Aalborg Universitet

AALBORG UNIVERSITY

Investigating the Role of Cultural Values in Adopting Large Language Models for Software Engineering

Lambiase, Stefano; Catolino, Gemma; Palomba, Fabio; Ferrucci, Filomena; Russo, Daniel

Published in:
ACM Transactions on Software Engineering and Methodology

DOI (link to publication from Publisher):
10.48550/arXiv.2409.05055
10.1145/3725529

Creative Commons License
CC BY 4.0

Publication date:
2026

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Lambiase, S., Catolino, G., Palomba, F., Ferrucci, F., & Russo, D. (2026). Investigating the Role of Cultural Values in Adopting Large Language Models for Software Engineering. *ACM Transactions on Software Engineering and Methodology*, *35*(1), Article 23. https://doi.org/10.48550/arXiv.2409.05055, https://doi.org/10.1145/3725529

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: 11 27, 2026

Check for updates

ACM DIGITAL LIBRARY | Association for Computing Machinery | acm open

Latest updates: https://dl.acm.org/doi/10.1145/3725529

PDF Download
3725529.pdf
13 January 2026
Total Citations: 2
Total Downloads: 678

Published: 12 December 2025
Online AM: 21 March 2025
Accepted: 17 March 2025
Revised: 17 March 2025
Received: 24 August 2024

Citation in BibTeX format

RESEARCH-ARTICLE

Investigating the Role of Cultural Values in Adopting Large Language Models for Software Engineering

STEFANO LAMBIASE, University of Salerno, Salerno, SA, Italy

GEMMA CATOLINO, University of Salerno, Salerno, SA, Italy

FABIO PALOMBA, University of Salerno, Salerno, SA, Italy

FILOMENA FERRUCCI, University of Salerno, Salerno, SA, Italy

DANIEL RUSSO, Aalborg University, Aalborg, Nordjylland, Denmark

Open Access Support provided by:

University of Salerno

Aalborg University

Investigating the Role of Cultural Values in Adopting Large Language Models for Software Engineering

STEFANO LAMBIASE, GEMMA CATOLINO, FABIO PALOMBA, and FILOMENA FERRUCCI, University of Salerno, Fisciano, Italy
DANIEL RUSSO, Aalborg University Copenhagen, Copenhagen, Denmark

As a socio-technical activity, software development involves the close interconnection of people and technology. The integration of Large Language Models (LLMs) into this process exemplifies the socio-technical nature of software development. Although LLMs influence the development process, software development remains fundamentally human-centric, necessitating an investigation of the human factors in this adoption. Thus, with this study we explore the factors influencing the adoption of LLMs in software development, focusing on the role of professionals' cultural values. Guided by the Unified Theory of Acceptance and Use of Technology (UTAUT2) and Hofstede's cultural dimensions, we hypothesized that cultural values moderate the relationships within the UTAUT2 framework. Using Partial Least Squares-Structural Equation Modelling and data from 188 software engineers, we found that habit and performance expectancy are the primary drivers of LLM adoption, while cultural values do not significantly moderate this process. These findings suggest that, by highlighting how LLMs can boost performance and efficiency, organizations can encourage their use, no matter the cultural differences. Practical steps include offering training programs to demonstrate LLM benefits, creating a supportive environment for regular use, and continuously tracking and sharing performance improvements from using LLMs.

CCS Concepts: • **Social and professional topics** → **Computing industry**; **Cultural characteristics**; **Management of computing and information systems**; **Project and people management**;

Additional Key Words and Phrases: UTAUT2;, Culture;, Hofstede;, Generative AI;, LLM;, Empirical Software Engineering

ACM Reference format:
Stefano Lambiase, Gemma Catolino, Fabio Palomba, Filomena Ferrucci, and Daniel Russo. 2025. Investigating the Role of Cultural Values in Adopting Large Language Models for Software Engineering. *ACM Trans. Softw. Eng. Methodol.* 35, 1, Article 23 (December 2025), 43 pages.
https://doi.org/10.1145/3725529

This work has been partially supported by the European Union through the Italian Ministry of University and Research, Project PRIN 2022 "*QUAL-AI*: Continuous Quality Improvement of AI-based Systems" (Grant No. 2022B3BP5S, CUP: H53D23003510006).

Authors' Contact Information: Stefano Lambiase (corresponding author), University of Salerno, Fisciano, Italy; e-mail: slambiase@unisa.it; Gemma Catolino, University of Salerno, Fisciano, Italy; e-mail: gcatolino@unisa.it; Fabio Palomba, University of Salerno, Fisciano, Italy; e-mail: fpalomba@unisa.it; Filomena Ferrucci, University of Salerno, Fisciano, Italy; e-mail: fferrucci@unisa.it; Daniel Russo, Aalborg University Copenhagen, Copenhagen, Denmark; e-mail: daniel.russo@cs.aau.dk.



This work is licensed under Creative Commons Attribution International 4.0.

© 2025 Copyright held by the owner/author(s).
ACM 1557-7392/2025/12-ART23
https://doi.org/10.1145/3725529

1 Introduction

> "Technology is nothing. What's important is that you have faith in people, that they're basically good and smart—and if you give them tools, they'll do wonderful things with them. Tools are just tools. They either work, or they don't work."
>
> —*Steve Jobs*

Generative AI has become increasingly pervasive in individuals' daily lives. In particular, the advent of **Large Language Models (LLMs)** has significantly disrupted various activities, notably transforming software development by automating code generation [6, 29, 43, 44, 54, 57, 90], assisting in everyday activities [8, 46, 69, 92], and enhancing code comprehension [9, 48]. In fact, a large body of knowledge is starting to grow in investigating the use of LLMs for supporting software development and engineering activities. For such a reason, the research community has taken a keen interest in this impact, reporting that professionals are already leveraging LLMs in multiple ways, including generating source code, creating entire software modules, providing decision support, and identifying problems to enhance maintenance activities [3, 23, 46, 48, 54, 71].

Despite the integration of LLMs into the software development process, this field remains (still) inherently socio-technical. It involves collaboration among individuals with diverse cultural backgrounds, skill sets, and ways of thinking, all working together and with technology to achieve a common objective [11, 14, 35, 58, 66]. At present, practitioners view LLMs as companions in their daily work activities; they leverage the capabilities of LLMs in various ways and to different extents, with these variations being influenced by a wide range of personal preferences and social factors [48, 54, 71]. Given the transformative potential of LLMs in software engineering, it is crucial to better understand their implications on both the technical and social aspects of development. As LLMs become more integrated into the workflow, they have the potential to reshape not only the tools and practices of software engineering but also the dynamics of team collaboration, decision-making processes, and the overall role of engineers. Understanding the factors that led to the adoption of LLMs is essential for optimizing their use and mitigating any potential drawbacks.

Among the human factors influencing individual choices, cultural values stand out as a significant determinant [7, 49, 61, 80] whose potential influence on LLMs adoption remains unexplored. Culture encompasses shared motives, values, beliefs, identities, and interpretations of significant events that arise from everyday experiences within groups and underlie human attitudes and behaviors [15, 37, 39, 40, 55]. Researchers in both software engineering and management have demonstrated that the values deriving from the cultural background—from now on, *cultural values*—can influence a wide range of variables, including individual personality traits and the choice to adopt a specific technology [7, 49, 61, 80]. As Im et al. [42] claim, technology adoption is as much a cultural issue as a rational decision-making process. Given that the adoption of LLMs, like other technologies, is influenced by both technical and social aspects, it is reasonable to assume that cultural background plays a role in this process. Thus, exploring how cultural factors impact the acceptance and use of LLMs could provide valuable insights, ultimately facilitating their seamless integration into the software development industry. Such a smooth integration ensures that LLMs are effectively leveraged, enhancing productivity, fostering innovation, and minimizing resistance or challenges that could arise from cultural misalignment or management misunderstandings.

Starting from what we mentioned above, the present research started with the following research question:

> **Research Question.** *What is the role of individuals' cultural values in adopting LLMs for supporting software engineering activities?*

To address the research question, our study was divided into two key objectives. First, we explored the individual factors associated with technology adoption as identified in the literature, using the **Unified Theory of Acceptance and Use of Technology (UTAUT2)** [86] as a guiding framework. Second, we examined the influence of individual cultural values on these factors and on technology adoption itself, operationalizing Hofstede's cultural dimensions to conceptualize participants' cultural backgrounds [37, 38].

The UTAUT2 aims to explain users' intentions and behaviors related to technology acceptance. It includes constructs such as **Performance Expectancy (PE)**, **Effort Expectancy (EE)**, **Social Influence (SI)**, **Facilitating Conditions (FC)**, **Hedonic Motivation (HM)**, **Price Value (PV)**, and **Habit (HB)**. On the other side, Hofstede's cultural values framework identifies six dimensions of culture: power distance, Individualism vs. **Collectivism (COL)**, **Masculinity (MAS)** vs. Femininity, uncertainty avoidance, **Long-Term Orientation (LTO)** vs. short-term normative orientation, and **Indulgence vs. Restraint (IVR)**. These dimensions help in understanding how cultural differences impact various behaviors and attitudes. Furthermore, it is important to emphasize that we do not rely on the cultural dimensions at national level—i.e., measuring the cultural values of an individual relying on their birth place—but at an individual level, measured through the largely used Yoo et al. CVSCALE [93]; this is the correct approach when studying a phenomenon related to individual behavior.

From the literature, we derived a theoretical model incorporating the two concepts mentioned above. In such a model, we integrated the constructs associated with Hofstede's cultural values as *moderators* in the UTAUT2 model[1]. The choice to use cultural values as moderators have been guided by similar literature on the field of cultural influence in technology adoption [7, 49, 61, 80].

We tested our theory using **Partial Least Squares-Structural Equation Modeling (PLS-SEM)** [31, 73]. PLS-SEM is a statistical technique for analyzing complex relationships between observed and latent variables, combining features of factor analysis and multiple regression to simultaneously examine multiple dependent and independent variables. It is particularly valuable in exploratory research and for models involving complex constructs, such as cultural values. Data collection was conducted through two questionnaires: one for pre-screening participants and another for measuring the construct. These questionnaires aimed to capture the adoption behavior of LLMs and the cultural background of 188 software engineers.

We argue that gaining a deeper understanding of the adoption of LLMs in software engineering could offer numerous actionable insights and lay the groundwork for broader research opportunities.

