as will be discussed in more detail in Section 4—one of the key advantages of using Hofstede's framework is the availability of well-established and highly validated measurement instruments. These instruments strengthened the methodological rigor of the study by enabling precise and reliable operationalization of cultural values.

Having defined the four cultural dimensions retained in our investigation, we now argue why they are expected to play a role in the adoption of fairness toolkits, before formulating the corresponding hypotheses. As discussed in Section 1, fairness in software is increasingly understood as a socio-technical phenomenon: deciding whether to adopt a fairness toolkit is not only a technical choice but also a value-laden one, shaped by how practitioners interpret fairness, how they relate to authority and to their team, and how they weigh long-term ethical considerations against short-term delivery pressures. Each of the four Hofstede dimensions we consider speaks directly to one of these aspects—Power Distance shapes responsiveness to hierarchical and community-endorsed norms; Uncertainty Avoidance shapes openness to practices perceived as novel or unstructured; Long-Term Orientation shapes the willingness to invest in initiatives whose benefits accrue over time; and Collectivism shapes the weight given to team norms and shared responsibility. In what follows, we therefore translate this rationale into testable hypotheses, modeling each cultural value first as an *antecedent* of UTAUT2 constructs (Section 3.2.2) and then as a *moderator* of UTAUT2 relationships (Section 3.2.3).

3.2.2 Hypothesis on Cultural Values as Antecedents in the UTAUT2 Model

Previous studies have attested to the direct influence of individual cultural values on users' acceptance of various technologies (Tamilmani et al. 2021; Lai et al. 2016; Nistor et al. 2014; Ribiere et al. 2010; Shiu et al. 2015).

Power Distance (PD) refers to the extent to which individuals accept and expect hierarchical differences in power and authority within social and professional settings. In the context of software development, practitioners with high levels of Power Distance tend to be more receptive to the opinions and directives of authority figures and peers. As a result, they are more likely to be influenced by social cues—particularly from superiors—which highlight a direct link with Social Influence (SI). Moreover, with the rapid advancement of AI technologies, there is growing attention within the SE community toward ensuring the reliability, fairness, and ethical alignment of these tools. As these values become more institutionally reinforced, individuals with higher Power Distance may be more inclined to follow community-endorsed norms and authoritative guidance. This dynamic could lead people with high levels of Power Distance to a greater Behavioral Intention (BI) to adopt fairness-focused technologies, and ultimately, to increased actual Use Behavior (UB) of fairness toolkits (Lai et al. 2016). Based on the above rationale, we hypothesize the following:

- H8a—Power Distance (PD) positively influences software practitioners' Social Influence (SI) regarding the adoption of fairness toolkits.

– H8b—Power Distance (PD) positively influences software practitioners' Behavioral Intention (BI) to adopt fairness toolkits.
– H8c—Power Distance (PD) positively influences the actual Use Behavior (UB) of fairness toolkits.

The adoption of new technologies is often associated with innovative thinking and a tolerance for ambiguity and uncertainty (Ford et al. 2003). This characteristic stands in contrast to the preferences of individuals with high levels of Uncertainty Avoidance (UA), who tend to favor predictability and clear structures. Such a value misalignment can act as a barrier to the adoption of technological innovations. Indeed, prior research has consistently demonstrated that both individuals and organizations characterized by high Uncertainty Avoidance are less inclined to embrace emerging technologies (Ribiere et al. 2010; Lai et al. 2016). Although fairness technologies do not represent radical innovations from a technical standpoint—being implemented within classical computing frameworks—they nonetheless embody novel practices and values related to ethical AI and socio-technical responsibility. Given that activities involving fairness technologies may still be perceived as unstructured or ambiguous, it is important to investigate the potential influence of Uncertainty Avoidance. Therefore, we hypothesize the following:

– H9a—Uncertainty Avoidance (UA) negatively influences software practitioners' Behavioral Intention (BI) to adopt fairness toolkits.
– H9b—Uncertainty Avoidance (UA) negatively influences the actual Use Behavior (UB) of fairness toolkits.

Considering the adoption of fairness technologies in light of their primary objective—namely, fostering fair and responsible behavior in AI systems—it is evident that this represents a long-term commitment requiring sustained effort and continuous attention. Individuals with a high Long-Term Orientation (LT) are generally more inclined to invest in initiatives with future-oriented benefits, and as such, may be more likely to recognize the value of adopting such technologies (Lai et al. 2016). Moreover, they may be more predisposed to developing consistent usage patterns over time (Lai et al. 2016; Nistor et al. 2014). Therefore, we hypothesize the following:

– H10a—Long-Term Orientation (LT) positively influences software practitioners' Performance Expectancy (PE) regarding the adoption of fairness toolkits.
– H10b—Long-Term Orientation (LT) positively influences software practitioners' Habit (HB) regarding the adoption of fairness toolkits.
– H10c—Long-Term Orientation (LT) positively influences software practitioners' Behavioral Intention (BI) to adopt fairness toolkits.

Regarding collectivistic behavior, it has been extensively demonstrated that individuals with high levels of Collectivism (CO) are more likely to perceive stronger Social Influence (SI), greater Facilitating Conditions (FC), and increased Effort Expectancy (EE) (Abbasi et al. 2015; Childers et al. 2001; Markus and Kitayama 2014). The first two associations are well-grounded: Collectivism is inherently tied to group norms and interdependence, making individuals more responsive to the opinions of peers and more aware of the support provided by their social and organizational environments. As for the perceived EE, prior studies suggest that collectivistic individuals often engage in cooperative problem-solving and knowledge-sharing behaviors, which can reduce individual cognitive load and foster a more supportive environment for learning new technologies. This shared effort can lead to a perception that the technology is easier to use, even if the actual complexity remains the same. Therefore, we hypothesize the following:

- H11a—Collectivism (CO) positively influences software practitioners' Effort Expectancy (EE) regarding the adoption of fairness toolkits.
- H11b—Collectivism (CO) positively influences software practitioners' Social Influence (SI) regarding the adoption of fairness toolkits.
- H11c—Collectivism (CO) positively influences software practitioners' Facilitating Conditions (FC) regarding the adoption of fairness toolkits.

3.2.3 Hypothesis on Cultural Values as Moderators in the UTAUT2 Model

There has been an increasing volume of literature evidencing the moderating effects of individual espoused cultural values on technology use (Lai et al. 2016; Tamilmani et al. 2021; Baptista and Oliveira 2015; Shiu et al. 2015; Lambiase et al. 2024a).

As previously outlined, the influence of Power Distance (PD) is closely intertwined with that of Social Influence (SI) (Lee et al. 2015; Lin 2014). Individuals with high PD tend to place significant value on the opinions of superiors and authority figures. In such cultural contexts, the endorsement of a technology by hierarchical figures may substantially amplify both the intention to adopt and the subsequent Usage Behavior. This perspective aligns with prior findings in mobile technology adoption research (Baptista and Oliveira 2015), where high PD cultures were more responsive to normative pressures. Therefore, it is hypothesized that:

- H8d—Power Distance (PD) positively moderates the relationship between software practitioners' Social Influence (SI) and their Behavioral Intention (BI) to adopt fairness toolkits.
- H8e—Power Distance (PD) positively moderates the relationship between software practitioners' Behavioral Intention (BI) and their actual Use Behavior (UB) of fairness toolkits.

Uncertainty Avoidance (UA) has been shown to exert a dual moderating effect on technology adoption processes in information technology context. Specifically, individuals with high UA are more likely to rely on social cues when forming intentions to adopt new technologies, thereby strengthening the effect of Social Influence (SI) on Behavioral Intention (BI). Conversely, the same individuals may be less inclined to translate such intentions into actual adoption behavior due to their aversion to uncertainty and risk (Straub et al. 1997; Lee et al. 2013). Given the previously established link between Behavioral Intention and the actual use of fairness toolkits, the following hypotheses are proposed:

- H9c—Uncertainty Avoidance (UA) positively moderates the relationship between software practitioners' Social Influence (SI) and their Behavioral Intention (BI) to adopt fairness toolkits.

- H9d—Uncertainty Avoidance (UA) negatively moderates the relationship between software practitioners' Behavioral Intention (BI) and their actual Use Behavior (UB) of fairness toolkits.

Among Hofstede's cultural dimensions, Long-Term Orientation (LT) is uniquely concerned with how individuals incorporate future-oriented considerations into their decision-making processes. Prior studies suggest that LT plays a pivotal moderating role in the context of technology acceptance and use at the individual level (Hoehle et al. 2015). Specifically, individuals with a stronger Long-Term Orientation tend to place greater emphasis on the anticipated benefits of a technology, thereby amplifying the influence of Performance Expectancy (PE) on Behavioral Intention (BI) (Lee et al. 2013; Zhou et al. 2015). Moreover, their future-focused mindset may strengthen the relationship between intention and actual behavior, as long-term goals enhance the persistence and enactment of intentions. Based on this, the following hypotheses are proposed:

- H10d—Long-Term Orientation (LT) positively moderates the relationship between software practitioners' Performance Expectancy (PE) and their Behavioral Intention (BI) to adopt fairness toolkits.
- H10e—Long-Term Orientation (LT) positively moderates the relationship between software practitioners' Behavioral Intention (BI) and their actual Use Behavior (UB) of fairness toolkits.

The cultural dimension of individualism versus Collectivism (CO) has received considerable attention in technology adoption research. This dimension shapes the extent to which individuals prioritize personal autonomy versus group affiliation when making decisions. Individuals with a strong collectivistic orientation tend to emphasize group harmony and shared goals over personal preferences, making them more susceptible to normative pressures such as Social Influence (SI) (Abbasi et al. 2015; Ford et al. 2003). Prior studies have consistently found that Collectivism can strengthen the effect of subjective norms on Behavioral Intention (BI) (Srite and Karahanna 2006; Lee et al. 2013; Faqih and Jaradat 2015). Moreover, collectivistic tendencies have also been shown to enhance the translation of Behavioral Intention into actual technology use, possibly due to a greater commitment to group-aligned behaviors (Flight et al. 2011; Faqih and Jaradat 2015). Based on this evidence, we propose the following hypotheses:

- H11d—Collectivism (CO) positively moderates the relationship between software practitioners' Performance Expectancy (PE) and their Behavioral Intention (BI) to adopt fairness toolkits.
- H11e—Collectivism (CO) positively moderates the relationship between software practitioners' Effort Expectancy (EE) and their Behavioral Intention (BI) to adopt fairness toolkits.
- H11f—Collectivism (CO) positively moderates the relationship between software practitioners' Social Influence (SI) and their Behavioral Intention (BI) to adopt fairness toolkits.

– H11g—Collectivism (CO) positively moderates the relationship between software practitioners' Behavioral Intention (BI) and their actual Use Behavior (UB) of fairness toolkits.

Figure 1 provides an integrated graphical view of the full research model.

4 Research Design

To evaluate the research model presented in Section 3, a survey study was conducted. The collected data were subsequently analyzed using *Partial Least Squares Structural Equation Modeling (PLS-SEM)* (Hair Junior et al. 2014).

The adoption of questionnaires and PLS-SEM in the context of our study is a natural methodological choice. First, the constructs of interest are inherently latent in nature—such as cultural dimensions and self-perceived usefulness—and are therefore not directly measurable. Capturing such variables typically requires a questionnaire-based approach, for which validated instruments are readily available in the literature. Furthermore, the hypothesized model entails multiple relationships among variables, including several dependent constructs, which exceeds the capabilities of standard regression analysis. In addition, our model incorporates both mediation and moderation effects, which cannot be adequately investigated through conventional regression techniques.