- First, while the interrelationship between culture and technology is a recognized research topic in many disciplines [7, 36, 49, 61, 78, 80], there is a lack of studies in software engineering [59, 74]. This gap in knowledge could lead researchers to incorrect conclusions, resulting in incomplete transfer of knowledge from research to practice. We argue that our work could serve as a milestone in examining the role of cultural values in the adoption of technology in software engineering. Furthermore, although culture has been shown to be important in several technology adoption processes, to date there is no research work covering its role in the adoption of LLMs in software development teams.
- Second, although some researchers have investigated culture and its implications, most have focused on organizational culture. For example, Weerd et al. [89] and Hameed et al. [32]

[1]A moderator effect occurs when the relationship between two variables changes depending on the level of a third variable [31], which in this case are the cultural values. For example, in cultures with high COL, the positive impact of PE on BI might be stronger because individuals in these cultures place more value on technologies that benefit the group. Conversely, in cultures with low COL, the influence of PE on BI might be weaker, as individual benefits are prioritized over group benefits.

demonstrated that strong support from top management is crucial for technology adoption, as it acts as an enabler by providing necessary resources and fostering a conducive environment for change. Moving to the concept of individual culture and focusing on the software engineering community, most have conceptualized individual culture at the national level. When studying phenomena at the individual level, this approach—often based on the scores provided by various cultural frameworks (such as those of Hofstede [38] or GLOBE [40])—fails to capture the substantial variations in cultural values among individuals within the same country, leading to questionable results [25, 26, 30, 60]. To avoid this fallacy, we heeded the call of scholars advocating for a better representation of culture [60, 78] and used individual-espoused cultural values in our study instead of national scores. To the best of our knowledge, this is a novel approach in the field of software engineering research, which still relies heavily on national scores.

— Third, combining cultural values with UTAUT2 constructs to investigate the adoption of LLMs in software engineering has the potential to provide concrete and actionable results [7, 36, 49, 61, 78, 80]. For example, Baptista and Oliveira [7] highlighted the short-term benefits of mobile banking, such as convenience and immediate advantages, while enhancing enjoyment and user experience over usability. Furthermore, Hoehle et al. [36] found that mobile app design and utility had a greater impact on individuals with high collectivistic, feminine, and low power distance values, informing public relations strategies. By identifying the key factors influencing the use of this disruptive technology, we aimed to offer a set of actions that organizations can adopt to develop culturally-sensitive approaches to fostering LLMs adoption. For example, organizations could implement tools, procedures, and cross-sectional processes to integrate the technology into their teams' workflows. This careful integration can mitigate ethical concerns and prevent disruptive situations with potentially catastrophic consequences for project outcomes [45, 63, 72, 95].

— Furthermore, our research could benefit other researchers by paving the way for new research agendas regarding generative AI in software development. It could also inspire others to investigate cultural values with greater awareness, leading to new research projects in similar areas.

In summary, this research makes the following contributions:

— Conducted a statistical evaluation using PLS-SEM to assess the impact of factors from the UTAUT2 model on the intention to use and actual adoption of LLMs for software engineering.
— Performed a moderating analysis using PLS-SEM to examine the role of cultural values in the intention to use and adoption of LLMs for software engineering.
— Made all tests and data publicly available in an online appendix [50] to ensure reliability and promote open-source collaboration.

The article is structured as follows. Section 2 presents the related work. Section 3 reports the theoretical constructs that lead to the definition of hypothesis 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 some reflections and implications of our study. Section 7 closes our research paper with some conclusions and our future research agenda.

2 Related Work

LLMs are advanced AI systems designed to understand and generate human-like text by learning statistical patterns from vast datasets. Originating from the field of natural language processing,

LLMs have demonstrated significant potential across various domains, including software development [12]. In the context of software development, LLMs are particularly useful for tasks such as code generation [6, 29, 43, 44, 54, 57, 90], code explanation [9, 48], and as development assistant [8, 46, 69, 92]. These models, such as OpenAI's GPT-3.5-Turbo and GPT-4-Turbo, as well as open source variants like Meta's Codellama and Deepseek's Deepseek Coder, have been explored for their capabilities in generating tests to validate and verify compiler implementations, showcasing their impressive code generation abilities. Overall, LLMs originate from the need to model and generate human language, and their application in software development and other fields underscores their broad utility and evolving capabilities.

The research community is beginning to explore the adoption of LLMs and the factors influencing this process more deeply. In the following section, we outline current research trends related to LLM adoption, which provide valuable insights for understanding our work and its position within the state-of-the-art, as the research gap we are addressing.

2.1 Use of LLMs in Software Development

The integration of LLMs into software development is revolutionizing the field, offering innovative solutions that enhance both efficiency and quality. Recent research has highlighted diverse applications of LLMs, including code generation [6, 29, 43, 44, 54, 57, 90], code explanation and documentation [9, 48], and conversational programming assistance [8, 46, 69, 92].

LLMs such as GPT-4 have demonstrated significant utility in safety-critical software development, where code quality is paramount. Liu et al. [54] introduced the Prompt-FDC method, which enhances code completeness, comment clarity, and overall output quality. Jiang et al. [44] proposed a structured self-planning framework for code generation, which includes planning and implementation phases. This approach has been shown to improve both the performance and quality of generated code. Fernandes [29] further highlighted LLMs' ability to handle tasks such as code translation, generation, and description, emphasizing their versatility in software development workflows. Wei [90] extended these capabilities by developing a tailored LLM for generating code snippets from well-structured requirements documents, using a progressive prompting technique to enhance efficiency and quality. In the context of testing and validation, Munley et al. [57] demonstrated LLMs' effectiveness in automatically generating tests to validate compiler implementations, including those for the OpenACC parallel programming paradigm. Models such as Deepseek-Coder-33b-Instruct and GPT-4-Turbo produced particularly promising results, showcasing the potential of LLMs in testing processes [57]. As a foundational contribution, Barke et al. [8] conducted a grounded theory study with 20 participants to characterize how developers interact with Copilot for source code generation. Their research resulted in the identification of a bimodal interaction model, highlighting two primary modes of engagement: acceleration and exploration. In the acceleration mode, developers leverage Copilot to expedite tasks when they already have a clear understanding of the next steps. In contrast, the exploration mode is employed when developers are uncertain about the subsequent steps, using Copilot to investigate and generate potential options.

LLMs have also shown a significant impact in enhancing code summarization and program comprehension. Kumar and Chimalakonda [48] successfully employed federated learning for this purpose, highlighting the value of distributed training methods for LLMs. Bhattacharya et al. [9] evaluated the effectiveness of various LLMs in generating natural language summaries for code snippets. Their findings demonstrated that models fine-tuned for code-related tasks consistently outperformed general-purpose models, underscoring the importance of domain-specific optimization.

LLMs are reshaping the workflows of software engineers by serving as conversational assistants. Khojah et al. [46] investigated the use of ChatGPT as a tool for guidance, learning, and

problem-solving, revealing its capacity to address complex engineering challenges. Ross et al. [69] conducted experiments with human participants to evaluate LLMs as programming assistants. Despite initial skepticism, participants were impressed by the assistant's breadth of capabilities, the quality of its responses, and its potential to improve productivity. Additionally, LLMs have demonstrated promise in enhancing code security, including the detection of vulnerabilities and protection of data privacy, outperforming traditional approaches in these areas. However, as Yao et al. [92] cautioned, these advanced capabilities also pose potential risks, such as misuse for user-level attacks due to their human-like reasoning abilities.

As this field continues to evolve [62], a growing body of research is expected to deepen the understanding of LLM applications in software engineering. These forthcoming studies aim to address existing challenges, expand the utility of LLMs, and refine their implementation. The anticipated advancements are likely to revolutionize the field further, paving the way for broader and more impactful applications.

2.2 Factors Influencing the Adoption of LLMs

The adoption of LLMs in professional contexts is shaped by several common factors identified across multiple studies.

Perceived usefulness and *frequency of use* consistently emerge as a critical factor, with users more likely to adopt LLMs if they find them valuable for their tasks [3, 46]. Agossah et al. [3] presented a survey study to examine the intention to use generative tools like LLMs among IT company employees, revealing that perceived usefulness significantly influences the intention to use these tools. The study suggests that the frequency of use is a key factor in understanding employee perceptions, indicating a need for further investigation into the specific requests made to these tools by different audiences [3]. Furthermore, Khojah et al. [46] investigated how professional software engineers use ChatGPT in their day-to-day tasks, focusing on applications beyond generating code. The study concluded that software engineers often use ChatGPT not just for coding tasks but also for seeking guidance on how to approach certain problems, learning about new topics, and receiving advice on software engineering tasks in more abstract terms. Moreover, the authors found that perceived usefulness and trust are protagonist factors in the adoption of LLMs [46].

HB also emerged as a fundamental factor. Draxler et al. [23] investigate the usage patterns of LLMs among 1,500 US citizens, revealing a gender gap in technology adoption and the importance of technology-related education in bridging this gap. Their findings indicate that expert users are more likely to use LLMs for professional tasks, underscoring the need for equitable access and education in AI to maximize workplace benefits [23].

Additionally, the *ease of integration* into existing workflows and *usability* are vital [71]. Recently, Russo has explored the adoption of Generative AI in software engineering. It introduces the **Human-AI Collaboration and Adaptation Framework (HACAF)**, merging various established theories to provide a nuanced understanding of such adoption. Specifically, Russo conducted a mixed-method study; first, a qualitative investigation was conducted to develop the theory, and then, a PLS-SEM statistical analysis was used to test the theory. Russo found that SI is not significant, while the facility to which the tool fits in the working activity is the main aspect to consider [71]. Moreover, Rasnayaka et al. [67] conducted a perception study to gain insights into the perceived usefulness, influencing factors, and future outlook of LLM from a computer science student's perspective. Their findings highlighted the necessity of shifting the educational focus toward preparing students for successful human-AI collaboration.

Trust in the technology is another essential factor, as highlighted by Khojah et al. [46], who found that professional reliance on LLMs for guidance and problem-solving depends heavily on the trustworthiness of these tools. In general, they noted that there is a surprisingly high level of trust

in ChatGPT responses. Regarding reasons for low trust, their study revealed that lack of sources is the main reason detrimental to trust in LLMs.

Related Work: Summary and Research Gap.

The aforementioned studies highlight the transformative potential of LLMs in software development, from enhancing code generation and summarization to improving user interactions and professional task efficiency. They also emphasize the importance of addressing educational and ethical considerations to ensure equitable and effective adoption. Building on this prior research, our study extends the existing knowledge by (1) identifying the most critical factors from the well-established UTAUT2 model that influence the adoption of LLMs, and (2) examining the moderating role of cultural values in this adoption process, thereby providing concrete insights applicable to similar research domains.

3 Hypothesis and Theory Development

This study aims to address two related research gaps. The first gap involves identifying the main individual factors influencing the adoption of LLMs in software development and engineering, a topic not yet thoroughly explored. The second gap focuses on understanding the role of individual cultural values in the adoption process of LLMs.

To achieve our objectives, we developed a set of hypotheses grounded in two theoretical frameworks: the UTAUT2 [86], widely used to explain technology adoption, and Hofstede's Cultural Dimensions framework [37, 38], which decomposes the complex concept of culture into distinct constructs. Below, we explain these two models and the hypotheses derived from them, tailored to the context of LLMs adoption in software development.

3.1 The UTAUT2

3.1.1 The Theoretical Framework. Several theoretical models have been developed to predict the adoption and use of technology. One such model is the UTAUT, created by Venkatesh et al. [85] to forecast technology acceptance within organizational contexts. UTAUT builds upon the key constructs of eight preexisting models that span fields from human behavior to computer science. It identifies four primary factors influencing the intention to use (BI) and actual usage (UB) of information technology: the belief that using the system will improve job performance PE, the ease of using the system EE, the perception that organizational and technical infrastructures are in place to support system use FC, and the perception that important others believe the system should be used SI.

Despite UTAUT's broad acceptance, Venkatesh et al. later introduced UTAUT2, an extended version of the previous model with three additional constructs aimed at capturing aspects related to the user in the role of client and customer rather than simple adopters of technology [86]. The three new constructs in UTAUT2 are the degree of pleasure or enjoyment derived from using technology HM, the user's cognitive tradeoff between the perceived benefits of the technology and the monetary cost of using it PV, and the extent to which people tend to perform behaviors automatically due to learning HB.

3.1.2 Motivation and Adoption Choices. Since our aim was to investigate the adoption of LLMs in software engineering and the moderating role of culture, UTAUT2 was an obvious choice. First, UTAUT2 is a widely used and validated instrument for examining technology adoption in numerous contexts [78, 80, 84–86]. Compared to its predecessors (TAM [21] and UTAUT [85]), it encompasses a broader set of individual-level factors that capture various aspects of technology

adoption [86]. Additionally, the SI and FC constructs enable us to account for environmental factors that may impact the adoption process [86]. Lastly, UTAUT2 allowed us to rely on well-established and validated measurement instruments, as well as contextualize our results within a large body of literature using the same theoretical framework.

In our study, we employed the UTAUT2 model to examine the developers' adoption of LLMs in software engineering. Our primary objective was to explore the influence of individual cultural values on this adoption. The standard version of UTAUT2 also involves three moderating variables, i.e., age, gender, and experience. For model parsimony reasons, previous works [7, 78] investigating cultural values in UTAUT decided not to include these three moderating constructs. Thus, we opted for the same choice.[2]

We imposed a conditional approach to incorporating the PV construct based on our data collection process. We identified three different modalities through which practitioners might use LLMs: provided by their company, accessed through a free version, or obtained via a personal subscription. The PV construct is relevant only in the case of personal subscriptions. Consequently, we collected data on the modality used by participants and decided to include the PV construct in our final model test only if the majority of participants fell into the third category.[3]

3.1.3 Hypothesis Development. In the following, we present the hypothesis we developed for the constructs of the UTAUT2 model.

PE refers to the extent to which an individual believes that using a particular technology will enhance their job performance [85]. This concept indicates that software engineers are more likely to adopt new technologies, such as LLMs, if they perceive these tools as beneficial for their software development tasks [18]. The positive outcomes associated with LLMs, such as increased coding efficiency, improved accuracy, and enhanced problem-solving capabilities, can significantly impact engineers' willingness to integrate these tools into their workflow [46, 48, 54, 71]. Given the potential for LLMs to streamline development processes and provide substantial performance benefits, it is reasonable to propose that PE will be a key determinant in software engineers' intention to adopt these technologies. Therefore, we hypothesize:

H1 *Performance expectancy (PE) positively influences software engineers' intention to adopt (BI) LLMs for software development.*

EE is defined as the degree of ease associated with the use of a particular technology [85]. It implies that individuals are more inclined to adopt new technologies if they find them easy to understand and use [20]. For software engineers, the ease of integrating and utilizing LLMs in their development processes can significantly influence their adoption decisions. If engineers perceive that using LLMs requires minimal effort and reduces the complexity of their tasks, they are more likely to embrace these tools. Given the importance of usability and reduced cognitive load in technology acceptance, it is expected that EE will positively influence software engineers' intention to adopt LLMs for software development. Therefore, we hypothesize:

H2 *Effort expectancy (EE) positively influences software engineers' intention to adopt (BI) LLMs for software development.*

[2]For reliability reasons and thoroughness, we conducted the analysis using the original UTAUT2 moderating variables (age, gender, and experience). The results, available in our online appendix [50], revealed that the original three moderating variables were not significant. Thus, we excluded them for model parsimony, consistent with the approach adopted by previous studies (e.g., [7, 78]).

[3]Our preliminary exploration revealed that most practitioners adopt the free version of LLMs.

SI is the degree to which an individual perceives that important others believe they should use a particular technology [85]. This concept suggests that individuals are more likely to adopt a new technology if they perceive that influential peers, superiors, or societal norms endorse its use. For software engineers, the encouragement and approval from colleagues, mentors, or industry leaders can significantly impact their decision to adopt LLMs. If engineers observe that LLMs are widely accepted and valued within their professional community, they are more likely to follow suit [84]. Given the strong role of peer and social pressure in technology adoption, it is anticipated that SI will positively affect software engineers' intention to adopt LLMs for software development. Therefore, we hypothesize:

H3 *Social influence (SI) positively influences software engineers' intention to adopt (BI) LLMs for software development.*

PV refers to engineers' cognitive evaluation of the tradeoff between the perceived benefits of using LLMs and the associated monetary costs [86]. As mentioned before, it makes sense to consider such an aspect only if the participant is paying the LLM on their own. The PV is considered positive when the perceived benefits of using LLMs outweigh the monetary costs. Therefore, we hypothesize:

H4 *Price Value (PV) positively influences software engineers' intention to adopt (BI) LLMs for software development.*

HM refers to the enjoyment or pleasure experienced when using LLMs for software engineering and development purposes [86] and plays a significant role in user technology acceptance [83]. It is reasonable to think that as the entertainment of the activity increases, so does the practitioner's intention to adopt it. Therefore, we hypothesize:

H5 *Hedonic Motivation (HM) positively influences software engineers' intention to adopt (BI) LLMs for software development.*

FC refer to the belief that an organizational and technical infrastructure exists to support the use of a particular technology [85]. For software engineers, having access to necessary tools, training, and technical support can significantly influence their intention to adopt LLMs [71]. When engineers feel confident that their organization provides sufficient support for integrating and using LLMs, they are more likely to adopt these tools [82]. Furthermore, FC also impacts the actual use of LLMs, as the availability of resources and support ensures effective integration into daily workflows. Therefore, we hypothesize:

H6a *Facilitating conditions (FC) positively influence software engineers' intention to adopt (BI) LLMs for software development.*

H6b *Facilitating conditions (FC) positively influence the actual use behavior (UB) of software engineers regarding LLMs for software development.*

HB refers to the extent to which people tend to perform behaviors automatically because of learning [86]. For software engineers, the familiarity and routine use of certain tools can significantly influence their intention to adopt LLMs. When engineers are accustomed to integrating new technologies into their workflows, they are more likely to intend to adopt LLMs due to their habitual acceptance of innovation [71, 86]. Moreover, HB also affects the actual **Use Behavior (UB)**, as engineers who regularly incorporate LLMs in their working routine are more likely to consistently use LLMs [3, 23, 46, 71]. Therefore, we hypothesize:

H7a *Habit (HB) positively influences software engineers' intention to adopt (BI) LLMs for software development.*

H7b *Habit (HB) positively influences the actual use behavior (UB) of software engineers regarding LLMs for software development.*

In line with models based on psychological theories, which assert that individual behavior is predictable and influenced by personal intention, UTAUT2 posits that **Behavioral Intention (BI)** significantly impacts technology use [85]. Therefore, we hypothesize:

H8 *Behavioral Intention (BI) to use LLMs positively influences the actual use behavior (UB) of software engineers regarding LLMs for software development.*

3.2 The Role of Individual Cultural Values in Technology Adoption

3.2.1 Cultural Values and Frameworks. Geert Hofstede has defined culture as *"The programming of the human mind by which one group of people distinguishes itself from another group."* [37]. The cultural background of individuals influences most of the choices made by humans in a plethora of aspects of everyday life. Undoubtedly, culture represents a complex topic to investigate and formalize, raising an (apparently) unsurmountable wall for researchers eager to embark on this path. To overcome this challenge, researchers from the cross-cultural research field started to conduct empirical investigations to identify values (or dimensions) capable of characterizing individuals' cultural backgrounds and differentiating them. This research agenda led to the development of *cultural dimension frameworks*, which are structured sets of dimensions designed to capture the shared values, beliefs, and behaviors that characterize cultural groups. Each dimension represents a specific aspect of culture, such as attitudes toward power distribution, individualism versus COL, or uncertainty avoidance, providing a systematic way to analyze and compare cultural differences across individuals or groups [37, 38]. Thus, this study adopted Hofstede's cultural dimensions to conceptualize individual-espoused cultural values.

Hofstede's cultural framework is composed of six cultural value dimensions [37, 38]. In this study, we considered only five of Hofstede's cultural dimensions, excluding the sixth dimension, i.e., *IVR*. The reasons for this exclusion are twofold: (1) there is a lack of studies exploring its role as a factor for hypothesis formulation, and (2) there are no validated scales available to measure it at the individual level, as existing scales primarily assess it at the national level.

Power Distance Index (PDI). This dimension measures the extent to which the less powerful members of a society accept and expect an unequal distribution of power. Individuals in high power distance cultures demonstrate strong obedience to and dependence on authority figures. They also tend to adhere closely to organizational norms.

Individualism vs. COL. This dimension assesses the degree to which individuals are integrated into groups. High COL indicates people prefer to take care of themselves and their immediate families rather than integrate themselves into strong, cohesive groups that offer protection in exchange for loyalty.

MAS vs. Femininity. This dimension contrasts societal preferences for achievement, heroism, assertiveness, and material success (high MAS) with preferences for cooperation, modesty, caring for the weak, and quality of life (low MAS).[4]

Uncertainty Avoidance Index (UAI). This dimension reflects the degree to which members of a society feel uncomfortable with uncertainty and ambiguity. Individuals with higher levels of

[4]The MAS dimension has often—particularly in the past—been associated with the gender dimension, which, as previously noted, was found to be non-significant in our preliminary analysis and was therefore excluded for model parsimony reasons. However, this association has weakened over time, particularly when considering cultural values rather than national culture. Recent trends even suggest renaming this dimension as **"Motivation towards Achievement and Success."** Despite this shift, the term "MAS" is retained in this article for its widespread recognition in academia, its clarity in relation to index nomenclature, and its consistent identification in the literature.

uncertainty avoidance exhibit elevated anxiety, perceive uncertainty as a threat, prefer familiar situations, and show intolerance towards unconventional ideas. They tend to be risk-averse and resistant to innovation and change.

Long-Term vs. Short-Term Orientation (LTO). This dimension measures the extent to which a society values long-term commitments and respect for tradition. Individuals with a LTO are known for their forward-thinking mindset, receptiveness to new ideas, and reduced emphasis on traditional practices.

IVR. This dimension gauges the degree of freedom allowed by societal norms in fulfilling human desires. Individuals in indulgent societies are characterized by a free expression of their desires and a tendency to enjoy life and have fun, while those in restrained societies are more likely to control and regulate the fulfillment of their needs through strict social norms.