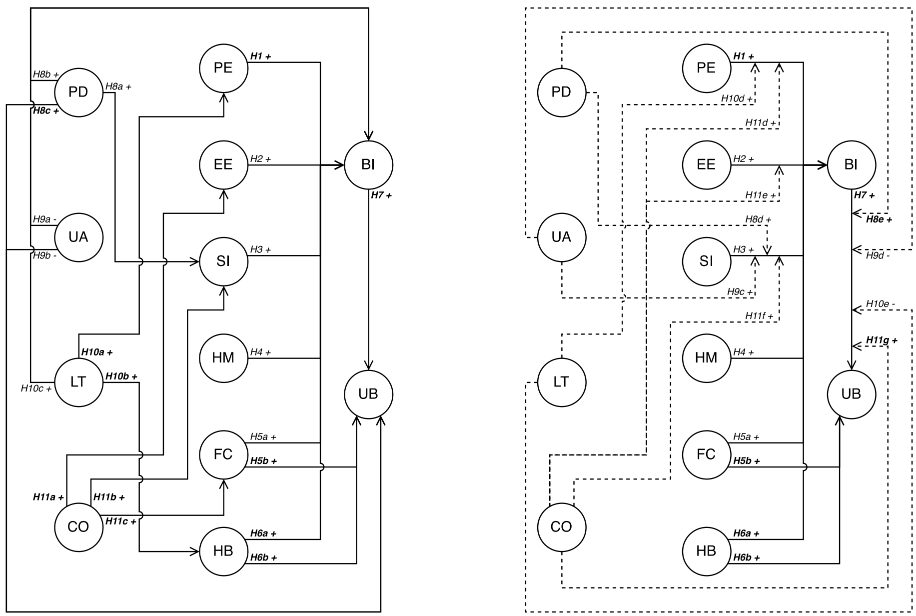

Fig. 1 Integrated graphical view of the research model. The upper panel shows cultural values acting as *antecedents* of UTAUT2 constructs; the lower panel shows them acting as *moderators* of the relationships between UTAUT2 constructs and Behavioral Intention or Use Behavior. The model is split into two panels purely for readability—having all arrows in a single diagram would have been hard to visualise—but it represents one single analysis model. A given cultural value may appear in both panels when it is tested in both roles (e.g., Power Distance as antecedent of SI and as moderator of SI → BI)

The research design followed a three-step process:

1. A questionnaire was developed based on measurement instruments validated in the literature, targeting the constructs specified in the research model introduced in Section 3.
2. The questionnaire was then administered to software practitioners with prior experience using fairness toolkits, in order to collect data for construct measurement.
3. Finally, the collected responses were analyzed using PLS-SEM to test the proposed hypotheses and evaluate the model.

The remainder of this section details each of these methodological steps.

4.1 Survey Instrument

The survey instrument was designed to measure both technology adoption factors and individual cultural values relevant to fairness toolkit adoption. The questionnaire was structured into two main parts.

The questionnaire items used to measure the core constructs of the UTAUT2 model were adapted from the original instrument proposed by Venkatesh et al. (2012), which has been widely validated in prior studies. The construct *Use Behavior* (UB), representing the frequency of actual usage, was assessed using a single-item, 6-point frequency Likert scale. The remaining seven predictor constructs were measured using a 7-point Likert agreement scale, with the following number of items per construct: *Performance Expectancy* (PE—5 items), *Effort Expectancy* (EE—6 items), *Social Influence* (SI—5 items), *Hedonic Motivation* (HM—3 items), *Facilitating Conditions* (FC—4 items), *Habit* (HB—4 items), and *Behavioral Intention* (BI—3 items).

For the measurement of cultural values, although Hofstede's framework originally included a questionnaire (Hofstede 2011), it was explicitly developed for group-level (i.e., national-level) analysis and is not suitable for assessing individual-level cultural traits. Using national scores or group-level instruments to characterize individuals is a well-documented methodological fallacy. To address this, various researchers have proposed alternative scales to operationalize Hofstede's dimensions at the individual level. In this study, the CVSCALE instrument developed by Yoo et al. (2011) was adopted, as it is widely validated for capturing individual-espoused cultural values. Specifically, the following dimensions were included: *Power Distance* (PD—5 items), *Uncertainty Avoidance* (UA—5 items), *Collectivism* (CO—6 items), and *Long-Term Orientation* (LT—6 items). The cultural dimensions were measured using a 5-point Likert agreement scale; for the LT dimension, the CVSCALE employs an importance-based scale, where 1 indicates "very unimportant" and 5 indicates "very important."

In addition to the main constructs, demographic information was collected, including gender, age, professional role, and years of experience in the software industry.

To improve clarity and usability, the survey instrument followed established guidelines for survey research in SE (Kitchenham and Pfleeger 2008). Questions were organized into clearly labeled sections and formatted using SurveyMonkey tools to improve readability.

The complete questionnaire—including the verbatim wording of all UTAUT2 and CVSCALE items, the screening items used to verify familiarity with fairness toolkits, and the four embedded attention checks—is provided in the replication package (Lambiase et al.

2026). The average completion time of approximately 10 minutes (versus 15 minutes in the first pilot) is consistent with the predominance of short Likert items adopted from previously validated scales.

An iterative pilot testing process was conducted prior to the main study. Three pilot rounds involving ten researchers from our professional network were carried out to evaluate the clarity of the questions and the overall survey structure. Based on their feedback, minor wording revisions and formatting adjustments were implemented. A final pilot test was then conducted on the Prolific platform with five eligible participants to ensure that the questionnaire functioned correctly within the data collection environment. Across all pilot rounds, we also monitored and optimized the survey completion time by refining the wording and reducing unnecessary complexity. The estimated completion time decreased progressively from 15 minutes in the first pilot, to 12 minutes in the second, and ultimately to approximately 10 minutes in the final version of the questionnaire.

4.2 Data Collection Procedure

Data collection was conducted through an online survey administered using SurveyMonkey, while participant recruitment and screening were managed through the Prolific platform.[8] This combination allowed us to leverage Prolific's participant filtering capabilities together with SurveyMonkey's survey management and data storage functionalities. Prolific has received increasing recognition within the SE research community as a reliable platform for recruiting high-quality participants, provided that rigorous filtering protocols are followed. As with any online recruitment platform, concerns about participant reliability persist. However, studies by Eyal et al. (2021) and Douglas et al. (2023) have demonstrated that Prolific—alongside Cloud Research—offers highly reliable respondent pools when appropriate methodological safeguards are in place. Notably, many of these safeguards overlap with those proposed by Russo (2022), whose protocol guided the recruitment process in this study.

The target population of this study consists of software practitioners who have exposure to and use fairness toolkits in the development or evaluation of machine learning systems. This population was selected because the research focuses on the adoption of fairness-related development tools, which require participants to possess both SE expertise and familiarity with fairness practices in AI-based systems. Consequently, the recruitment strategy aimed to identify practitioners with relevant technical backgrounds rather than a general population of software developers.

The recruitment process followed a structured multi-stage procedure. First, potential participants were identified on Prolific using the platform's built-in filtering tools to target individuals with relevant backgrounds. The inclusion criteria required participants to: (1) work in the Information Technology sector, (2) be employed either full-time or part-time, and (3) possess programming skills. Additional quality-related criteria were applied to ensure reliability and comprehension, including: (4) fluency in English (the language of the survey), (5) a 100% approval rate on Prolific, and (6) at least 25 previously approved submissions on the platform. Moreover, despite not directly addressed using filters built-in prolific, participants were required to have (7) previous experience with Fairness toolkit use; this was verified using the pre-screening questionnaire and through the introduction of it. Moreover,

[8] Prolific (www.prolific.com) [March 2025]

several measures were adopted to ensure participant reliability. An iterative pre-screening strategy was used to refine and confirm participant eligibility. Screening questions were embedded to verify respondents' professional experience as software developers (Danilova et al. 2021). In addition, a tiered compensation model was implemented to enhance participant engagement and response quality, offering £9 (rated as "good" by Prolific) for the pre-screening phase and £12 ("great") for completing the main survey.[9]

Participants who satisfied the first six criteria were invited to complete the survey hosted on SurveyMonkey. The survey consisted of two questionnaires administered within the same session: the first measuring the UTAUT2 constructs and the second assessing individual cultural values. Before starting the survey, participants were presented with an introduction describing the purpose of the study, the voluntary nature of participation, and the data privacy guarantees. Informed consent was required to proceed.

Several measures were implemented to monitor response quality and reduce potential misrepresentation. First, four attention-check items were embedded throughout the questionnaire to verify that participants were reading and answering the questions carefully. Second, SurveyMonkey settings were configured to prevent duplicate submissions and ensure that each participant could complete the survey only once. Third, the survey platform allowed monitoring of response completeness and response patterns as well as the identification of low quality responses.

The final dataset contained no missing values. Because of this configuration, no imputation procedures were required. The 181 responses included in the analysis correspond to the total set of submissions received: no responses were excluded post-hoc on the basis of the attention checks or other quality criteria. All submitted responses satisfied the embedded attention checks, which were used as a real-time quality control mechanism through Prolific; submissions failing these checks would have been rejected at submission time and would not have entered the dataset. We did not perform additional post-hoc checks targeting more subtle forms of low-quality or suspicious participation beyond those provided by SurveyMonkey's monitoring tools. All invited participants completed the survey, resulting in a response rate of 100%.

The survey followed a cross-sectional design, capturing participants' perceptions and behaviors at a single point in time. The entire data collection process was conducted in accordance with established survey guidelines in SE research (Kitchenham and Pfleeger 2008; Andrews et al. 2007). Given that the survey was conducted online, best practices outlined by Andrews et al. (2007) were also followed. To provide an example, Survey Monkey was used to (1) ensure compatibility across different devices, (2) prevent multiple submissions, (3) provide an adaptive template for various browsers and screen sizes, and (4) allow participants to leave feedback.

From an ethical standpoint, at the time of data collection, no formal instrument—such as an ethics review board—was available or required in the authors' country to obtain ethical approval. Nevertheless, the study was designed and conducted with careful attention to ethical principles, particularly those relevant to survey-based research (Hall and Flynn 2001). Participants' privacy was prioritized throughout the process. A detailed privacy statement was provided, and no sensitive business or personally identifiable information was requested. Participants were explicitly informed that their data would be used exclusively

[9] Prolific classifies participant compensation into four tiers—low, fair, good, and great—based on hourly payment rates.

for research purposes in line with the study's objectives. Participation was voluntary, limited to individuals over the age of 18, and required informed consent. Respondents were also made aware that they could withdraw at any time without penalty. Furthermore, participants were transparently informed that their anonymized responses would be published and permanently stored in the online appendix of this paper (Lambiase et al. 2026).

4.3 Participants and Sample Characteristics

The minimum required sample size was estimated through an a priori power analysis using G*Power (Faul et al. 2009). Assuming an effect size of 15%, a significance level of 5%, and a statistical power of 95%, the recommended sample size for a model with 11 predictors was 178 participants. Since all 181 participants from the prior study met the defined eligibility criteria, they were all invited to participate in the current investigation, yielding a final sample that exceeds the minimum threshold for statistical validity.

In terms of gender identity, 80% of respondents identified as male, 19% as female, and 1% as non-binary. Participants were located across 15 countries, with the majority residing in the United States (39%) and the United Kingdom (24%). Additional representation included individuals from Australia, Belgium, Canada, Finland, France, Germany, Ireland, Italy, the Netherlands, Portugal, Singapore, Spain, and Sweden, contributing to a globally diverse dataset.

A wide range of professional roles within the software and IT industry was represented. The most common occupations were Software Developer or Programmer (25%), followed by Project Manager (15%), Data Analyst, Data Engineer, or Data Scientist (14%), and Software Engineer (13%). This variety in roles ensured coverage of multiple perspectives across technical and managerial domains.

Participants also varied in age and professional experience. The largest age group was between 30 and 44 years (52%), followed by those aged 18 to 29 (25%), and 45 to 59 (20%). Only a small minority (3%) were aged 60 or older. Regarding professional experience in the software industry, 30% reported more than 10 years of experience, 28% reported 1–2 years, and 24% had 4–6 years. Respondents with 7–9 years of experience accounted for 16% of the sample, while only 1% had less than one year of industry experience.

4.4 Data Analysis

The data analysis process began with a *preliminary evaluation* aimed at ensuring data quality and reliability. This step involved checking for missing values, detecting anomalous response patterns, identifying outliers, and assessing the distribution of the data. Although PLS-SEM is more flexible than traditional statistical methods and does not require normality, these checks were performed to ensure the robustness of the dataset. No missing values or suspicious patterns were identified. Skewness and kurtosis values were examined, and all indicators fell within the commonly accepted threshold range (–2 to +2), suggesting no significant deviations from normality.

Following this validation, the dataset was imported into SmartPLS for statistical analysis using PLS-SEM (Ringle et al. 2024).[10] The analysis was conducted in two main phases: *Measurement Model Evaluation* and *Structural Model Evaluation*.

During the *Measurement Model Evaluation* phase, the reliability and validity of the constructs were assessed to ensure that the observed items adequately represented their respective latent variables. This step served as a prerequisite for the structural analysis, verifying the adequacy of the measurement instruments used in the study.

In the subsequent *Structural Model Evaluation* phase, the hypothesized relationships among constructs were tested, and the model's explanatory and predictive capabilities were assessed. As previously noted, the primary objective of this study was to investigate the role of individual cultural values in fairness toolkit adoption, rather than to re-examine the effects of all UTAUT2 constructs, which were already analyzed in earlier work (Voria et al. 2025a).