3.2.2 Hypothesis Development. From the various proposed frameworks, Hofstede's emerged as one of the most adopted in the field of software engineering research [2, 10, 13]. Hofstede's cultural values have also been used in conjunction with the construct of UTAUT2 to investigate the potential influence of culture on the adoption of technology [7, 49, 61, 80]. Some papers from the management research field explored the influence of cultural values as *moderators* for the UTAUT2 constructs [7, 49, 61, 80]; a *moderator* is a construct able to change the direction of correlation relationships [31]. When the moderator amplifies or strengthens the relationship between the independent and dependent variables, it is referred to as positive moderation; conversely, when the moderator weakens or diminishes this relationship, it is referred to as negative moderation. Since this is a well-accreditated way to integrate the two models (UTAUT2 and Hofstede dimensions), we adopted such a choice. In the following, we present the dimensions composing the framework and how we hypothesized they influence the adoption of LLMs for software engineering, supporting our claims using relevant literature.

Previous research has demonstrated that uncertainty avoidance (UAI) positively moderates the effect of SI and negatively moderates the effect of PE on technology usage [79]. For individuals with high uncertainty avoidance, given their resistance to change, SI plays a more significant role in shaping their adoption intentions [4, 41, 78]. Moreover, their inclination to avoid unstructured or uncertain situations can counteract the perceived usefulness of the technology [41]. Furthermore, as Hofstede defines it, UAI is strictly related to people's intention to exit their comfort zone. LLMs are still a young technology, and there is still discussion on the data usage used by the companies developing them. Thus, for individuals with high uncertainty avoidance, intention to use a new technology may be less likely to result in actual use. For the reasons mentioned above, we hypothesized that:

H9 *The UAI value of software engineers would moderate the influence of (H9a) Performance Expectancy on Behavioral Intention (decrease), (H9b) Social Influence on Behavioral Intention (increase), and (H9c) Behavioral Intention on Use Behavior (decrease).*

LTO value is related to the tendency of people to take into account future and prospective events in decision-making [38]. Previous research demonstrated that LTO could have a significant moderating effect when integrated into the Technology Acceptance Model. Individuals with a strong LTO may downplay the importance of HM when considering the use of technology, as they prioritize future benefits over immediate pleasure [96]. Research indicates that the greater an individual's LTO, the more significantly PE affects their intention to use technology [52, 96]. Additionally, those with a long-term focus are more likely to see their BIs translate into actual

actions, as their commitment to long-term goals strengthens their intentions. Therefore, we propose the following hypothesis:

H10 *The LTO value of software engineers would moderate the influence of (H10a) Hedonic Motivation on Behavioral Intention (decrease), (H10b) Performance Expectancy on Behavioral Intention (increase), and (H10c) Behavioral Intention on Use Behavior (increase).*

People's attitudes to being collectivistic or individualistic largely determine the importance of SI on choices. Collectivistic people naturally tend to be primarily influenced by their peers; conversely, PE plays a minor role [1, 30]. Moreover, collectivists tend to consider more EE when adopting a new technology because collective resource and skill utilization can enhance EE [1, 16, 55]. Furthermore, easier technology use aligns with group decisions and supports individuals' sense of belonging [1]. Additionally, Fagih and Jaradat [27] reported that a high collectivistic value negatively moderates the relationship between the intention to use and actual technology use. Consequently, we hypothesize that:

H11 *The COL value of software engineers would moderate the influence of (H11a) Performance Expectancy on Behavioral Intention (decrease), (H11b) Effort Expectancy on Behavioral Intention (increase), (H11c) Social Influence on Behavioral Intention (increase), and (H11d) Behavioral Intention on Use Behavior (decrease).*

As for individualism vs. COL, the Power Distance (PDI) attitude is closely related to SI. Indeed, people with high power distance values are more likely to follow what their superiors or authoritative figures command them to do [51, 53]. The same discourse can be applied to the intention to use technology; people with high power distance could be pushed in their intention to use an instrument by their peers' suggestions [88]. For such reasons, we hypothesize that:

H12 *The PDI value of software engineers would moderate the influence of (H12a) Social Influence on Behavioral Intention (increase) and (H12b) Behavioral Intention on Use Behavior (increase).*

MAS vs. Femininity is the value reported as the less influential in technology adoption. Some papers demonstrated that people with high MAS values could amplify the effect of PE on BI [87]; this is reasonable since MAS is generally associated with competitiveness. Thus, we made the following hypothesis:

H13 *The MAS value of software engineers would moderate the influence of (H13a) Performance Expectancy on Behavioral Intention (increase) and (H13b) Behavioral Intention on Use Behavior (increase).*

We integrated the hypothesis mentioned above with the one of UTAUT2 to investigate the adoption of LLMs for Software Engineering purposes. Thus, we developed the model shown in Figure 1. In the next section, we explain the research steps to test our hypothesis, thus our theoretical model, and answer the research question guiding the investigation.

4 Research Design and Theory Validation

To test our hypotheses and uncover the role of culture in the adoption of LLMs for software development, we conducted a survey study targeting a sample of software engineers. We collected data using a set of questionnaires, each consisting of items validated by existing literature. Subsequently, we applied *PLS-SEM* [31] to test our hypotheses and validate the theoretical framework described in the previous Section 3.

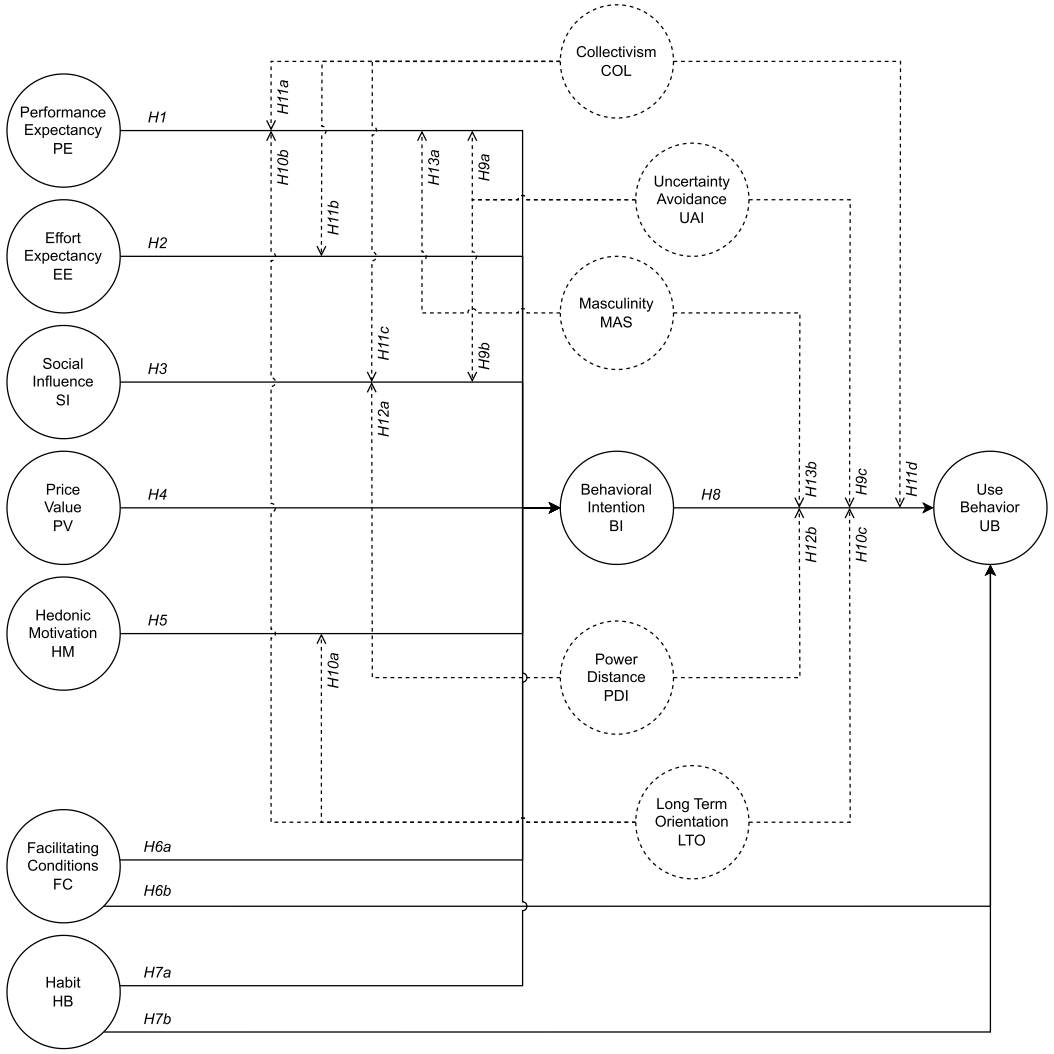

Fig. 1. Theoretical model and hypothesis. Circles indicate constructs of the model; solid lines indicate antecedent relationships between constructs, while dashed lines indicate moderating relationships.

PLS-SEM is a statistical technique used to analyze complex relationships between complex constructs [31]. It combines factor analysis and multiple regression, allowing researchers to model the relationships among multiple independent and dependent constructs simultaneously. PLS-SEM is particularly useful for predictive analysis and theory development when the research model is complex, the sample size is small, or the data distribution is non-normal. It focuses on maximizing the explained variance of the dependent constructs rather than fitting the model to the data.

Figure 2 reports the methodological steps we defined for conducting our investigation. First, we identified a set of criteria for selecting the study participants, as well as the necessary number based on the theoretical model complexity and G*Power [28]. Moreover, starting from the constructs of the theoretical model, we identified validated instruments in the literature to measure them through questionnaires [86, 93]. As a first step, we developed a questionnaire to screen our participants and collect data on their cultural values. Then, we administered a second questionnaire to the

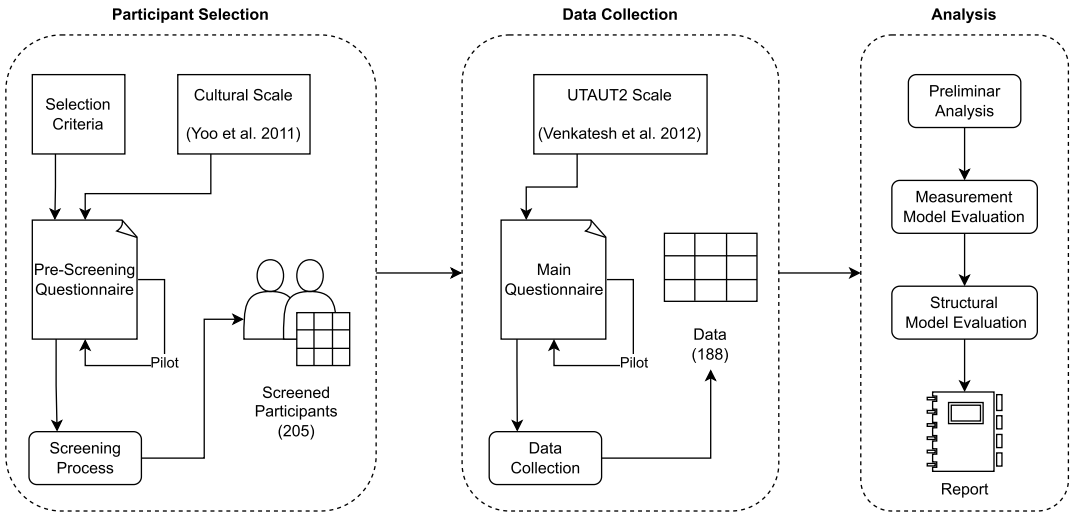

Fig. 2. Research method.

participants who passed the scrutinizing phase to measure the UTAUT2 constructs. Last, we used PLS-SEM to analyze the data and answer our research questions. Regarding the questionnaire development, we used guidelines in the literature and conducted iterative pilots to ensure their quality [5, 47]. Moreover, since our study involved human participation, we sought and received approval from the ethical board committee of Aalborg University, where this research has been conducted.

The section is organized as follows. First, in Section 4.1 we summarize the data collection protocol and the questionnaire employed. Then, in Section 4.2, we describe the indicators (i.e., the questions) used to measure the constructs in our theoretical model, specifically the UTAUT2 [86] and Cultural Values [93] constructs. Next, in Section 4.3, we outline the criteria used to identify the target population for our survey study. This is followed by a description of participants' demographics in Section 4.4. Finally, we provide details on the analysis method in Section 4.5.

4.1 Survey and Data Collection

To collect data, we developed two questionnaires: the first one (from now on referred to as the "Pre-Screening questionnaire") aimed at identifying the ideal participants for our study from the ones already filtered using the Prolific platform; the second one (from now on referred to as the "Main questionnaire") aimed to measure the UTAUT2 constructs in the population obtained from the pre-screening survey.

To design both surveys, we relied on the well-known guidelines developed by Kitchenham and Pfleeger [47], along with Andrews et al. [5], which are widely recognized in software engineering research. Moreover, we adopted the SIGSOFT Empirical Standard for Questionnaire Surveys to design the investigation [65]. The questionnaires were made completely anonymous, and we inserted an intro description that provided all the details that were useful to participants in understanding what they had to do. Moreover, we included a closing question asking for any feedback on the questionnaire. Furthermore, for participants' reliability, we included attention check questions aimed at verifying it.[5]

[5]Ulterior details on the use of the guidelines are reported in Section 4.6.

The questionnaires were designed using Qualtrics, a professional platform for designing questionnaires. We took advantage of various instruments provided by the platform: we checked if the respondent were bots and bias patterns in their responses, and we randomized the order of the questions measuring the constructs to reduce response bias.

We conducted an iterative pilot strategy to (1) test the quality and comprehensibility and (2) estimate the completion time of both surveys. First, we performed three pilots with 10 researchers from our network; after each pilot, we modified the survey to fix typos, reorganize questions within sections, and adjust the introductory text based on the feedback received. These modifications aimed to enhance participant understanding by providing varying levels of detail in the introduction to find the optimal balance. Since the surveys relied on widely validated scales, no changes were made to the content of the survey questions themselves. Then, we conducted a pilot (for each questionnaire) directly in Prolific. We administered the pre-screening questionnaire to five participants who met our criteria, followed by the main questionnaire (we obtained four answers since one of the participants did not complete the second survey). Feedback from this pilot was collected through an open-ended question at the end of each survey, where participants could report any issues or provide suggestions for improvement. No significant issues or patterns indicating problems with the survey design emerged, and the survey flow was confirmed to be effective. Consequently, no further changes were made, and the responses from the Prolific pilot participants were included in the final dataset.

The pre-screening questionnaire aimed at collecting demographic information, participant reliability, and participants' programming skills and experience. Moreover, we included in the pre-screening questions for measuring cultural values constructs. It was estimated to have a duration of 6 minutes and was administered on 5 June 2024; it took two days to gather 200 answers.

The main questionnaire contained questions to measure the UTAUT2 constructs. It was estimated to have a duration of 5 minutes and was administered on 7 June 2024; it took 5 days to gather 184 answers.

Since our study involved human participants, we sought and received approval from the Aalborg University Research Ethics Committee in May 2024. All participants were over 18 years old, provided informed consent before participating, and were informed of their right to withdraw at any time. The questionnaire and additional information are available in our online Appendix [50].

4.2 Measurement Instruments

As previously mentioned, data collection was conducted through a survey study. All constructs in the theoretical model described in Section 3 were measured at the individual level using items validated in the literature, ensuring the reliability of our measurement process. We explain the items used in the following sections, with a comprehensive table of all items and references provided in Appendix A (Table A1).

Questionnaire items measuring the core constructs of the UTAUT2 model were adapted from the original authors [86]. The dependent variable, UB, was measured using a single-item frequency scale. The eight predictors, including seven used in the final analysis and *PV* (excluded due to missing values), were measured using a seven-point Likert scale. These predictors included *PE* (measured with five items), *EE* (measured with six items), *SI* (measured with five items), *HM* (measured with three items), *FC* (measured with four items), *HB* (measured with four items), *PV* (measured with three items), and *BI* (measured with three items). In addition to these items, as mentioned in Section 3, we included a conditional question asking the modality to which the participants of the questionnaire accessed LLMs (free, paid by their company, bought by themselves). We asked for the *PV* indexes only in the second case since, in the others, it would make no sense.

Regarding the measurement of cultural values, Hofstede developed a framework that decomposes culture into specific constructs and provides a questionnaire instrument to measure these constructs [38]. However, Hofstede noted that these instruments are validated at the group level, effectively capturing the cultural characteristics of groups rather than individuals. Moreover, the scores provided by Hofstede are associated with country-level data, meaning that using them to characterize individuals based on their country of origin can lead to incorrect results. Therefore, the instrument provided by Hofstede cannot be used to characterize the cultural values of an individual, as was the case in our study.

Given these limitations, various researchers have developed their own scales based on Hofstede's dimensions to measure cultural values at the individual level. In this study, we adopted the scale proposed by Yoo et al. [93], known as CVSCALE scale, which has been extensively validated for this purpose. The questionnaire items were measured on a five-point Likert scale. For the *LTO* dimension, a score of 1 indicated "very unimportant," and a score of five indicated "very important." For all other dimensions, a traditional agreement Likert scale was used. Specifically, we measured *Power Distance* (PDI, measured with five items), *Uncertainty Avoidance* (UAI, measured with five items), *COL* (measured with six items), *LTO* (measured with six items), and *MAS* (measured with four items). In addition to the variables mentioned above, some demographic background data were collected, including the participants' age, gender, working role, and years of experience in the software industry. The rationale was that we wanted to characterize our sample with respect to other survey works in the literature.

4.3 Participants

To collect data for our investigation, we conducted a survey study. We used a cluster sampling strategy through Prolific, an academic data collection platform, to recruit participants.[6] Prolific enabled us to specify a set of criteria to filter potential participants according to our requirements. Below, we explain the criteria imposed for participant selection, the rationale behind them, and how we ensured these criteria were met through Prolific or other ways—our approach was inspired by Russo's paper on recruiting participants on Prolific [70].

Based on our objectives, we defined the ideal participant profile. Since we aimed to gather software engineers' opinions on using LLMs for software development, we sought practitioners proficient in computer programming, employed full-time or part-time in the software industry, and working as software engineers or in similar roles. Prolific allowed us to set filters for each of these criteria. Additionally, we titled the survey *"[Software Engineering] Pre-Screening—The Role of Culture in the Adoption of LLMs for Supporting Software Engineering"* and clearly stated our goal in the opening description. This helped ensure that only those meeting the pre-screening criteria, even if not actively working as software engineers, participated.

To ensure participants were reliable and capable of answering our questionnaire correctly, we added two further filters. First, participants had to be fluent in English, as the questionnaire was in English. Second, participants needed a 100% approval rate for previous surveys. In Prolific, a researcher can reject a participant's response based on various criteria and checks, and the approval rate represents the percentage of a participant's submissions that have been approved by researchers, indicating their reliability and consistency in providing quality responses.

In addition to the criteria mentioned, we included questions developed by Danilova et al. [19] to assess participants' programming knowledge, ensuring competency in programming. Finally, since our focus was on LLMs, we sought participants with at least some familiarity with LLMs in a work context. To assess this, we included a question about the frequency of LLM use.

[6]Prolific (www.prolific.com) [May 2024].

Based on the criteria mentioned above, Prolific identified a potential pool of 849 participants. The minimum sample size was determined by conducting an a priori power analysis using G*Power [28]. With an effect size of 15%, a significance level of 5%, and a power of 95%, the smallest required sample size for twelve predictors was calculated to be 184 participants. To account for potential attrition between the pre-screening and the main survey, we initially collected 205 responses for the pre-screening survey; for the main one, we gathered 188 responses. The temporal gap between the two surveys allowed us time to screen participants thoroughly. Anticipating that some would not continue to the main survey, we began with a larger number of participants to ensure a sufficient sample size for our analysis.

4.4 Participants Demographics

Our survey featured a varied group of 188 participants, including 73.4% men, 25.5% women, and 1.1% non-binary individuals. The respondents work in 13 distinct countries, with the highest numbers of participants from the United Kingdom (27.1%), the United States (26.6%), Canada (10.1%), Germany (9.6%), and Spain (6.4%). Other countries represented in the survey include Italy, the Netherlands, France, Austria, Ireland, Australia, Switzerland, and Sweden, reflecting the diverse backgrounds of our respondents. In terms of born country, the majority of respondents were born in North America (28.7%), with a significant number from the United States (22.3%) and Canada (6.4%). Western Europe also had a substantial representation at 28.2%, primarily from the United Kingdom (19.7%), Germany (3.7%), and Spain (3.2%). Asian countries accounted for 17.0% of the participants, including India (3.2%), and China (2.1%). Eastern Europe contributed 2.7% to the survey, with notable mentions from Bulgaria and Hungary. Additionally, 4.8% of participants were born in African nations, 1.1% in South America, and 1.1% in Oceania. This wide-ranging geographic distribution underscores the global nature of our respondent pool.

Our survey featured a range of work positions among the 188 respondents. The most common roles were Software Developer/Programmer (33.1%) and Software Engineer (22.8%). Other significant positions included Data Analyst/Data Engineer/Data Scientist (10.6%), Team Leader (7.5%), Project Manager (6.6%), and Tester/QA Engineer (5.6%). UX/UI Designer and those in other specified roles each accounted for 4.4%. Additionally, Software Architects comprised 3.4% of respondents, while CIO/CEO/CTO roles represented 1.6%.

The survey responses reflect a diverse range of experience levels among participants. In terms of working role experience, the majority have between 1 and 3 years (34.0%), followed by 4–6 years (28.7%), and more than 10 years (20.2%). A smaller portion of respondents have 7–9 years (11.7%) and less than 1 year (5.3%) of experience. When considering experience within the software industry, a significant portion of participants have over 10 years (34.0%) of experience. This is followed by those with 1–3 years (26.1%), 4–6 years (23.4%), 7–9 years (11.2%), and less than 1 year (5.3%). These statistics highlight a broad spectrum of expertise within the surveyed group, with a notable presence of both early-career and highly experienced professionals.

4.5 Analysis Overview

In terms of analysis, we first conducted a *preliminary examination* of the data collected. Although PLS-SEM does not impose strict constraints, we checked for missing data, suspicious response patterns, outliers, and data distribution to ensure data quality.