Following the approach proposed by Lai et al. (2016), we estimated three progressively extended model configurations: (1) a baseline UTAUT2-only model, (2) an extended model incorporating cultural values as *antecedents*, and (3) a full model where cultural values acted both as *antecedents* and as *moderators* of the relationships between UTAUT2 constructs and the outcome variables. The first two models were estimated primarily to illustrate how the progressive inclusion of cultural variables affects the explanatory characteristics of the model (e.g., explanatory power). The final model (Model 3) represents the complete theoretical specification and is therefore the model used for the interpretation of the structural relationships.

Within this final model, both **mediation** and **moderation** mechanisms were examined. Mediation analysis was used to assess whether the effects of cultural values on the dependent variables—Behavioral Intention (BI) and Use Behavior (UB)—were indirectly transmitted through UTAUT2 constructs. Moderation analysis, instead, evaluated whether cultural values influenced the strength or direction of the relationships between UTAUT2 predictors and the outcome variables. In PLS-SEM, moderation is modeled by introducing interaction constructs (interaction terms) between the moderator and the relevant predictor variables, which are then included in the structural model and evaluated alongside the direct effects. Together, these analyses offered a nuanced perspective on how cultural values shape fairness toolkit adoption.[11]

5 Analysis of the Results

The following section presents the results of the PLS-SEM analysis. For readability and clarity, detailed information can be found in the online appendix (Lambiase et al. 2026), while the main findings are reported here.

[10] For an in-depth overview of the PLS-SEM methodology, refer to Hair Junior et al. (2014) and Russo and Stol (2021).

[11] As discussed by Hair Junior et al. (2014), PLS-SEM is particularly suitable for models involving both mediation and moderation relationships within the same structural framework.

5.1 Measurement Model Evaluation

As a first step in the evaluation of the theoretical model, it is paramount to evaluate the reliability of the constructs of the model (Hair Junior et al. 2014; Russo and Stol 2021). Consequently, we analyze the indicator reliability, internal consistency reliability, convergent validity, and discriminant validity. This section presents the obtained results for each of the steps mentioned above.

5.1.1 Indicator Reliability

The evaluation of the measurement model began with an assessment of *indicator reliability*, which examines how well each questionnaire item (indicator) reflects its associated latent construct. Outer loadings represent the strength of the relationship between an indicator and its construct. A widely accepted threshold is 0.708, indicating that the indicator shares more than 50% of its variance with the construct. Indicators with loadings below 0.40 are typically discarded, while those between 0.40 and 0.70 are subject to removal if doing so improves the model's internal consistency or convergent validity (Hair Junior et al. 2014).

All indicators associated with the UTAUT2 constructs exceeded the recommended threshold, confirming their reliability. In contrast, a few indicators related to the cultural values constructs did not meet the criteria—an outcome consistent with prior studies employing cultural dimensions (Lambiase et al. 2024a; Lai et al. 2016). Accordingly, all indicators with loadings below 0.40 were removed, and those falling between 0.40 and 0.70 were reviewed using the aforementioned criteria. Ultimately, five indicators were excluded from the model: LT2, LT4, LT5, PD3, and PD5.

5.1.2 Internal Consistency Reliability

The second step involved assessing *internal consistency reliability* to determine whether the indicators reliably measure their respective latent constructs. We reported three widely used metrics: *Cronbach's Alpha* (α), *Composite Reliability* (ρ_c), and the *Reliability Coefficient* (ρ_A).

As shown in Table 3, all constructs met or exceeded the commonly recommended threshold of 0.70 for *Composite Reliability* for each of these measures (Hair Junior et al. 2014).[12] The overall results confirm an acceptable level of internal consistency across constructs, providing a reliable foundation for further analysis.

[12] As noted by Hair Junior et al. (2014), in PLS-SEM the recommended reliability indicator is the Composite Reliability, rather than Cronbach's Alpha. While Cronbach's Alpha is traditionally used in covariance-based SEM, it relies on the assumption of tau-equivalence (i.e., that all indicators have equal loadings), which is rarely met in practice and often leads to an underestimation of construct reliability. In contrast, Composite Reliability provides a more accurate reliability estimate for PLS-SEM because it accounts for the different loadings of indicators within a construct and is therefore considered a more consistent reliability estimator in variance-based structural equation modeling. For this reason, methodological guidelines for PLS-SEM recommend evaluating construct reliability primarily through Composite Reliability rather than relying solely on Cronbach's Alpha.

5.1.3 Convergent Validity

Convergent validity refers to the extent to which the indicators of a given construct are positively correlated, reflecting a high degree of shared variance (Hair Junior et al. 2014). As all constructs in this model were specified using reflective measurement, a substantial level of convergence among indicators was expected. The primary metric used to assess convergent validity is the *Average Variance Extracted* (AVE), with values of 0.50 or higher indicating that a construct explains more than half of the variance in its indicators.

As reported in Table 3, all constructs exhibited AVE values exceeding the 0.50 threshold. These results confirm the presence of strong convergent validity and support the adequacy of the measurement model.

5.1.4 Discriminant Validity

Discriminant validity assesses whether a construct is sufficiently distinct from other constructs in the model, both conceptually and empirically. To evaluate this, we adopted the *Heterotrait-Monotrait Ratio of Correlations* (HTMT), a method introduced by Henseler et al. (2015), which has been shown to offer greater reliability than earlier approaches. HTMT values above 0.90 suggest a lack of discriminant validity, values between 0.85 and 0.90 are considered acceptable, and values below 0.85 indicate strong discriminant validity.

Complete HTMT results are provided in the online appendix (Lambiase et al. 2026). All constructs showed HTMT values below the recommended threshold of 0.85, with the exception of one value slightly exceeding it (0.875). To verify the robustness of these results, a bootstrapping procedure with 10 000 subsamples was performed, using a one-tailed test at a significance level of 0.05. The analysis confirmed that all HTMT values remained statistically below the critical cutoffs, providing additional support for satisfactory discriminant validity across the model.

5.2 Structural Model Evaluation

Following the evaluation of the measurement model, the structural model was assessed to examine the relationships between constructs and validate the hypotheses.

Table 3 Measurement model evaluation metrics

Construct	Cronbach α	ρ_A	ρ_c	AVE
Performance Expectancy (PE)	0.915	0.921	0.936	0.746
Effort Expectancy (EE)	0.868	0.883	0.899	0.599
Social Influence (SI)	0.866	0.883	0.902	0.648
Hedonic Motivation (HM)	0.920	0.937	0.950	0.863
Facilitating Conditions (FC)	0.817	0.827	0.879	0.647
Habit (HB)	0.841	0.866	0.894	0.681
Behavioral Intention (BI)	0.888	0.889	0.931	0.817
Power Distance (PD)	0.563	0.561	0.775	0.536
Uncertainty Avoidance (UA)	0.801	0.816	0.861	0.554
Long-Term Orientation (LT)	0.570	0.681	0.763	0.524
Collectivism (CO)	0.822	0.830	0.869	0.526

5.2.1 Collinearity Analysis

The initial step involved assessing collinearity between constructs to ensure reliable path estimations. To detect potential multicollinearity, the *Variance Inflation Factor* (VIF) was used, a widely accepted measure in regression analysis. A VIF value below 3 is preferred, while values under 5 are typically considered acceptable.

In the analysis, most of the VIF values fell below the preferred threshold of 3, indicating minimal collinearity among variables. Only four values slightly exceeded this threshold, with the highest reaching 3.989, but all remained well within the acceptable upper limit of 5. These findings confirm that multicollinearity does not pose a significant issue in the structural model.

5.2.2 Significance and Relevance

The structural model analysis was conducted to assess the significance and relevance of the hypothesized relationships. Given the complexity of the research model, the presentation of the results was organized as follows: first, the significance and relevance of each direct hypothesis are examined based on path coefficients and corresponding *p*-values; second, indirect effects were analyzed in order to find potential relevant mediating effects; third, the moderating effects are discussed, with a focus on the significant interaction terms and their effect sizes; finally, an overview of the total effects of the predictors on the two main dependent variables—Behavioral Intention (BI) and Use Behavior (UB)—is provided.

Significance of the Hypothesis Significance testing was performed using the bootstrapping method with 10 000 sub-samples, employing a complete approach and two-tailed test type (Hair Junior et al. 2014). The analysis yielded *T*-statistics, *p*-values, and the original path coefficients, which indicate the relative strength of each relationship in the structural model. Moderation effects were assessed through the inclusion of interaction terms, enabling the identification of conditional effects on the dependent variables.

The results (reported in Table 4, and graphically represented in Fig. 2) revealed that among the UTAUT2 predictors, Performance Expectancy ($\beta = 0.429$, $p < 0.001$) and Habit ($\beta = 0.375$, $p < 0.001$) had a statistically significant and positive effect on Behavioral Intention. In contrast, Effort Expectancy, Facilitating Conditions, Social Influence, and Hedonic Motivation were not significant predictors of Behavioral Intention. Regarding Use Behavior, both Habit ($\beta = 0.426$, $p < 0.001$) and Behavioral Intention ($\beta = 0.183$, $p = 0.032$) emerged as significant positive predictors. Additionally, the cultural dimension Power Distance had a weak but significant positive influence on both Social Influence ($\beta = 0.185$, $p = 0.007$) and Use Behavior ($\beta = 0.152$, $p = 0.031$). Long-Term Orientation was significantly associated with Habit ($\beta = 0.278$, $p = 0.010$) and PE ($\beta = 0.312$, $p = 0.000$), suggesting a key role in shaping user engagement. Furthermore, Collectivism showed significant associations with both Effort Expectancy ($\beta = 0.324$, $p < 0.001$), Social Influence ($\beta = 0.283$, $p < 0.001$) and Facilitating Conditions ($\beta = 0.221$, $p < 0.001$). These findings indicate that, in addition to core UTAUT2 constructs, cultural dimensions—particularly Power Distance, Long-Term Orientation, and Collectivism—substantially contributed to explaining both intention and actual use behavior.

Table 4 Significance and Relevance of the Hypotheses. Significance levels are denoted as follows: **(.)**$p < 0.1$, **(*)**$p < 0.05$, **(**)**$p < 0.01$, **(***)**$p < 0.001$

ID	Hypotheses	Coeff. (β)	T	p-value
H1	PE → BI	0.429	4.958	0.000 (***)
H2	EE → BI	0.062	0.641	0.522
H3	SI → BI	-0.033	0.445	0.656
H4	HM → BI	-0.08	0.981	0.327
H5a	FC → BI	0.047	0.604	0.546
H5b	FC → UB	0.072	1.200	0.230
H6a	HB → BI	0.375	4.159	0.000 (***)
H6b	HB → UB	0.426	4.852	0.000 (***)
H7	BI → UB	0.183	2.150	0.032 (*)
H8a	PD → SI	0.185	2.675	0.007 (**)
H8b	PD → BI	-0.059	1.121	0.262
H8c	PD → UB	0.152	2.159	0.031 (*)
H9a	UA → BI	0.042	0.755	0.450
H9b	UA → UB	-0.071	1.088	0.277
H10a	LT → PE	0.312	5.396	0.000 (***)
H10b	LT → HB	0.278	4.011	0.000 (***)
H10c	LT → BI	0.103	1.568	0.117
H11a	CO → EE	0.324	4.899	0.000 (***)
H11b	CO → SI	0.283	3.787	0.000 (***)
H11c	CO → FC	0.221	3.517	0.000 (***)

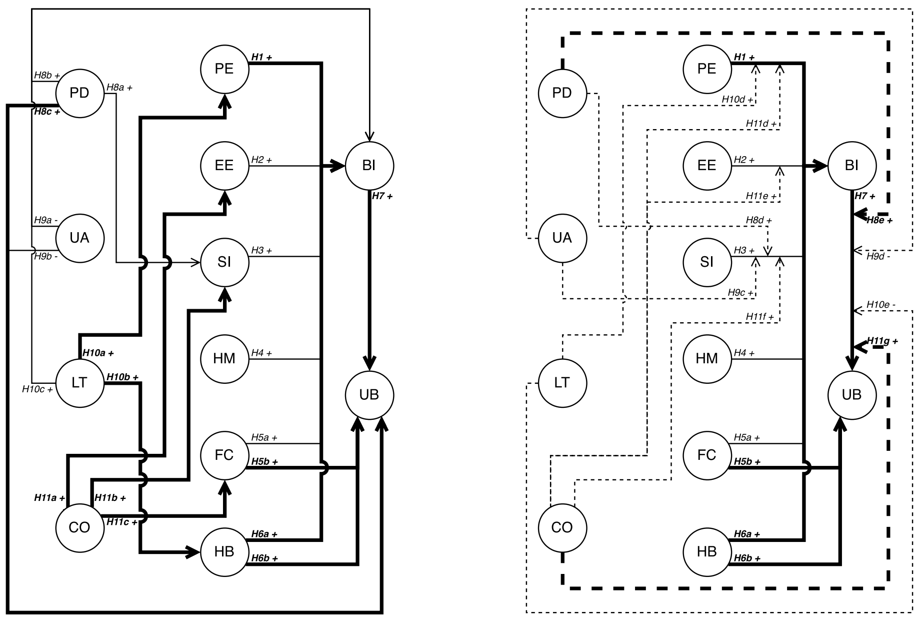

Fig. 2 Research model with results. The upper panel reports the hypotheses in which cultural values act as *antecedents* of UTAUT2 constructs; the lower panel reports those in which they act as *moderators* of UTAUT2 relationships. Solid bold lines denote relationships that are statistically supported by the analysis (path coefficients reported on each arrow); dashed lines denote hypothesized relationships that were not supported. A given cultural value can appear in both panels because the same construct is tested in both roles (e.g., Power Distance as antecedent of SI and as moderator of SI → BI). The two panels are shown separately only for readability; they represent a single analysis model