Once the data was validated, we imported the dataset into SmartPLS, a software tool designed for conducting PLS-SEM statistical analysis [68]. Here is a summary of how analysis with PLS-SEM works. Due to the complexity of the process, we recommend referring to Hair et al. [31] and Russo and Stol [73] for a full and comprehensive explanation. First, we created the measurement model (also known as the outer model), which combines the construct to be measured with its associated

indicators. For each construct in the theoretical model (i.e., UTAUT2 and cultural values), we created a latent variable representing a non-observable construct. We then associated the indicators, i.e., the participants' questionnaire responses, with each corresponding construct. Next, we created the structural model (also known as the inner model), which represents the relationships between the constructs based on our theoretical hypotheses. Finally, we incorporated the moderating effect of cultural values into the model. A moderating effect hypothesizes that the value of one construct can change the direction of the relationship between two other constructs. This was achieved by associating each cultural value construct with the corresponding relationships discussed in our hypotheses.

After constructing both the measurement and structural models within the tool and running the PLS-SEM algorithm, we initially evaluated the measurement models. Given that all our indicators had a reflective relationship with the constructs, we assessed the following:

— *Indicator Reliability*: Ensures that each item consistently represents the intended construct.
— *Internal Consistency Reliability*: Assessed using composite reliability to ensure the consistency of the constructs.
— *Convergent Validity*: Verified that the indicators of a construct share a high proportion of variance through the **Average Variance Extracted (AVE)**.
— *Discriminant Validity*: Confirmed that constructs are distinct from one another.

Following the measurement model assessment, we proceeded to evaluate the structural model. We assessed:

— *Collinearity*: Ensured that predictor variables in the model are not highly correlated, using **Variance Inflation Factor (VIF)**.
— *Significance of Relationships*: Determined the statistical significance of path coefficients using bootstrapping.
— *Explanatory Power*: Evaluated how well the model explains the variance in the dependent variables using R-squared values.
— *Predictive Power*: Assessed the model's ability to predict outcomes using measures such as Q-squared.

Since our structural model included cultural values as moderators, we performed a moderator analysis. In PLS-SEM, conducting a moderator analysis requires introducing an additional construct for each moderating effect, known as the interaction term. This involves multiplying each standardized indicator of the moderator variable with those of the independent variable and creating a relationship between the interaction term and the dependent variable. Both the moderating construct and the interaction term must be used in the model evaluation: the moderating construct is assessed in the measurement model, while the interaction term is evaluated in the structural model.

To ensure replicability and reliability, we included all materials used during the analysis in the replication package accompanying this article [50].

4.6 Design Choices and Justification

To design the survey study, we relied on the well-known guidelines developed by Kitchenham and Pfleeger [47], along with Andrews et al. [5], which are widely recognized in software engineering research. As suggested by Kitchenham and Pfleeger [47], the objective of the survey was defined and categorized within the cross-sectional survey typology. Additionally, validated instruments from the literature were identified instead of designing new ones, allowing us to leverage robust and validated tools. The guidelines provided by Kitchenham and Pfleeger also informed the presentation

of the questionnaire: the formatting tools in Qualtrics were used to emphasize important elements (e.g., bold and italic fonts), and the questions were organized into sections. Respondents were also provided with essential information to understand and complete the survey, presented in its introductory section. Finally, the target population was identified using filters available on Prolific and through pre-screening procedures. The guidelines of Andrews et al. [5], which are more specific to online surveys, were also followed. To adhere to these, Qualtrics was used to ensure (1) maximum compatibility across devices, (2) prevention of multiple submissions, (3) provision of an adaptive template for different devices and browsers, and (4) the ability for respondents to leave feedback. Furthermore, pilot testing was conducted through Prolific, as recommended by Andrews et al. [5].

Moreover, we adopted the SIGSOFT Empirical Standard for Questionnaire Surveys to design the investigation [65]. Here, we present examples of how we followed them:

—Essential attributes were rigorously applied throughout the study. The target population was identified, and the sampling strategy was defined using specific criteria through the Prolific platform, as detailed in Section 4.3. Measurement instruments were developed based on validated scales from the literature, including the UTAUT2 constructs and Yoo et al.'s CVSCALE [93] (Section 4.2). Iterative pilots were conducted to test and refine the questionnaires prior to their administration, ensuring their quality and comprehensibility. Data handling included systematic actions for addressing missing values, suspicious response patterns, and outliers, demonstrating a systematic and replicable methodology (Section 5.1). Furthermore, all materials used, including the questionnaires, were made available in an online appendix to promote transparency and replicability [50].

—Desirable attributes were also considered to enhance the robustness of our study. The demographic characteristics of the target population were reported, providing a detailed characterization of the participants (Section 4.4). Ethical principles were carefully addressed, with approval obtained from the ethics committee and informed consent collected from all participants. Pilots and iterative testing ensured the appropriateness of the instrument in terms of layout, duration, and comprehensibility. Additionally, efforts were made to minimize response bias through—for example—randomization of question order and inclusion of attention-check items.

—Finally, the study incorporated certain extraordinary attributes to elevate its methodological rigor. For example, construct validity analysis was performed to assess the convergent and discriminant validity of the constructs, confirming their reliability and distinctiveness (Section 5.2). Advanced statistical techniques, such as PLS-SEM, were employed to ensure the robustness of the analysis and to validate the theoretical framework [31].

5 Analysis of the Results

The following section presents the results obtained from the PLS-SEM analysis, offering insights into the structural model's path coefficients, the measurement model's reliability and validity, and the overall model fit. Through this analysis, we aim to understand the causal relationships and underlying patterns within the data, providing a comprehensive evaluation of the hypothesized model.

5.1 Preliminary Examination

Before conducting the principal PLS-SEM investigation, we performed a preliminary data analysis. Notably, we did not encounter any missing data, which can be attributed to the quality of our questionnaire and the reliability of our sample, ensured by the approval rate filter. Furthermore, we checked for suspicious response patterns in the answers and did not find any. It is also worth noting that none of the participants failed the attention check questions.

Regarding outliers, an analysis using boxplots and statistical tools revealed very few. Since our population meets all the requirements and criteria, we chose not to eliminate the outliers as they represent actual data. Additionally, we examined data distribution for skewness and kurtosis. Our analysis indicated a nonnormal distribution in the data for four of the indicators for cultural values. As a general guideline, values beyond −2 and +2 indicate nonnormality. Despite normality is not required in PLS-SEM, nonnormal distribution is in general problematic [31]. For such a reason and since only four indicators suffered from this problem—i.e., one for PDI (PDI_3), one for UAI (UAI_3), and two for LTO (LTO_1 and LTO_6)—we excluded them from the PLS-SEM analysis.

As a final check, we evaluated the PV construct. The majority of participants (51.9%) use a free version of the software, 33.7% have their software provided by their company, and only 14.4% purchased the software independently. Including the PV construct in our analysis was meaningful only for participants who bought the tool independently. Given that most participants did not, we excluded this construct from the final investigation.

5.2 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 [31, 73]. 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.

5.2.1 Step 1—Indicator Reliability. As indicated in the work of Hair et al. [31], the first step in assessing the measurement model involves checking the reliability of the indicators, specifically the *outer loadings* of the indicators. High outer loadings on a construct indicate that the associated indicators share a significant amount of commonality, which is captured by the construct.

Guided by a practical rule of thumb, indicators with outer loadings exceeding 0.708 are typically retained in the model, while those below 0.40 are usually discarded. For indicators falling between 0.40 and 0.70, the decision to eliminate them should be based on whether their removal enhances internal consistency reliability or convergent validity.

Tables 1 and 2 report all the outer loadings for the indicators of our model. As observed, some of the indicators, particularly those associated with cultural values, presented outer loadings less than 0.70. This is not surprising, as similar results have been reported in other studies using cultural values. This is because the constructs represent culture as a contraposition between two ways of thinking and behaving. We eliminated all the indicators with outer loadings less than 0.40 and evaluated the remaining indicators based on the aforementioned rules.

5.2.2 Step 2—Internal Consistency Reliability. The second step involved testing internal consistency reliability, ensuring that the indicators consistently and reliably measure the constructs. To assess this, we referred to three measures: *Cronbach's alpha*, *composite reliability* (ρ_c), and the *reliability coefficient* (ρ_A).

The results are presented in Table 3. All the values for these measures exceeded the recommended threshold of 0.60 [31], except for Cronbach's alpha for LTO. Nevertheless, we retain this construct for three reasons: (1) the lower value can be attributed to the inherent complexity and variability in measuring LTO, and is coherent with similar works using the CVSCALE [49]; (2) the scale used for measuring LTO is largely adopted and validated; (3) the other values to measure Internal Consistency Reliability, i.e., the *composite reliability* (ρ_c) and the *reliability coefficient* (ρ_A), are above the treshold.

5.2.3 Step 3—Convergent Validity. Convergent validity is the extent to which a measure correlates positively with alternative measures of the same construct [31]. Since we selected reflective

Table 1. Outer Loadings of the UTAUT2 Constructs

Indicators	EE	FC	HB	HM	PE	SI	BI	UB
EE_1	0.769							
EE_2	0.814							
EE_3	0.827							
EE_4	0.831							
EE_5	0.714							
EE_6	0.755							
FC_1		0.848						
FC_2		0.889						
FC_3		0.730						
HB_1			0.869					
HB_2			0.786					
HB_3			0.734					
HB_4			0.876					
HM_1				0.933				
HM_2				0.933				
HM_3				0.858				
PE_1					0.870			
PE_2					0.915			
PE_3					0.877			
PE_4					0.865			
PE_5					0.744			
SI_1						0.937		
SI_2						0.951		
BI_1							0.889	
BI_2							0.905	
BI_3							0.936	
UB								1.000

BI, Behavioral Intention; EE, Effort Expectancy; FC, Facilitating Conditions; HB, Habit; HM, Hedonic Motivation; PE, Performance Expectancy; SI, Social Influence; UB, Use Behavior.

indicators for all the constructs, we expected that the indicators of the same construct converge or share a high proportion of variance. A standard measure used to assess this is the AVE. In general, an AVE value of 0.50 or higher indicates that the construct explains more than half of the variance of its indicators, which is desirable.

Table 3 reports the AVE values for each construct in the model. As it is possible to observe, all the constructs vastly exceed the value of 0.50, allowing us to pass this step confidently.

5.2.4 Step 4—Discriminant Validity. The last test aimed at checking the discriminant validity, i.e., the extent to which a construct is truly distinct from other constructs by empirical standards. To assess the discriminant validity, Henseler et al. [34] proposed the **Heterotrait-Monotrait (HTMT)** ratio. HTMT is computed for each construct by means of the PLS-SEM algorithm. Typically, a value above 0.90 indicates a lack of discriminant validity, while a value below 0.85 seems warranted. Moreover, it is suggested to rely on a *bootstrapping* procedure to check if the HTMT values are significantly different from the threshold values. Bootstrapping is a nonparametric procedure that allows testing the statistical significance of various PLS-SEM results, such as path coefficients, Cronbach's alpha, and HTMT values.