Table 5 Effect size (f^2) of the interaction terms for moderation analysis. Significance levels are denoted as follows: **(.)**$p < 0.1$, **(*)**$p < 0.05$, **(**)**$p < 0.01$, **(***)**$p < 0.001$. Interpretation of f^2 effect sizes for moderation remains debated. Following Kenny and Judd (2019), effect sizes are annotated as: **(+)** small ($f^2 \geq 0.005$), **(++)** medium ($f^2 \geq 0.01$), **(+++)** large ($f^2 \geq 0.025$)

ID	Hypotheses	f^2	p-value
H8d	PD × SI → BI	0.002	0.523
H8e	PD × BI → UB	0.031 (+++)	0.031 (*)
H9c	UA × SI → BI	0.001	0.666
H9d	UA × BI → UB	0.008	0.193
H10d	LT × PE → BI	0.028 (++)	0.114
H10e	LT × BI → UB	0.003	0.467
H11d	CO × PE → BI	0.008	0.256
H11e	CO × EE → BI	0.013	0.182
H11f	CO × SI → BI	0.000	0.905
H11g	CO × BI → UB	0.019 (++)	0.023 (*)

Mediation Analysis To explore potential mediation effects within the structural model, we analyzed the indirect influence of key predictors on Use Behavior through Behavioral Intention (reported in Table 4). The results confirmed a significant mediating role of Behavioral Intention in the relationship between Performance Expectancy and Use Behavior, with a significant indirect effect ($\beta = 0.106$, $p < 0.001$). A similar mediating effect was found for Habit, which influenced Use Behavior both directly and indirectly via Behavioral Intention ($\beta = 0.093$, $p < 0.001$). These findings highlight the central role of Behavioral Intention as a mediating mechanism in the model, through which Performance Expectancy and Habit exert their influence on actual Usage Behavior. No significant mediation effects were observed for Effort Expectancy, Facilitating Conditions, Social Influence, or Hedonic Motivation.

Moderation Analysis Regarding the moderation analysis (results reported in Table 5), interpretation requires an understanding of the *interaction terms*, which represent moderation effects. For example, the hypothesis *PD* × *SI* → *BI* refers to the moderating effect of Power Distance on the relationship between Social Influence and Behavioral Intention, where *PD* × *SI* is the interaction term. According to Hair Junior et al. (2014), in moderation analysis, the relevance of a moderating effect should not be assessed via the path coefficient but rather through the f^2 effect size, which is computed as part of the PLS-SEM algorithm. Statistical significance, instead, is evaluated via p-values derived from the same bootstrapping procedure adopted for direct effects.

Among all tested moderating effects, only two yielded statistically significant results. The interaction term *PD* × *BI* → *UB* showed a significant effect ($f^2 = 0.031, p = 0.031$), with a large effect size according to traditional guidelines ($f^2 > 0.025$). Similarly, *CO* × *BI* → *UB* was also significant ($f^2 = 0.019, p = 0.023$), with a medium effect size ($f^2 > 0.01$).[13] These results suggest that high levels of Power Distance strengthen the effect of Behavioral Inten-

[13] The interpretation of f^2 effect sizes for moderation is still debated (Hair Junior et al. 2014). While Cohen (2013) proposed the classical thresholds of 0.02, 0.15, and 0.35 for small, medium, and large effects respectively, more recent work by Kenny and Judd (2019), building on Aguinis et al. (2005), suggests more realistic thresholds of 0.005, 0.01, and 0.025; we used them in this study.

tion on Use Behavior, and that high levels of Collectivism enhance the relationship between Behavioral Intention and Use Behavior.

Analysis of the Total Effect The total effect analysis was conducted to estimate the overall influence of each construct on the two dependent variables—Behavioral Intention (BI) and Use Behavior (UB)—taking into account both direct effects and those transmitted through mediation or moderation. To enable a structured and comparative interpretation, the analysis was conducted across three progressively extended models: the first included only the core UTAUT2 predictors; the second integrated additional cultural antecedents (i.e., Power Distance (PD), Uncertainty Avoidance (UA), Long-Term Orientation (LT), and Collectivism (CO)); and the third incorporated interaction terms to account for moderating effects. This incremental modeling strategy enabled the isolation of cultural influences and the evaluation of whether these act directly on outcomes or through conditional mechanisms.

As shown in Table 6, Performance Expectancy, Habit, and Long-Term Orientation consistently exhibited a significant and positive total effect on Behavioral Intention across all model specifications. In the most comprehensive model (Antecedents + Moderators), Performance Expectancy ($\beta = 0.421$, $p < 0.001$), Habit ($\beta = 0.365$, $p < 0.001$), and Long-Term Orientation ($\beta = 0.345$, $p < 0.001$) emerged as the most influential factors on Behavioral Intention.

With respect to Use Behavior, Habit was the strongest and most consistent predictor, yielding a total effect of $\beta = 0.492$ ($p < 0.001$) in the final comprehensive model. Behav-

Table 6 Significance and relevance of the Total Effect of the factors on the dependent variable. Significance levels are denoted as follows: **(.)**$p < 0.1$, **(*)**$p < 0.05$, **(**)**$p < 0.01$, **(***)**$p < 0.001$

	Cst.	UTAUT2		Antecedents		Ant. + Mod.	
		Coeff.	***p*-value**	Coeff.	***p*-value**	Coeff.	***p*-value**
BI	PE	0.465	0.000 (***)	0.443	0.000 (***)	0.421	0.000 (***)
	EE	0.056	0.531	0.044	0.607	0.048	0.622
	SI	0.011	0.862	-0.001	0.985	-0.019	0.801
	HM	-0.067	0.394	-0.058	0.452	-0.073	0.371
	FC	0.090	0.242	0.072	0.335	0.067	0.390
	HB	0.323	0.000 (***)	0.334	0.000 (***)	0.365	0.000 (***)
	PD	*NA*	*NA*	-0.071	0.124	-0.077	0.173
	UA	*NA*	*NA*	0.034	0.519	0.042	0.449
	LT	*NA*	*NA*	0.350	0.000 (***)	0.345	0.000 (***)
	CO	*NA*	*NA*	0.030	0.354	0.043	0.461
UB	PE	0.075	0.042 (*)	0.085	0.019 (*)	0.077	0.056 (.)
	EE	0.009	0.570	0.008	0.631	0.009	0.672
	SI	0.002	0.874	-0.000	0.986	-0.004	0.827
	HM	-0.011	0.439	-0.011	0.476	-0.013	0.428
	FC	0.106	0.054 (.)	0.095	0.094 (.)	0.084	0.161
	HB	0.505	0.000 (***)	0.461	0.000 (***)	0.492	0.000 (***)
	BI	0.161	0.024 (*)	0.193	0.007 (**)	0.183	0.032 (*)
	PD	*NA*	*NA*	0.188	0.007 (**)	0.138	0.042 (*)
	UA	*NA*	*NA*	-0.036	0.565	-0.063	0.316
	LT	*NA*	*NA*	0.178	0.000 (***)	0.078	0.331
	CO	*NA*	*NA*	0.024	0.149	0.100	0.175

ioral Intention also showed a consistently significant and positive total effect on Use Behavior ($\beta = 0.183$, $p = 0.032$), confirming its mediating role. Among the cultural variables, Power Distance had a weak but statistically significant total effect on Use Behavior ($\beta = 0.138$, $p = 0.042$), indicating that higher levels of Power Distance may positively influence actual usage. Performance Expectancy also approached significance in relation to Use Behavior in the final model ($\beta = 0.077$, $p = 0.056$).

Notably, Long-Term Orientation had a significant total effect on Use Behavior in the antecedents-only model ($\beta = 0.178$, $p < 0.001$), but this effect became non-significant when moderation terms were added. None of the interaction terms involving Long-Term Orientation reached statistical significance, suggesting that Long-Term Orientation does not exert a moderating influence on the relationships between Performance Expectancy or Behavioral Intention and Use Behavior. The reduction in significance is therefore likely attributable to increased model complexity rather than a lack of underlying effect. As such, Long-Term Orientation should be interpreted as a direct antecedent of Use Behavior rather than a moderator.

Overall, these results confirm the central role of Performance Expectancy, Habit, and Long-Term Orientation in shaping Behavioral Intention, and emphasize the importance of Habit, Performance Expectancy, and Behavioral Intention in predicting Use Behavior. Cultural dimensions exhibited weaker and less consistent effects, with Power Distance and Long-Term Orientation emerging as the only cultural factors showing a significant total influence on Use Behavior.

5.2.3 Explanatory Power

The third step of the analysis involved assessing the model's explanatory power, specifically its ability to account for variance in the dependent variables by quantifying the strength of the associations—i.e., how well the model fits the data (Hair Junior et al. 2014; Russo and Stol 2021). This was primarily evaluated using the *coefficient of determination* (R^2), which ranges from 0 to 1 and indicates the proportion of variance explained by the predictors. Although there is no universally accepted threshold for what constitutes a "good" R^2 , values of 0.10 are often considered minimally acceptable, while values above 0.19 are commonly seen as indicative of moderate explanatory power (Chin et al. 1998; Raithel et al. 2012; Hair Junior et al. 2014).

The model that includes only the UTAUT2 constructs yielded an R^2 of 0.632 for Behavioral Intention and 0.467 for Use Behavior. A similar level of explanatory power was observed in the most comprehensive model—integrating both cultural antecedents and interaction terms—which reported the same R^2 values for Behavioral Intention (0.632) and Use Behavior (0.467). The intermediate model, incorporating only cultural antecedents without interaction terms, resulted in an R^2 of 0.629 for Behavioral Intention and 0.428 for Use Behavior. These results indicate that the proposed models are capable of explaining approximately 63% of the variance in Behavioral Intention and between 43% and 47% of the variance in actual use behavior. All values fall within acceptable thresholds and suggest robust explanatory power. Moreover, as none of the R^2 values approach 0.90, concerns related to model overfitting can be reasonably dismissed.

Following the assessment of the R^2, the f^2 *effect size* was analyzed to evaluate the relative contribution of each construct to the variance explained in the dependent variables. The

Table 7 Effect size (f^2) for hypotheses in order to assess the explanatory power of the model with cultural values both as antecedents and moderators. Significance levels are denoted as follows: **(.)**$p < 0.1$, **(*)**$p < 0.05$, **(**)**$p < 0.01$, **(***)**$p < 0.001$. Interpretation of f^2 effect sizes is different between antecedents and for moderation. For antecedents (above the separator line), following classical values (Cohen 2013), effect sizes are annotated as: **(+)** small ($f^2 \geq 0.02$), **(++)** medium ($f^2 \geq 0.15$), **(+++)** large ($f^2 \geq 0.35$). For moderation (below the seperator line), following Kenny and Judd (2019), effect sizes are annotated as: **(+)** small ($f^2 \geq 0.005$), **(++)** medium ($f^2 \geq 0.01$), **(+++)** large ($f^2 \geq 0.025$)

ID	Hypotheses	f^2	p-value
H1	PE → BI	0.162 (++)	0.000 (***)
H2	EE → BI	0.002	0.522
H3	SI → BI	0.000	0.656
H4	HM → BI	0.007	0.327
H5a	FC → BI	0.005	0.546
H5b	FC → UB	0.006	0.230
H6a	HB → BI	0.114 (+)	0.000 (***)
H6b	HB → UB	0.127 (+)	0.000 (***)
H7	BI → UB	0.025 (+)	0.032 (*)
H8a	PD → SI	0.036 (+)	0.007 (**)
H8b	PD → BI	0.012	0.262
H8c	PD → UB	0.038 (+)	0.031 (*)
H9a	UA → BI	0.003	0.450
H9b	UA → UB	0.007	0.277
H10a	LT → PE	0.108 (+)	0.000 (***)
H10b	LT → HB	0.084 (+)	0.000 (***)
H10c	LT → BI	0.024 (+)	0.117
H11a	CO → EE	0.118 (+)	0.000 (***)
H11b	CO → SI	0.085 (+)	0.000 (***)
H11c	CO → FC	0.051 (+)	0.000 (***)
H8d	PD × SI → BI	0.002	0.523
H8e	PD × BI → UB	0.031 (+++)	0.031 (*)
H9c	UA × SI → BI	0.001	0.666
H9d	UA × BI → UB	0.008 (+)	0.193
H10d	LT × PE → BI	0.028 (++)	0.114
H10e	LT × BI → UB	0.003	0.467
H11d	CO × PE → BI	0.008 (+)	0.256
H11e	CO × EE → BI	0.013 (++)	0.182
H11f	CO × SI → BI	0.000	0.905
H11g	CO × BI → UB	0.019 (++)	0.023 (*)

f^2 metric measures the change in R^2 when a given predictor is omitted from the model, thereby indicating the individual impact of that construct.[14]

Table 7 reports all the f^2 values for the structural paths considered, including interaction terms, which were already analyzed in the context of moderation effects (Table 5).