Table 2. Outer Loadings of the Cultural Constructs as Moderators

Indicators	COL	LTO	MAS	PDI	UAI
COL_2	0.634				
COL_3	0.879				
COL_4	0.913				
COL_6	0.723				
LTO_2		0.927			
LTO_4		0.668			
MAS_1			0.796		
MAS_2			0.933		
MAS_3			0.797		
PDI_1				0.900	
PDI_2				0.815	
PDI_4				0.787	
UAI_2					0.920
UAI_5					0.811

COL, Collectivism; LTO, Long Term Orientation; MAS, Masculinity; PDI, Power Distance Index; UAI, Uncertainty Avoidance Index.

Most of the HTMT values fall beneath the threshold of 0.85, and some are slightly above it. We used SmartPLS to run a bootstrapping procedure with 10,000 subsamples, one-tailed as the test type, and a 0.05 significance level; also in this case, all the values fall beneath the thresholds. This result suggests that every construct in the model represents a distinct phenomenon and allows us to proceed with assessing the structural model.[7]

Measurement Model Evaluation: Summary of the results.

The measurement model evaluation confirmed the reliability and validity of the constructs. Indicator reliability was established, with most outer loadings exceeding the threshold of 0.708. Internal consistency reliability was supported by Cronbach's alpha, composite reliability, and reliability coefficients; most of the metrics were above the thresholds. Convergent validity was verified with AVE values above 0.50 for all constructs. Discriminant validity was demonstrated using HTMT values, all of which were below the critical threshold of 0.85, supported by bootstrapping results. Overall, the constructs exhibit strong reliability and validity, ensuring the robustness of the measurement model.

5.3 Structural Model Evaluation

After assessing the quality of the measurement model, the next step is to test the structural model. While the measurement model assessment focused on the moderating effect to test the influence of cultural values, this phase centers on the interaction term, which encompasses the concrete impact of the moderating effect [31].

5.3.1 Step 1—Collinearity Analysis. The first step in assessing the structural model involves verifying the collinearity between the exogenous and endogenous variables. Ensuring that our

[7]The table with HTMT values is reported in the appendix.

Table 3. Internal Consistency Reliability of the Constructs

Constructs	Cronbach's alpha	ρ_A	ρ_c	AVE
Performance Expectancy	0.908	0.918	0.932	0.733
Effort Expectancy	0.877	0.888	0.906	0.618
Facilitating Conditions	0.764	0.789	0.864	0.681
Social Influence	0.878	0.887	0.942	0.891
Habit	0.836	0.860	0.890	0.670
Hedonic Motivation	0.894	0.905	0.934	0.826
Behavioral Intention	0.897	0.897	0.936	0.829
Long Term Orientation	0.507	0.663	0.785	0.653
Masculinity	0.830	0.987	0.881	0.713
Power Distance	0.788	0.853	0.873	0.698
Uncertainty Avoidance	0.682	0.758	0.858	0.753
Collectivism	0.813	0.942	0.871	0.633

The black line in the middle of the table separates the UTAUT2 constructs and the Cultural Values constructs.

model does not suffer from multicollinearity issues is crucial for accurate path estimation. The structural model's estimation relies on ordinary least squares regressions of each endogenous construct on its corresponding predecessor constructs. Therefore, low collinearity is essential to avoid bias in the path estimation.

To assess collinearity, we employed the VIF, a widely recognized measure for detecting the presence and severity of multicollinearity in multiple regression models. Generally, a VIF value less than three indicates no collinearity and is desirable; however, values less than five are also considered acceptable.

In our case, most values were below three (the greatest among them was 2.47), with only two slightly exceeding this ideal threshold (PE -> BI and EE -> BI). Consequently, we concluded that multicollinearity is not a significant issue in our model.

5.3.2 Step 2—Significance and Relevance of the Relationships. The second step in the analysis involved testing for the significance of the relationships of the structural model as well as for their relevance. Such a test occurs by means of *bootstrapping* (for the significance) and observing the *path coefficients* of the relationships (for the relevance); such path coefficients are standardized (their values can range from −1 to +1) and thus can be analyzed relative to each other; a higher path coefficient indicates a greater relevance.

Regarding the significance, we used a two-tailed bootstrapping method with 10,000 sub-samples. T-values, p-values, or bootstrap confidence intervals can be used to test the significance. Hair et al. [31] suggest using the last one as a measure. Tables 4 and 5 report the values for testing the significance of our relationships; for the moderating variables (Table 5), one has to see the values associated with the interaction term (the ones expressed as products of two constructs) rather than the one associated with the moderating construct. As it is possible to observe, PE and HB have a significant relationship with BI. Moreover, Facilitating Condition, HB, and BI are significantly related to UB. Furthermore, none of the cultural values constructs is significant.

We then turned our attention to the relevance of the significant relationships. We did this by closely examining the path coefficients of these paths in the PLS-SEM algorithm computation; such values are also reported in the second column of Tables 4 and 5. PE is the most relevant construct for the intention to use LLMs in software development, followed by HB. Moreover, focusing on the

Table 4. Significance and Relevance of the UTAUT2 Constructs

Hypotheses	Path Coefficients	Bootstrap Mean	St. Dev	T statistics	P values	Significance
EE → BI	0.087	0.097	0.081	1.074	0.283	
FC → UB	0.173	0.174	0.062	2.799	0.005	**
FC → BI	0.008	0.017	0.067	0.116	0.908	
HB → UB	0.378	0.379	0.082	4.613	0.000	***
HB → BI	0.274	0.260	0.083	3.292	0.001	***
HM → BI	0.106	0.115	0.070	1.513	0.130	
PE → BI	0.463	0.441	0.091	5.060	0.000	***
SI → BI	−0.034	−0.024	0.057	0.600	0.549	
BI → UB	0.176	0.171	0.091	1.940	0.050	*

$***: p < 0.001; **: p < 0.01; *: p < 0.05$.
Significant paths are highlighted in bold. BI, Behavioral Intention; EE, Effort Expectancy; FC, Facilitating Conditions; HB, Habit; HM, Hedonic Motivation; PE, Performance Expectancy; SI, Social Influence; UB, Use Behavior.

Table 5. Significance and Relevance of the Cultural Values Moderators

Hypotheses	Path Coefficients	Bootstrap Mean	St. Dev	T statistics	P values
COL → UB	−0.098	−0.090	0.070	1.399	0.162
COL → BI	−0.031	−0.043	0.063	0.491	0.624
LTO → UB	−0.010	−0.001	0.079	0.130	0.897
LTO → BI	0.025	0.017	0.065	0.381	0.703
MAS → UB	−0.044	−0.043	0.079	0.555	0.579
MAS → BI	−0.039	−0.034	0.057	0.688	0.491
PDI → UB	−0.115	−0.122	0.070	1.639	0.101
PDI → BI	−0.061	−0.072	0.057	1.056	0.291
UAI → UB	0.002	−0.001	0.071	0.024	0.981
UAI → BI	−0.007	−0.003	0.058	0.113	0.910
UAI × SI → BI	−0.067	−0.060	0.060	1.112	0.266
MAS × PE → BI	0.021	0.021	0.052	0.393	0.694
LTO × BI → UB	−0.035	−0.030	0.068	0.515	0.607
UAI × BI → UB	−0.021	−0.018	0.065	0.328	0.743
UAI × PE → BI	−0.038	−0.046	0.060	0.642	0.521
COL × SI → BI	0.072	0.073	0.068	1.058	0.290
COL × PE → BI	0.058	0.050	0.087	0.669	0.503
PDI × BI → UB	0.022	0.025	0.064	0.344	0.731
LTO × HM → BI	−0.009	−0.012	0.079	0.117	0.907
LTO × PE → BI	−0.067	−0.056	0.080	0.834	0.404
MAS × BI → UB	0.062	0.022	0.082	0.748	0.454
COL × BI → UB	−0.064	−0.079	0.074	0.861	0.389
COL × EE → BI	−0.077	−0.057	0.072	1.067	0.286
PDI × SI → BI	−0.009	0.000	0.061	0.141	0.888

The hypotheses listed below the black line in the middle of the table represent the interaction terms generated for the moderating analysis. COL, Collectivism; LTO, Long Term Orientation; MAS, Masculinity; PDI, Power Distance Index; UAI, Uncertainty Avoidance Index.

concrete use of LLMs, HB has the greatest relevance, followed by the intention to use the LLM and Facilitating Condition.

5.3.3 Step 3—Explanatory Power. The third step of our analysis involved assessing the model's explanatory power, specifically its ability to fit the data by quantifying the strength of the associations [31, 73]. This is generally measured using the *coefficient of determination* (R^2) of the dependent variables, which ranges from 0 to 1. Higher R^2 values indicate greater explanatory power. While there are no universally agreed-upon standards for R^2 values, in some contexts, a value of 0.10 may be satisfactory, though a commonly accepted benchmark is 0.19 or greater [17, 31, 64].[8]

Our model yields an R^2 value of 0.64 for BI and 0.41 for UB. This indicates that our model explains 64% of the variance in the intention to use LLMs and 41% of the actual use. Additionally, since our R^2 values are below 0.90, we can dismiss concerns about overfitting.

After assessing the coefficient of determination, it is useful to quantify the strength of the relationships using the F^2 *effect size*. This measure indicates the impact on R^2 if a specific construct is omitted from the model, thus providing insight into the concrete impact of each construct on the dependent variables.

Focusing on the intention to use LLMs, PE exhibits the greatest effect size (0.17), followed by HB (0.09). For the actual use of LLMs, HB has the highest effect size (0.12), followed by FC (0.04) and Intention Behavior (0.02). These results highlight the predominant role of PE in influencing the intention to use LLMs and the importance of habitual usage in predicting actual use.

5.3.4 Step 4—Predictive Power. To assess the model's utility for managerial decision-making, we tested whether the results of our PLS-SEM algorithm are applicable beyond the data used in the estimation process. We employed the $\text{PLS}_{\text{predict}}$ procedure [76], which involves separating the overall dataset into training and holdout samples. The measures used in this analysis include Stone-Geisser's Q^2 *statistic*, the **Mean Absolute Error (MAE)**, and the **Root Mean Square Error (RMSE)**. These statistics were tested against a benchmark, with Shmueli et al. [76, 77] proposing the use of linear regression model (LM) results as the benchmark.

Positive Q^2 values indicate that the model's prediction error is smaller than that of the benchmark. Additionally, lower values of MAE and RMSE suggest that the model possesses high predictive power.

Table 6 reports the values for the dependent variables in the model. All variables exhibit better values compared to the benchmark, indicating that the PLS-SEM model demonstrates high predictive power.

Structural Model Evaluation: Summary of the results.

The structural model evaluation confirmed the robustness of the proposed model. Collinearity analysis showed acceptable VIF values, indicating no significant multicollinearity issues. Path coefficients and bootstrapping results showed significant relationships, particularly for PE and HB with BI, and FC, HB, and BI with UB. Furthermore, cultural values moderating effect does not seem to be significant. The model exhibited strong explanatory power with R^2 values of 0.64 for BI and 0.41 for UB. Predictive power assessment through $\text{PLS}_{\text{predict}}$ indicated high predictive accuracy, with all variables showing better values compared to the linear regression benchmark.