Among the UTAUT2 constructs, Performance Expectancy ($f^2 = 0.162$ for Behavioral Intention) and Habit ($f^2 = 0.114$ for Behavioral Intention, $f^2 = 0.127$ for Use Behavior) had the largest effect sizes, confirming their substantial role in shaping both Behavioral Intention and Use Behavior. Behavioral Intention also had a meaningful effect size on Use Behavior ($f^2 = 0.025$), just above the small-effect threshold. Facilitating Conditions exhibited marginal contributions ($f^2 = 0.005$–0.006), and the effect sizes of Effort Expectancy, Social Influence, and Hedonic Motivation were negligible, aligning with their non-significant path coefficients.

[14] Hair Junior et al. (2014) noted that f^2 values are to some extent redundant when path coefficients are reported, yet they remain valuable when discussing model explanatory power and for moderators evaluation in general.

Table 8 Predictive power of the model

Indicator	PLS-SEM			Benchmark	
	Q^2	RMSE	MAE	RMSE	MAE
BI1	0.090	1.049	0.800	1.005	0.791
BI2	0.008	1.441	1.114	1.279	0.981
BI3	0.029	1.268	0.959	1.257	0.981
UB1	0.081	1.309	1.036	1.223	0.970

Among the cultural constructs, Long-Term Orientation and Collectivism demonstrated the strongest effects. Long-Term Orientation showed a meaningful impact on Performance Expectancy ($f^2 = 0.108$) and Habit ($f^2 = 0.084$), indicating that individuals with a future-oriented mindset are more likely to perceive the technology as beneficial and to integrate it into their routines over time. Furthermore, Collectivism had a medium effect on Effort Expectancy ($f^2 = 0.118$), and smaller but relevant effects on Social Influence ($f^2 = 0.085$) and Facilitating Conditions ($f^2 = 0.051$). These results suggest that collectivist values enhance the perception that the technology is easy to use, socially endorsed, and well-supported by the surrounding environment. Power Distance also showed small yet significant effects on Social Influence ($f^2 = 0.036$) and Use Behavior ($f^2 = 0.038$), underscoring the role of hierarchical structures in shaping perceived social norms and actual usage patterns.

In summary, the f^2 analysis reinforces the central importance of Performance Expectancy, Habit, and Long-Term Orientation, while highlighting the indirect but relevant role of cultural variables such as Power Distance and Collectivism in shaping user perceptions and behavior. The findings also underscore the importance of including moderating effects, particularly those involving cultural dimensions, as they help capture contextual variations in the strength of adoption mechanisms. Specifically, both Power Distance and Collectivism significantly moderate the impact of Behavioral Intention on Use Behavior, indicating that cultural traits not only affect antecedents but also influence the translation of intention into actual behavior. These results support the integration of both direct and moderating cultural factors in models of technology adoption, offering a more nuanced understanding of user behavior across diverse cultural settings.

5.2.4 Predictive Power

To evaluate the model's applicability for managerial decision-making, its out-of-sample predictive power was assessed—i.e., its ability to generalize beyond the data used in the estimation process. This was achieved using the $\text{PLS}_{\text{predict}}$ procedure (Shmueli et al. 2016), which involves partitioning the dataset into training and holdout subsets. The assessment relied on several metrics, including Stone-Geisser's Q^2 statistic, *mean absolute error* (MAE), and *root mean square error* (RMSE). These were evaluated against a benchmark model, as recommended by Shmueli et al. (2016, 2019), which uses results from a linear regression model (LM) for comparison.

Positive Q^2 values indicate that the prediction error of the PLS-SEM model is lower than that of the benchmark, while smaller MAE and RMSE values also reflect stronger predictive performance. Table 8 summarizes the predictive metrics for the dependent variables. All Q^2 values are positive and exceed those of the benchmark model, indicating meaningful predictive relevance. Although the RMSE and MAE values are largely comparable between the models, the PLS-SEM approach demonstrates satisfactory predictive accuracy overall, supporting its utility for practical, data-driven decision-making.

Summary of the Results.

The analysis confirmed that the measurement model satisfied key validity and reliability criteria. In the structural model, Performance Expectancy and Habit were the strongest predictors of Behavioral Intention and Use Behavior. Cultural values—especially Long-Term Orientation, Power Distance, and Collectivism—showed significant but often indirect effects, influencing intermediate constructs or acting as moderators. Notably, Power Distance and Collectivism significantly enhanced the effect of Behavioral Intention on Use Behavior. The explanatory power of the model was robust (R^2 = 0.632 for intention; R^2 = 0.467 for use), and its predictive accuracy was supported by positive Q^2 values. These findings suggest that both traditional adoption constructs and individual-level cultural values play an important role in fairness technology adoption, offering theoretical and practical insights into socio-technical dynamics in software development.

6 Discussion and Implications

This section discusses our findings in light of the theoretical model and empirical results presented in Section 5, providing implications for both practitioners (indicated using the symbol) and researchers (using the symbol).

As a preliminary remark, it is important to highlight that the structural model demonstrated good predictive power, as shown in Section 5.2, with relevant path coefficients and statistically significant total effects on both Behavioral Intention and Use Behavior. This result indicates that the model can serve as a valuable analytical tool to support decision-making processes regarding the integration of fairness toolkits into software development workflows. It also provides a robust foundation for the practical implications that are discussed in the remainder of this section.

6.1 Technology Adoption Factors

Turning to the UTAUT2 constructs, the results (see Table 6) are—as expected—largely consistent with our previous study on the adoption of fairness toolkits (Voria et al. 2025a). In line with that work, *Performance Expectancy* and *Habit* emerged as the strongest predictors of both intention and actual use, highlighting the central role of perceived utility and routine integration in shaping adoption behavior.

Our results suggest that fairness-toolkit adoption differs from traditional technology adoption contexts in that it is less driven by usability- or context-related factors and more by perceived impact and internalized motivations. Indeed, the significance of Performance Expectancy and Habit indicates that practitioners are primarily motivated by the effectiveness of fairness toolkits in mitigating bias and by their integration into established development workflows. Unlike many end-user-oriented technologies, fairness toolkits are typically used by experienced practitioners within complex development environments, where tools are more likely to be selected based on their ability to produce meaningful improvements and to fit within established pipelines (Deng et al. 2022; Holstein et al. 2019). In this context, effectiveness and workflow compatibility become the primary drivers of adoption: practitioners may be willing to engage with tools that require additional effort if they provide tangible benefits, and repeated integration into development activities naturally fosters habitual use.

Conversely, the non-significance of Effort Expectancy suggests that ease of use is not a primary determinant in this setting, likely due to the professional profile of participants, who are experienced practitioners accustomed to interacting with complex tools and development environments. Similarly, the lack of significance of Social Influence and Facilitating Conditions indicates that fairness-tool adoption is not yet strongly shaped by organizational norms or infrastructural support. This is likely due to the still-emerging nature of fairness practices within software engineering, where formal processes, guidelines, and shared standards are not yet consistently established across organizations (Ferrara et al. 2024b). As a result, practitioners may not experience strong normative pressure from peers or management, nor rely on well-defined supporting infrastructures, leading to a more individual and self-driven adoption process. Finally, the non-significance of Hedonic Motivation reinforces the utilitarian and responsibility-driven nature of fairness toolkits, where adoption is driven by perceived necessity and impact rather than by experiential or affective factors.

To sum up, these findings suggest that fairness-toolkit adoption is characterized by a shift from traditional technology acceptance mechanisms toward a more impact- and value-driven process, where adoption decisions are less influenced by usability, social, or experiential factors, and more by the perceived effectiveness of the tool and its alignment with professional responsibilities and development practices. Within this landscape, cultural values play a distinctive role, which is discussed in the following section.

> 💡 From a research perspective, these findings highlight the need for future investigations to explore alternative or complementary theoretical lenses—such as the Diffusion of Innovation (DoI) theory (Rogers 2003) or the Technology Acceptance Model (TAM) (Davis et al. 1989)—to achieve a more nuanced understanding of the cognitive, social, and contextual mechanisms that shape fairness toolkit adoption.

6.2 The Role of Cultural Values on Fairness Toolkits Adoption

The model exhibited strong explanatory power, with high R^2 values for both Behavioral Intention (BI) and Use Behavior (UB). Notably, the inclusion of cultural values—both as antecedents and as moderators—led to a consistent increase in R^2 across both dependent variables, reinforcing the conclusion that cultural dimensions contribute meaningfully to the overall explanatory capability of the model. These findings underscore the importance of considering individual-level cultural factors when modeling technology adoption in socio-technical domains.

In terms of statistical significance and actual contribution, the role of cultural constructs proves both complex and theoretically rich. Three out of the four cultural dimensions examined—Power Distance, Long-Term Orientation, and Collectivism—exhibited statistically significant effects in the structural model. These effects were not limited to direct influences on Behavioral Intention or use behavior, but also included mediation pathways and moderation interactions, thus revealing a multifaceted relationship between culture and fairness technology adoption.

> ☞ Managers should consider administering validated cultural value questionnaires to monitor individual orientations and use this information to inform task delegation and team composition. Doing so may enable more culturally aligned decision-making processes, improving both the likelihood of fairness toolkit adoption and the overall effectiveness of fairness-oriented practices.

> 💡 Researchers should explore the influence of software developers' individual cultural values on various aspects of fairness, beyond just fairness toolkits, to enrich the body of knowledge and support better decision-making.

6.2.1 On the Role of Power Distance

Power Distance (PD) emerged as a significant factor both as an antecedent and as a moderator. As an antecedent, PD positively influenced Social Influence (SI) and Use Behavior (UB), albeit with small effect sizes. As a moderator, it significantly strengthened the relationship between Behavioral Intention (BI) and Use Behavior (UB), with a large effect size. These findings are aligned with prior research on the adoption of other technologies and suggest that individuals with higher levels of PD—those who value hierarchical structures and are more inclined to follow authority—are more likely to perceive strong normative pressures from peers and team leaders to adopt fairness toolkits. However, this perceived social influence does not directly translate into increased adoption behavior, which explains the non-significance of PD in the total effect analysis.

> ☞ Although not directly supporting Behavioral Intention or use, the perception of strong social expectations related to fairness may nonetheless contribute to broader awareness. Thus, fostering a culture where fairness is publicly recognized as a collective concern may help disseminate knowledge and encourage dialogue about fairness toolkits among practitioners.

Moreover, the strong and significant moderation effect of PD highlights its role in shaping how behavioral intention translates into actual use, suggesting that individuals with stronger hierarchical orientations may be more responsive to formal adoption initiatives or leadership-driven mandates. In this sense, PD appears to play a meaningful role in facilitating the enactment of fairness-related practices within software development teams. While the direct relationship between PD and BI is positive but not statistically significant in the final model—and therefore cannot be interpreted as conclusive evidence—it may still provide some contextual indication that aligns with this broader pattern, pointing to a potential tendency that could be further explored in future work.

> ☞ Despite the lack of a significant direct effect on BI, PD may still play a role in supporting fairness toolkit adoption through its moderating influence on the relationship between intention and actual use. This suggests PD as a contextual factor shaping how individuals respond to organizational expectations, though further research is needed to better understand these effects.

6.2.2 On the Role of Uncertainty Avoidance

Our results did not identify any statistically significant effects for Uncertainty Avoidance (UA), neither as an antecedent nor as a moderator. This outcome suggests that UA may not play a substantial role in the adoption of fairness toolkits. A possible explanation lies in the nature of these technologies: while fairness toolkits are relatively novel and situated within an emerging socio-technical context, they are not inherently associated with high levels of perceived risk or ambiguity. Therefore, individuals high in UA may not view their adoption as threatening or disruptive, which could explain the lack of influence. Nevertheless, the absence of statistical significance should not be overinterpreted and does not imply that UA is irrelevant in all related contexts.

> Further empirical investigations are needed to assess the role of UA more conclusively, particularly in different stages of adoption or in more dynamic and uncertain organizational settings.

6.2.3 On the Role of Long-Term Orientation

Similar to Power Distance, Long-Term Orientation (LT) yielded noteworthy results, with two out of three hypothesized antecedent relationships found to be statistically significant, each with small effect sizes. Specifically, LT positively influenced both Performance Expectancy (PE) and Habit (HB), suggesting that individuals who tend to invest in long-term goals perceive fairness toolkits as valuable assets capable of contributing to future-oriented outcomes. This interpretation is further supported by the total effect analysis, where LT exhibited a significant indirect influence on Behavioral Intention (BI), mediated through PE and HB.

Regarding moderation effects, LT did not exhibit any significant moderating roles within the model. This absence indicates that the influence of LT is likely expressed through direct pathways rather than by altering the strength of other relationships. Such a configuration may also explain the limited influence of LT on Use Behavior (UB), as indirect effects tend to attenuate across multiple mediating steps.

> Managers should consider assigning fairness-related tasks to individuals with high levels of LT, as they are more likely to recognize the strategic value of fairness toolkits and are more inclined to integrate their use into habitual practices.

> Future research could explore the temporal dimension of fairness toolkit adoption more explicitly, for instance by assessing whether Long-Term Orientation correlates with sustained usage over time, institutionalization of fairness practices, or integration into organizational routines.

6.2.4 On the Role of Collectivism

Collectivism (CO) emerged as the most influential cultural variable in our model, with all three antecedent relationships found to be statistically significant, each exhibiting a small effect size, and one significant moderating effect with a medium effect size. Specifically, CO positively influenced Social Influence (SI), Effort Expectancy (EE), and Facilitating Conditions (FC). The effect on SI confirms that collectivist individuals tend to attribute greater importance to peer encouragement and support, which aligns with the broader notion that fairness is a value more salient to those with strong group-oriented orientations. Regarding EE and FC, the significant relationships—when considered alongside the effect on SI—suggest that fairness toolkits are increasingly seen as supported and collaborative tools. As a result, collectivist individuals may perceive them as less cognitively or logistically demanding to adopt.

As a moderator, CO significantly strengthened the relationship between Behavioral Intention (BI) and Use Behavior (UB), aligning it with Power Distance (PD) as one of the only two cultural values exhibiting significant moderation effects. This result further reinforces the interpretation that Collectivism is a distinguishing trait of individuals who are more inclined to promote and implement fairness-oriented practices within software development teams.

However, when examining the total effects on the two main dependent variables (BI and UB), CO did not demonstrate any statistically significant influence. This outcome can be explained by the fact that none of the UTAUT2 constructs directly influenced by CO—namely, SI, EE, and FC—were themselves significant predictors of BI or UB.

> ☞ Despite the lack of direct influence on BI and UB, practitioners can still derive value from identifying individuals with high CO orientations. Given CO's consistent positive influence on several intermediate variables, we recommend prioritizing such individuals for fairness-related initiatives within development teams.

> 💡 Further studies should delve into the group dynamics of fairness technology adoption in collectivist environments, potentially through team-level analyses or ethnographic studies. Understanding how group cohesion, consensus-building, and peer reinforcement shape adoption processes could enrich current models.

6.2.5 A Broader View on the Role of Cultural Values

Our results indicate that cultural values in fairness toolkit adoption do not primarily act as direct drivers of Behavioral Intention. Instead, they shape the pathways through which adoption is enacted in practice. While traditional adoption models assume that individual perceptions directly translate into intention, our findings suggest that intention is mainly driven by instrumental considerations (e.g., Performance Expectancy and Habit), whereas cultural values operate at a different level.

These findings can be more clearly interpreted when situated within the emerging body of work that conceptualizes fairness in software engineering as a socio-technical phenomenon rather than a purely technical property of ML models. Prior studies have shown that fairness-related concerns arise not only from algorithmic choices, but also from the interaction among team roles, development processes, and organizational priorities (Holstein et al. 2019; Ferrara et al. 2024b). Our results are consistent with this perspective: the effects of Power Distance (PD) and Collectivism (CO) on Social Influence, Effort Expectancy, Facilitating Conditions, and on the BI → UB relationship suggest that fairness-tool adoption is influenced by how practitioners position themselves within team structures, authority relations, and collaborative norms (Ferrara et al. 2024b; de Souza Santos et al. 2024). This helps explain why these cultural dimensions do not significantly increase Behavioral Intention directly: fairness adoption decisions appear to be primarily motivated by perceived effectiveness rather than by social or cultural alignment.

Importantly, our study extends the existing socio-technical fairness literature in a distinctive way. While prior research has mainly focused on external factors such as team dynamics, organizational barriers, or workflow constraints, our results show that these mechanisms should also be examined through the lens of individual cultural value orientations. In other words, the same team or organizational setting may be interpreted and enacted differently depending on practitioners' cultural dispositions. For example, individuals with higher Power Distance (PD) may be more likely to translate fairness-related intentions into concrete use when such practices are perceived as aligned with hierarchical expectations, whereas more collectivistic individuals may enact fairness practices more readily when they are embedded in shared team goals. This provides a complementary explanation to prior work: fairness adoption is shaped not only by the structure of the team, but also by how practitioners culturally interpret and respond to that structure.

Long-Term Orientation (LT), in contrast, operates through a more instrumental pathway, influencing Performance Expectancy (PE) and Habit (HB), and thus indirectly supporting adoption. Future-oriented individuals may perceive fairness toolkits as long-term investments contributing to sustainable software quality (De Martino et al. 2025). Uncertainty Avoidance (UA) shows limited relevance, suggesting that fairness tool adoption is not strongly perceived as a risk-related or uncertainty-driven decision.

From a generalizability standpoint, these results should not be interpreted as universally fixed effects of culture, but rather as mechanisms whose manifestation likely depends on the maturity of fairness practices within a given organization. In organizations where fairness is already institutionalized through clear processes and formal tool support, cultural effects may interact differently with social and organizational determinants. In less mature contexts, where fairness remains discretionary and weakly formalized, individual cultural orientations may play a stronger role in determining whether intentions are translated into actual use.

A Broader View.

Taken together, our findings indicate that cultural values primarily act as socio-technical conditioners of adoption behavior, rather than as direct motivators of intention. The novelty of our study lies in bridging team-level and individual-level perspectives: rather than treating team dynamics and individual values as separate explanations, we show that cultural values partly explain how practitioners perceive team norms, authority structures, and collaborative support when engaging with fairness toolkits. This extends prior socio-technical fairness research by adding an individual-level cultural mechanism as a previously underexplored explanatory layer.

6.3 Limitations and Threats to Validity

As outlined earlier, this study relied primarily on quantitative data supported by statistical analyses. In line with standard guidelines for empirical SE, we address potential threats to validity by considering four key dimensions defined by Wohlin et al. (2012): statistical conclusion validity, internal validity, construct validity, and external validity. Additionally, given the involvement of human participants, we also examine the reliability of responses as a separate concern.

6.3.1 Conclusion Validity

Conclusion validity concerns the extent to which the identified relationships between independent and dependent variables are statistically reliable (Wohlin et al. 2012). Threats in this domain are primarily associated with the appropriateness and rigor of the statistical techniques employed. This study applied Partial Least Squares Structural Equation Modeling (PLS-SEM), a method recognized for its robustness across various research contexts. Methodological rigor was maintained by strictly following the established guidelines of Hair Junior et al. (2014), and analyses were conducted using SmartPLS, a tool widely adopted in empirical studies and cited in over 1000 peer-reviewed publications, thereby enhancing the credibility and reproducibility of the findings.

> Future work could replicate this model using alternative statistical techniques, such as Covariance-Based SEM or Bayesian approaches, to compare stability and robustness across different analytical paradigms. Such triangulation would enhance the reliability of fairness toolkit adoption research.

6.3.2 Internal Validity

Internal validity refers to the degree to which observed effects can be attributed to the modeled constructs rather than to confounding variables (Wohlin et al. 2012). To address this concern, the study was grounded in theoretically validated frameworks from the technology adoption literature. In addition, participant selection was governed by rigorous inclusion criteria to ensure that the sample accurately represented the intended population—namely, software practitioners with experience in using fairness toolkits—while also maintaining sufficient heterogeneity to reflect realistic variation across individuals.

> To further address internal validity, future research should consider longitudinal designs or experimental manipulations to better infer causal pathways between cultural values, adoption constructs, and fairness-related behaviors.

6.3.3 Construct Validity

Construct validity deals with the degree to which the study accurately measures the theoretical constructs of interest (Wohlin et al. 2012). To ensure high construct validity, all latent variables were operationalized using established and validated measurement instruments drawn from prior research (Venkatesh et al. 2012; Yoo et al. 2011). The questionnaire followed state-of-the-art survey design practices, including randomized question ordering and attention-check mechanisms to minimize response bias and increase internal consistency (Kitchenham and Pfleeger 2008; Ralph et al. 2020; Danilova et al. 2021). Furthermore, established tools and recruitment strategies were adopted to strengthen the quality of measurement and respondent engagement (Eyal et al. 2021; Douglas et al. 2023; Alami et al. 2024; Russo 2022).

> 💡 Future studies should explore the development of domain-specific measurement instruments for fairness technology adoption, including reflective and formative indicators tailored to socio-technical contexts. Qualitative pre-studies or cognitive interviews could help refine construct definitions.

6.3.4 External Validity

External validity refers to the extent to which the study's results are generalizable beyond the sampled participants (Wohlin et al. 2012). To enhance generalizability, participants were recruited through Prolific using stringent pre-screening filters, ensuring alignment with the study's target population. Additionally, sample size requirements were calculated using G*Power (Faul et al. 2009), ensuring sufficient statistical power to detect meaningful effects. Nevertheless, generalizability remains a known challenge, particularly in fast-evolving domains such as software development and AI adoption. As technologies and practices evolve, replication in future contexts will be necessary to validate the stability of these findings over time.

> 💡 Future research could test the generalizability of these findings across industries, cultural regions, and software domains, using stratified sampling and cross-country comparative designs to better understand context-specific adoption dynamics.

A second potential concern relates to the partial overlap between the participant pool used in this study and that of a previous, related study (Voria et al. 2025a). In principle, repeated participation in studies examining the same constructs could raise questions about sensitization effects—whereby prior exposure influences subsequent responses—or about the limited generalizability of findings derived from a narrow, recurring sample. However, in the present study this concern is mitigated by the fact that all data were collected at a single point in time, with no longitudinal re-measurement of the same constructs. As such, the design does not entail the temporal dependencies characteristic of a panel study, and systematic response biases attributable to prior participation are unlikely.

6.3.5 Participant Reliability

Participant reliability was carefully considered given its critical role in ensuring data quality in survey-based research. To mitigate the risks associated with inattentive or unqualified respondents, multiple quality control measures were employed. Recruitment was conducted via Prolific, applying filters such as a 100% prior approval rate in past studies to select highly reliable participants (Eyal et al. 2021; Douglas et al. 2023). A pre-screening survey was used to verify participants' professional background and experience with LLMs in software development. In addition, validated programming knowledge questions (Danilova et al. 2021) were included to ensure a baseline level of technical competence. Attention-check items were embedded within the survey to detect inattentive behavior, and responses failing these checks were excluded from the analysis. Finally, an iterative pilot study was conducted to validate the clarity, duration, and structure of the instrument (Kitchenham and Pfleeger 2008), ensuring the reliability of data collection procedures.

> Further research should assess the influence of participant reliability on adoption modeling outcomes by comparing results across platforms (e.g., Prolific, MTurk, industry partners) and by analyzing behavioral traces or triangulated data.

7 Conclusion

This study revealed that both technological and cultural factors shape the adoption of fairness toolkits in software development. Performance expectancy and Habit emerged as the most significant predictors of practitioners' intention to adopt such tools, while Habit and Behavioral Intention were key determinants of actual use. Among cultural values, Long-Term Orientation and Power Distance showed direct positive effects on adoption, whereas Collectivism moderated the relationship between Behavioral Intention and use behavior. These results suggest that developers' cultural orientations play a significant role in the uptake of fairness-focused practices.