6 Discussion and Implications

This study set out to explore the impact of cultural values on the adoption of LLMs in software engineering, employing the UTAUT framework. Our primary objective was to understand whether

[8]The Standardized Root Mean Square Residual (SRMR) was calculated to evaluate the discrepancy between observed and predicted correlations, yielding a value of 0.069, below the suggested threshold of 0.08 [33].

Table 6. Prediction Summary (PLS-SEM) and Comparison with Linear Regression (LM)

Indicators	$Q^2_{predict}$	PLS-SEM_RMSE	PLS-SEM_MAE	LM_RMSE	LM_MAE
UB	0.312	1.088	0.820	1.222	0.929
BI_1	0.431	0.694	0.531	0.728	0.539
BI_2	0.409	1.022	0.787	1.085	0.838
BI_3	0.447	0.873	0.657	0.956	0.710

BI, Behavioral Intention; UB, Use Behavior.

cultural background plays a significant role in influencing software engineers' decisions to adopt LLMs, a pertinent inquiry given the increasing integration of generative AI in the field.

In the following section we discussed the statistical results of our investigation and provide a set of implications and lessons learned both for practitioners and researchers.

6.1 Discussion on Hypotheses Significance

The findings revealed that cultural values did not have a significant impact on the intention to adopt or the actual adoption of LLMs. In contrast, several constructs from the UTAUT2 model—namely HB, PE, FC, and BI—were found to significantly influence the dependent variables. These results suggest that individual and organizational factors, rather than cultural backgrounds, play a more critical role in the adoption process of LLMs in software engineering.

With these findings in mind, the following sections will provide a detailed discussion of each UTAUT2 construct and the cultural values construct, examining their relevance and the rationale behind our results.

6.1.1 On the Direct Effect of Technology Acceptance Constructs. First, we focused on the results regarding the UTAUT2 constructs. Table 7 summarizes our findings.

PE emerged as the most influential factor affecting the *BI to use LLMs* for software engineering. This result underscores the critical role that perceived benefits play in adopting new technologies [18, 85]. In the context of software engineering, practitioners are highly motivated by tools that can demonstrably enhance their productivity and efficiency. In line with the related literature [46, 48, 54], with their capability to generate code, assist in debugging, and provide decision support, LLMs directly address core aspects of software engineers' tasks, fulfilling their performance expectations. This alignment with crucial performance goals makes LLMs particularly attractive to users who are constantly seeking ways to improve their work outcomes. Moreover, the clear and immediate benefits of LLMs in streamlining complex processes probably reinforce the positive perception of their utility. For example, LLMs have been demonstrated to benefit code development by translating well-structured requirements into functional code snippets, effectively emulating the expertise of experienced software engineers [90], and DevOps practices by automating GitHub workflows, ensuring syntax correctness and improving the agility of software delivery pipelines [56, 94]. These benefits could lead to a strong intention to integrate these models into daily workflows [3, 46, 71]. This finding aligns with the broader technology adoption literature, which consistently identifies PE as a primary driver of BI, particularly when the technology in question offers significant enhancements to job performance [61]. Moreover, these results are aligned with the findings of Agossah et al. [3] and Khojah et al. [46], who both identified ease of use and frequency of use as fundamental factors in the adoption of LLMs by developers.

HB was identified as the second most influential factor on *BI* to use LLMs for software engineering, and it also showed outstanding significance. Furthermore, HB exhibited the highest path coefficient with *actual UB*, and mediation analysis revealed that this influence is direct and not mediated by *BI*.

Table 7. UTAUT2—Summary of Findings and Implications

Hypotheses	Findings	Implications
H1: Performance Expectancy → Behavioral Intention	Supported. This relationship is the strongest of the model, with a path coefficient of 0.46 and an effect size of 0.17	Software engineers' intentions to adopt LLMs are heavily influenced by their expectations of the technology's performance.
H2: Effort Expectancy → Behavioral Intention	Not supported. The relationship is not significant.	The perceived effort in learning how to use and use LLMs alone does not instigate the intention to adopt the tool in working context.
H3: Social Influence → Behavioral Intention	Not supported. The relationship is not significant.	The peers opinion on the use of LLMs in the working context alone does not instigate the intention to adopt the tool.
H4: Price Value → Behavioral Intention	Not investigated. The lack of a considerable number of individuals having purchased LLMs with their own money did not allow the hypothesis to be investigated.	N/A
H5: Hedonic Motivation → Behavioral Intention	Not supported. The relationship is not significant.	Developers do not intend to adopt LLMs on the basis of the fun and joy their use causes.
H6a: Facilitating Conditions → Behavioral Intention	Not supported. The relationship is not significant.	Organizational support and supporting resources do not significantly influence developers' intention to adopt LLM tools in the working context.
H6b: Facilitating Conditions → Use Behavior	Supported. This relationship is the third most significant regarding Use Behavior, with a path coefficient of 0.17 and an effect size of 0.04. A mediation analysis revealed that the effect is direct and not mediated by the relationship with Behavioral Intention.	While engineers may not initially consider the availability of resources and support systems in their decision to adopt LLMs, these factors become crucial once they start using the technology.
H7a: Habit → Behavioral Intention	Supported. This relationship is the second most significant regarding Behavioral Intention, with a path coefficient of 0.27 and an effect size of 0.09.	Developers that regularly utilized LLMs in their work strengthen their intention to rely on these tools for their tasks.
H7b: Habit → Use Behavior	Supported. This relationship is the most significant regarding Use Behavior, with a path coefficient of 0.38 and an effect size of 0.12. A mediation analysis revealed that the effect is direct and not mediated by the relationship with Behavioral Intention.	The habitual use of LLMs can lead to higher adoption rate.
H8: Behavioral Intention → Use Behavior	Supported. This relationship is the second most significant regarding Use Behavior, with a path coefficient of 0.18 and an effect size of 0.02.	As expected, the intention to adopt LLMs results in an actual adoption of the technology.

This result highlights the pivotal role of habitual behavior in the adoption and continued use of LLMs and are consistent with previous literature on the matter [3, 23, 46, 71]. The direct influence of HB on UB suggests that when engineers incorporate LLMs into their regular workflows, the technology becomes an integral part of their daily activities. This habitual use is likely driven by the consistent and tangible benefits that LLMs provide, such as increased efficiency and ease of performing repetitive tasks. The strong habitual behavior observed underscores the importance of initial exposure and ease of integration into existing processes, as these factors can lead to the sustained use of LLMs.

FC emerged as the third most significant variable influencing *actual UB of LLMs*, with a very strong path coefficient and direct relationship, unmediated by *BI*. This result underscores the critical role that the availability of resources, support, and infrastructure plays in the practical adoption and utilization of technology [82, 85]. Moreover, these findings are consistent with the work of Russo [71], which shows that facilitating the integration of LLMs in the development lifecycle is pivotal. Furthermore, this direct relationship indicates that even if engineers do not initially intend to use LLMs, the ease of access and the supportive environment can drive actual use. The lack of a significant relationship between FC and BI suggests that practical considerations and immediate usability factors outweigh initial user intentions when it comes to the adoption of LLMs [49]. In summary, engineers are more likely to use LLMs when they feel adequately supported and equipped to do so, regardless of their prior intention. By ensuring that these conditions are met, organizations can effectively promote the use of LLMs, even among those who may not have initially planned to adopt the technology. Additionally, the result is consistent with the observations of Rasnayaka et al. [67], who emphasized in their work that educating professionals (starting from their student years) on the use of these technologies is essential to ensure their proper adoption.

BI is the second most significant variable influencing the *actual UB of LLMs*. This result highlights that a strong intention to use LLMs translates directly into their actual adoption. This aligns with the broader technology acceptance literature, which shows that BI is a key predictor of actual use [85, 86]. By fostering positive intentions through training, awareness programs, and demonstrating the benefits of LLMs, organizations can effectively encourage their adoption and integration into daily workflows. Interestingly, other variables such as *FC* and *HB* show even stronger direct effects on actual use than BI does. This suggests that while having the intention to use LLMs is important, the practical aspects of support and resources (FC) and the integration of LLMs into daily routines (HB) play an even more critical role in ensuring actual adoption [3, 23, 46, 71]. These findings emphasize the need for organizations to not only build positive intentions but also provide robust support systems and encourage habitual use to maximize the adoption and effective utilization of LLMs.

EE was found to be non-significant in its relationship with *BI to use LLMs in software engineering* despite the broader technology acceptance literature [85, 86]. This result suggests that the perceived ease of use of LLMs does not strongly influence engineers' intentions to adopt these tools and finds support in the results of Russo [71]. One possible justification for this finding is that software engineers, as highly specialized professionals, might not view the effort required to learn and use LLMs as a major barrier. Given their technical expertise and familiarity with complex tools, they may prioritize the functionality and performance benefits of LLMs over the effort required to use them. Additionally, the strong influence of other factors such as *PE* and *HB* indicates that the potential improvements in job performance and the integration of LLMs into daily routines are more critical in shaping their adoption intentions; this is also supported by related works investigating LLMs in software development [3, 23, 46, 71]. This highlights the importance of focusing on the tangible benefits and seamless integration of LLMs rather than solely on their ease of use when promoting these tools to software engineers.

SI was found to be non-significant in its relationship with *BI to use LLMs in software engineering*. This result indicates that the opinions and behaviors of colleagues and peers do not strongly impact an engineer's intention to adopt LLMs and is consistent with the work of Russo [71]. One possible justification is that software engineers tend to base technology adoption decisions more on their individual assessment of a tool's utility and performance rather than on social factors. This is further explained by the Gartner Hype Cycle's latest report [62], which indicates that generative AI is still in its nascent stage and perceived as a disruptive technology. Consequently, individual perception plays a more significant role in the intention to adopt such technology. Based on this

Table 8. Cultural Values—Summary of Findings and Implications

Hypotheses	Findings	Implications
H9: Uncertainty Avoidance moderates PE → BI, SI → BI, and BI → UB	Not supported. The relationships are not significant.	The tendency of developers to prefer or avoid uncertainty does not moderate the intention to adopt or the actual use of LLMs in software engineering, suggesting that the disruptive nature of LLMs and their potential to significantly enhance perceived productivity diminishes the impact of individual uncertainty preferences on their adoption.
H10: Long Term Orientation moderates HM → BI, PE → BI, and BI → UB	Not supported. The relationships are not significant.	The LTO of developers does not act as a moderator for the intention to adopt or the actual use of LLMs in software engineering, suggesting that the immediate and practical benefits of LLMs, such as enhanced performance and usability, take precedence over long-term cultural values in decision-making.
H11: Collectivism moderates PE → BI, EE → BI, SI → BI, and BI → UB	Not supported. The relationships are not significant.	Collectivistic or individualistic tendencies do not moderate the intention to adopt or the actual use of LLMs in software engineering, suggesting that in this field, the practical benefits and enhancements to individual performance and efficiency provided by LLMs overshadow the influence of social and cultural contexts on adoption decisions.
H12: Power Distance moderates SI → BI and BI → UB	Not supported. The relationships are not significant.	