The implications are twofold. For practitioners, assessing individual cultural predispositions can inform the composition of development teams to foster fairness-oriented behaviors. For researchers, the findings open pathways for further investigation into the interplay between cultural traits and ethical software practices, expanding the socio-technical lens in fairness research.

The contributions of this work to the SE field include:

- The first empirical integration of individual-level cultural values into a model of fairness toolkit adoption.
- A validated structural model that explains how cultural and technological factors jointly influence fairness practices.
- Practical recommendations for fairness-aware team composition and tool adoption strategies.
- Methodological guidance for future studies examining cultural factors in SE technology adoption.

These contributions advance the understanding of how fairness-related tools can be more effectively integrated into development workflows by accounting for the socio-cultural context in which developers operate.

Acknowledgements The authors thank the questionnaire participants for their valuable responses and for providing the essential data required to conduct this research. The authors would also like to thank the reviewers for their constructive comments, which have helped to improve the quality of the work.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Stefano Lambiase, Gianmario Voria, and Maria Concetta Schiavone. The first draft of the manuscript was written by Stefano Lambiase and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Aalborg University. This work has been partially supported by the European Union — NextGenerationEU through the Italian Ministry of University and Research, Projects PRIN 2022 PNRR "FRINGE: context-aware FaiRness engineerING in complex software systEms" (grant n. P2022553SL, CUP: D53D23017340001). This work has been partially supported by the EMELIOT national research project, which has been funded by the MUR under the PRIN 2020 program (Contract 2020W3A5FY).

Data Availability The data that support the findings of this study are openly available in our online appendix (Lambiase et al. 2026) and at the following .

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics approval At the time of data collection, no formal instrument—such as an ethics review board—was available or required in the authors' country to obtain ethical approval. Nevertheless, the study was designed and conducted with careful attention to ethical principles, particularly those relevant to survey-based research (Hall and Flynn 2001). Participants' privacy was prioritized throughout the process. A detailed privacy statement was provided, and no sensitive business or personally identifiable information was requested. Participants were explicitly informed that their data would be used exclusively for research purposes in line with the study's objectives. Participation was voluntary, limited to individuals over the age of 18, and required informed consent. Respondents were also made aware that they could withdraw at any time without penalty. Furthermore, participants were transparently informed that their anonymized responses would be published and permanently stored in the online appendix of this paper (Lambiase et al. 2026).

Informed Consent Study participants received informed consent to participation and data were treated anonymously.

Clinical Trial Number Not applicable.

References

Abbasi MS, Tarhini A, Elyas T, Shah F (2015) Impact of individualism and collectivism over the individual's technology acceptance behaviour: A multi-group analysis between pakistan and turkey. J Enterp Inf Manag 28(6):747–768

Abufardeh S, Magel K (2010) The impact of global software cultural and linguistic aspects on global software development process (gsd): Issues and challenges. In: 4th International conference on new trends in information science and service science, IEEE, pp 133–138
Aguinis H, Beaty JC, Boik RJ, Pierce CA (2005) Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. J Appl Psychol 90(1):94
Alami A, Zahedi M, Ernst N (2024) Are you a real software engineer? best practices in online recruitment for software engineering studies. In: Proceedings of the 1st IEEE/ACM international workshop on methodological issues with empirical studies in software engineering, pp 52–57
Andrews D, Nonnecke B, Preece J (2007) Conducting research on the internet:: Online survey design, development and implementation guidelines
Ayling J, Chapman A (2022) Putting ai ethics to work: are the tools fit for purpose? AI and Ethics 2. https://doi.org/10.1007/s43681-021-00084-x
Balayn A, Yurrita M, Yang J, Gadiraju U (2023) "fairness toolkits, a checkbox culture?" on the factors that fragment developer practices in handling algorithmic harms. In: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery, New York, NY, USA, AIES '23, p 482–495. https://doi.org/10.1145/3600211.3604674
Baptista G, Oliveira T (2015) Understanding mobile banking: The unified theory of acceptance and use of technology combined with cultural moderators. Comput Hum Behav 50:418–430
Bellamy RK, Dey K, Hind M, Hoffman SC, Houde S, Kannan K, Lohia P, Martino J, Mehta S, Mojsilović A et al (2019) Ai fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM J Res Dev 63(4/5):4–1
Borchers G (2003) The software engineering impacts of cultural factors on multi-cultural software development teams. In: 25th International Conference on Software Engineering, 2003. Proceedings., IEEE, pp 540–545
Brun Y, Meliou A (2018) Software fairness. In: Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering, pp 754–759
Calmon F, Wei D, Vinzamuri B, Natesan Ramamurthy K, Varshney KR (2017) Optimized pre-processing for discrimination prevention. Adv Neural Inf Process Syst 30
Cannavale A, Voria G, Scognamiglio A, Giordano G, Catolino G, Palomba F (2025) Fairness set and forgotten: Mining fairness toolkit usage in open-source machine learning projects. Inf Softw Technol, p 107957
Casey V (2011) Imparting the importance of culture to global software development. ACM inroads 1(3):51–57
Chakraborty J, Majumder S, Yu Z, Menzies T (2020) Fairway: a way to build fair ml software. In: Proceedings of the 28th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering, pp 654–665
Chakraborty J, Majumder S, Menzies T (2021) Bias in machine learning software: why? how? what to do? In: Proceedings of the 29th ACM joint meeting on European Software engineering conference and symposium on the foundations of software engineering, pp 429–440
Che FN, Vajjhala NR, Strang KD (2021) Perceived student learning, attitudes and experiential learning in software engineering education. In: 2021 IEEE World Conference on Engineering Education (EDUNINE), IEEE, pp 1–6
Chen Z, Zhang JM, Sarro F, Harman M (2024) Fairness improvement with multiple protected attributes: How far are we? In: Proceedings of the IEEE/ACM 46th international conference on software engineering, pp 1–13
Childers TL, Carr CL, Peck J, Carson S (2001) Hedonic and utilitarian motivations for online retail shopping behavior. J Retail 77(4):511–535
Chin WW et al (1998) The partial least squares approach to structural equation modeling. Modern methods for business research 295(2):295–336
Choudhuri R, Trinkenreich B, Pandita R, Kalliamvakou E, Steinmacher I, Gerosa M, Sanchez C, Sarma A (2024) What guides our choices? modeling developers' trust and behavioral intentions towards genai. arXiv:2409.04099
Cohen J (2013) Statistical power analysis for the behavioral sciences. routledge
Danilova A, Naiakshina A, Horstmann S, Smith M (2021) Do you really code? designing and evaluating screening questions for online surveys with programmers. In: 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), IEEE, pp 537–548
Davis FD, Bagozzi R, Warshaw P (1989) Technology acceptance model. J Manag Sci 35(8):982–1003
De Martino V, Voria G, Troiano C, Catolino G, Palomba F (2025) Examining the impact of bias mitigation algorithms on the sustainability of ml-enabled systems: A benchmark study. J Syst Softw 230:112458

Deng WH, Nagireddy M, Lee MSA, Singh J, Wu ZS, Holstein K, Zhu H (2022) Exploring how machine learning practitioners (try to) use fairness toolkits. In: 2022 ACM Conference on Fairness, Accountability, and Transparency, ACM, FAccT '22. https://doi.org/10.1145/3531146.3533113

Dolata M, Feuerriegel S, Schwabe G (2021) A sociotechnical view of algorithmic fairness. Inf Syst J 32:754–818. https://doi.org/10.1111/isj.12370

Douglas BD, Ewell PJ, Brauer M (2023) Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. PLoS ONE 18(3):e0279720

Eyal P, David R, Andrew G, Zak E, Ekaterina D (2021) Data quality of platforms and panels for online behavioral research. Behav Res Methods, pp 1–20

Faqih KM, Jaradat MIRM (2015) Assessing the moderating effect of gender differences and individualism-collectivism at individual-level on the adoption of mobile commerce technology: Tam3 perspective. J Retail Consum Serv 22:37–52

Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. Behavior research methods 41(4):1149–1160

Ferrara C, Casillo F, Gravino C, De Lucia A, Palomba F (2024a) Refair: Toward a context-aware recommender for fairness requirements engineering. In: Proceedings of the IEEE/ACM 46th international conference on software engineering, pp 1–12

Ferrara C, Sellitto G, Ferrucci F, Palomba F, De Lucia A (2024b) Fairness-aware machine learning engineering: how far are we? Empir Softw Eng 29(1):9

Flight RL, Allaway AW, Kim WM, D'Souza G (2011) A study of perceived innovation characteristics across cultures and stages of diffusion. Journal of Marketing Theory and Practice 19(1):109–126

Ford DP, Connelly CE, Meister DB (2003) Information systems research and hofstede's culture's consequences: an uneasy and incomplete partnership. IEEE Trans Eng Manage 50(1):8–25

Galhotra S, Brun Y, Meliou A (2017) Fairness testing: testing software for discrimination. In: Proceedings of the 2017 11th Joint meeting on foundations of software engineering, pp 498–510

Gelfand M, Higgins M, Nishii L, Raver JL, Alexandria D, Murakami F, Yamaguchi S, Toyama M (2002) Culture and egocentric perceptions of fairness in conflict and negotiation. J Appl Psychol 87(5):833–45. https://doi.org/10.1037/0021-9010.87.5.833

Hair Junior JF, Hult GTM, Ringle CM, Sarstedt M (2014) A primer on partial least squares structural equation modeling (pls-sem)

Hall T, Flynn V (2001) Ethical issues in software engineering research: a survey of current practice. Empir Softw Eng 6(4):305–317

Henseler J, Ringle CM, Sarstedt M (2015) A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 43:115–135

Hoda R (2021) Socio-technical grounded theory for software engineering. IEEE Trans Software Eng 48(10):3808–3832

Hoehle H, Zhang X, Venkatesh V (2015) An espoused cultural perspective to understand continued intention to use mobile applications: a four-country study of mobile social media application usability. Eur J Inf Syst 24(3):337–359

Hofstede G (1980) Culture and organizations. International studies of management & organization 10(4):15–41

Hofstede G (2011) Dimensionalizing cultures: The hofstede model in context. Online readings in psychology and culture 2(1):8

Holstein K, Vaughan J, Daumé I H, Dudík M, Wallach H (2019) Improving fairness in machine learning systems: What do industry practitioners need? Association for Computing Machinery. https://doi.org/10.1145/3290605.3300830

Hort M, Chen Z, Zhang JM, Harman M, Sarro F (2024) Bias mitigation for machine learning classifiers: A comprehensive survey. ACM Journal on Responsible Computing 1(2):1–52

House RJ, Hanges PJ, Javidan M, Dorfman PW, Gupta V (2004) Culture, leadership, and organizations: The GLOBE study of 62 societies. Sage publications

Javed I, Janjua UI, Muhammad Abdulhamid S, Madni TM, Akhunzada A (2023) The impact of mitigation strategies for socio-cultural distance issues in gsd: An empirical study. IEEE Access 11:99499–99518. https://doi.org/10.1109/ACCESS.2023.3300836

Johnson B, Pidd H (2009) 'gay writing' falls foul of amazon sales ranking system. The Guardian https://www.theguardian.com/culture/2009/apr/13/amazon-gay-writers

Kamiran F, Calders T (2012) Data preprocessing techniques for classification without discrimination. Knowl Inf Syst 33(1):1–33

Kamishima T, Akaho S, Asoh H, Sakuma J (2012) Fairness-aware classifier with prejudice remover regularizer. In: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012. Proceedings, Part II 23, Springer, pp 35–50

Kenny DA, Judd CM (2019) The unappreciated heterogeneity of effect sizes: Implications for power, precision, planning of research, and replication. Psychol Methods 24(5):578

Khattak KN, Qayyum F, Naqvi SSA, Mehmood A, Kim J (2023) A systematic framework for addressing critical challenges in adopting devops culture in software development: A pls-sem perspective. IEEE Access 11:120137–120156

Kim TY, Leung K (2007) Forming and reacting to overall fairness: A cross-cultural comparison. Organ Behav Hum Decis Process 104:83–95. https://doi.org/10.1016/J.OBHDP.2007.01.004

Kitchenham BA, Pfleeger SL (2008) Personal opinion surveys. In: Guide to advanced empirical software engineering, Springer, pp 63–92

Lai C, Wang Q, Li X, Hu X (2016) The influence of individual espoused cultural values on self-directed use of technology for language learning beyond the classroom. Comput Hum Behav 62:676–688

Lambiase S (2024) Cultural and socio-technical aspects in software development. In: Proceedings of the 28th international conference on evaluation and assessment in software engineering, pp 482–487

Lambiase S, De Lucia A (2026) Quantifying adoption: A sem study of quantum software technology in software development. Empir Softw Eng 31(3):69

Lambiase S, Catolino G, Pecorelli F, Tamburri DA, Palomba F, Van Den Heuvel WJ, Ferrucci F (2022a) “there and back again?” on the influence of software community dispersion over productivity. In: 2022 48th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp 177–184. https://doi.org/10.1109/SEAA56994.2022.00035

Lambiase S, Catolino G, Tamburri DA, Serebrenik A, Palomba F, Ferrucci F (2022b) Good fences make good neighbours? on the impact of cultural and geographical dispersion on community smells. In: Proceedings of the 2022 ACM/IEEE 44th international conference on software engineering: software engineering in society, pp 67–78

Lambiase S, Catolino G, Torre M, Tamburri DA, Palomba F, Ferrucci F (2022c) Fright not and be dispersed! evaluating cultural dimensions versus software communication and collaboration activities. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4210197, https://ssrn.com/abstract=4210197

Lambiase S, Catolino G, Palomba F, Ferrucci F, Russo D (2024a) Investigating the role of cultural values in adopting large language models for software engineering. arXiv:2409.05055

Lambiase S, Catolino G, Pecorelli F, Tamburri DA, Palomba F, van den Heuvel WJ, Ferrucci F (2024b) An empirical investigation into the influence of software communities’ cultural and geographical dispersion on productivity. J Syst Softw 208:111878

Lambiase S, Catolino G, Palomba F, Ferrucci F, Russo D (2025) Exploring individual factors in the adoption of llms for specific software engineering purposes. arXiv preprint arXiv:2504.02553

Lambiase S, Voria G, Schiavone MC, Catolino G, Palomba F (2026) From values to adoption: On the role of individual cultural values on fairness toolkit adoption in software development — online appendix. https://doi.org/10.6084/m9.figshare.28938572

Lee MSA, Singh J (2021) The landscape and gaps in open source fairness toolkits. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, CHI ’21. https://doi.org/10.1145/3411764.3445261

Lee SG, Trimi S, Kim C (2013) Innovation and imitation effects’ dynamics in technology adoption. Ind Manag Data Syst 113(6):772–799

Lee H, Chung N, Jung T (2015) Examining the cultural differences in acceptance of mobile augmented reality: Comparison of south korea and ireland. In: Information and Communication Technologies in Tourism 2015: Proceedings of the International Conference in Lugano, Switzerland, February 3-6, 2015, Springer, pp 477–491

Lin HC (2014) An investigation of the effects of cultural differences on physicians’ perceptions of information technology acceptance as they relate to knowledge management systems. Comput Hum Behav 38:368–380

Lund DJ, Scheer L, Kozlenkova IV (2013) Culture’s impact on the importance of fairness in interorganizational relationships. J Int Mark 21:21–43. https://doi.org/10.1509/jim.13.0020

Mac R (2021) Facebook apologizes after a.i. puts ’primates’ label on video of black men. The New York Times https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html

Majumder S, Chakraborty J, Bai GR, Stolee KT, Menzies T (2023) Fair enough: Searching for sufficient measures of fairness. ACM Trans Softw Eng Methodol 32(6). https://doi.org/10.1145/3585006

Markus HR, Kitayama S (2014) Culture and the self: Implications for cognition, emotion, and motivation. In: College student development and academic life, Routledge, pp 264–293

Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2021) A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR) 54(6):1–35

Microsoft, contributors (2019) “fairlearn”. https://fairlearn.github.io/

Miller S (2019) Machine learning, ethics and law. Australas J Inf Syst 23:1–13

Nistor N, Lerche T, Weinberger A, Ceobanu C, Heymann O (2014) Towards the integration of culture into the unified theory of acceptance and use of technology. Br J Edu Technol 45(1):36–55

Pagano TP, Loureiro RB, Lisboa FV, Peixoto RM, Guimarães GA, Cruz GO, Araujo MM, Santos LL, Cruz MA, Oliveira EL et al (2023) Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. Big data and cognitive computing 7(1):15

Parziale A, Voria G, Giordano G, Catolino G, Robles G, Palomba F (2025) Fairness on a budget, across the board: A cost-effective evaluation of fairness-aware practices across contexts, tasks, and sensitive attributes. Inf Softw Technol, p 107858

Pessach D, Shmueli E (2022) A review on fairness in machine learning. ACM Computing Surveys (CSUR) 55(3):1–44

Raithel S, Sarstedt M, Scharf S, Schwaiger M (2012) On the value relevance of customer satisfaction. multiple drivers and multiple markets. J Acad Mark Sci 40:509–525

Rakova B, Yang J, Cramer H, Chowdhury R (2021) Where responsible ai meets reality: Practitioner perspectives on enablers for shifting organizational practices. Proc ACM Hum-Comput Interact 5(CSCW1). https://doi.org/10.1145/3449081

Ralph P, Ali Nb, Baltes S, Bianculli D, Diaz J, Dittrich Y, Ernst N, Felderer M, Feldt R, Filieri A et al (2020) Empirical standards for software engineering research. arXiv preprint arXiv:2010.03525

Ribiere VM, Haddad M, Vande Wiele P (2010) The impact of national culture traits on the usage of web 2.0 technologies. Vine 40(3/4):334–361

Ringle CM, Wende S, Becker JM (2024) Smartpls 4. https://www.smartpls.com/

Rogers EM (2003) Diffusion of Innovations, 5th edn. Simon and Schuster

Russo D (2022) Recruiting software engineers on prolific. arXiv preprint arXiv:2203.14695

Russo D (2024) Navigating the complexity of generative ai adoption in software engineering. ACM Transactions on Software Engineering and Methodology 33(5)

Russo D, Stol KJ (2021) Pls-sem for software engineering research: An introduction and survey. ACM Computing Surveys (CSUR) 54(4):1–38

Sesari E, Sarro F, Rastogi A (2024) Understanding fairness in software engineering: Insights from stack exchange. arXiv preprint arXiv:2402.19038

Sharma S, Zhang Y, Ríos Aliaga JM, Bouneffouf D, Muthusamy V, Varshney KR (2020) Data augmentation for discrimination prevention and bias disambiguation. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp 358–364

Shiu E, Walsh G, Hassan LM, Parry S (2015) The direct and moderating influences of individual-level cultural values within web engagement: A multi-country analysis of a public information website. J Bus Res 68(3):534–541

Shmueli G, Ray S, Estrada JMV, Chatla SB (2016) The elephant in the room: Predictive performance of pls models. J Bus Res 69(10):4552–4564

Shmueli G, Sarstedt M, Hair JF, Cheah JH, Ting H, Vaithilingam S, Ringle CM (2019) Predictive model assessment in pls-sem: guidelines for using plspredict. Eur J Mark 53(11):2322–2347

Sotolani R, Freire S, Fronchetti F, Santos RDS, Spinola R (2026) Exposing hidden bias: A study of fairness debt in gray literature. J Syst Softw, p 112866. https://doi.org/10.1016/j.jss.2026.112866, https://www.sciencedirect.com/science/article/pii/S0164121226000993

de Souza SR, Fronchetti F, Freire S, Spínola RO (2024) Software fairness debt: Building a research agenda for addressing bias in ai systems. ACM Transactions on Software Engineering and Methodology 34:1–21. https://doi.org/10.1145/3709357

Srite M, Karahanna E (2006) The role of espoused national cultural values in technology acceptance. MIS quarterly, pp 679–704

Starke C, Baleis J, Keller B, Marcinkowski F (2021) Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. 2103.12016

Storey MA, Ernst NA, Williams C, Kalliamvakou E (2020) The who, what, how of software engineering research: a socio-technical framework. Empir Softw Eng 25:4097–4129

Straub D, Keil M, Brenner W (1997) Testing the technology acceptance model across cultures: A three country study. Information & management 33(1):1–11

Tamilmani K, Rana NP, Wamba SF, Dwivedi R (2021) The extended unified theory of acceptance and use of technology (utaut2): A systematic literature review and theory evaluation. Int J Inf Manage 57:102269

Trinkenreich B, Stol KJ, Steinmacher I, Gerosa MA, Sarma A, Lara M, Feathers M, Ross N, Bishop K (2023) A model for understanding and reducing developer burnout. In: 2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), pp 48–60. https://doi.org/10.1109/ICSE-SEIP58684.2023.00010

Trinkenreich B, Gerosa MA, Steinmacher I (2024) Unraveling the drivers of sense of belonging in software delivery teams: Insights from a large-scale survey. In: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering, Association for Computing Machinery, New York, NY, USA, ICSE '24. https://doi.org/10.1145/3597503.3639119

Udeshi S, Arora P, Chattopadhyay S (2018) Automated directed fairness testing. In: Proceedings of the 33rd ACM/IEEE international conference on automated software engineering, pp 98–108

Vasudevan S, Kenthapadi K (2020) Lift: A scalable framework for measuring fairness in ml applications. In: Proceedings of the 29th ACM international conference on information & knowledge management, pp 2773–2780

Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: Toward a unified view. MIS Quarterly 27(3):425–478. http://www.jstor.org/stable/30036540

Venkatesh V, Thong JYL, Xu X (2012) Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly 36(1):157–178. http://www.jstor.org/stable/41410412

Verma S, Rubin J (2018) Fairness definitions explained. In: 2018 ieee/acm international workshop on software fairness (fairware), IEEE, pp 1–7

Voria G, Lambiase S, Schiavone MC, Catolino G, Palomba F (2025a) From expectation to habit: Why do software practitioners adopt fairness toolkits? In: 2025 IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), IEEE, pp 94–105

Voria G, Scala B, Todisco L, Venditto C, Giordano G, Catolino G, Palomba F (2025b) Fair and square? evaluating fairness of llm-generated synthetic datasets. Inf Softw Technol, p 107980

Voria G, Sellitto G, Ferrara C, Abate F, De Lucia A, Ferrucci F, Catolino G, Palomba F (2025c) A catalog of fairness-aware practices in machine learning engineering. In: Proceedings of the 2025 29th international conference on evaluation and assessment in software engineering companion, pp 72–81

Voria G, Sellitto G, Ferrara C, Abate F, De Lucia A, Ferrucci F, Catolino G, Palomba F (2025d) Fairness-aware practices from developers' perspective: A survey. Inf Softw Technol, p 107710

Warmerdam VD (2020) scikit-fairness: Fairness metrics, transformers, and models for scikit-learn. https://github.com/koaning/scikit-fairness, gitHub repository. Accessed 12 March 2026

Wei M, Zhou Z (2022) Ai ethics issues in real world: Evidence from ai incident database. 2206.07635

Wexler J, Pushkarna M, Bolukbasi T, Wattenberg M, Viégas F, Wilson J (2019) The what-if tool: Interactive probing of machine learning models. IEEE Trans Visual Comput Graphics 26(1):56–65

Wohlin C, Runeson P, Höst M, Ohlsson MC, Regnell B, Wesslén A et al (2012) Experimentation in software engineering, vol 236. Springer

Yasin A, Fatima R, Ali Khan J, Liu L, Ali R, Wang J (2024) Counteracting sociocultural barriers in global software engineering using group activities. J Softw Evol Process 36(5):e2587

Yoo B, Donthu N, Lenartowicz T (2011) Measuring hofstede's five dimensions of cultural values at the individual level: Development and validation of cvscale. J Int Consum Mark 23(3–4):193–210

Zhang JM, Harman M (2021) "ignorance and prejudice" in software fairness. In: 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), IEEE, pp 1436–1447

Zhou Z, Jin XL, Fang Y, Vogel D (2015) Toward a theory of perceived benefits, affective commitment, and continuance intention in social virtual worlds: cultural values (indulgence and individualism) matter. Eur J Inf Syst 24(3):247–261

Zhang BH, Lemoine B, Mitchell M (2018) Mitigating unwanted biases with adversarial learning. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp 335–340

Authors and Affiliations

Stefano Lambiase[1] · **Gianmario Voria**[2] · **Maria Concetta Schiavone**[2] · **Gemma Catolino**[2] · **Fabio Palomba**[2]

✉ Stefano Lambiase
stla@cs.aau.dk

Gianmario Voria
gvoria@unisa.it

Maria Concetta Schiavone
mariaconcetta@studenti.unisa.it

Gemma Catolino
gcatolino@unisa.it

Fabio Palomba
fpalomba@unisa.it

1 Aalborg University, Aalborg University in Copenhagen, A. C. Meyers Vænge 15, 2450 København, Denmark

2 University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy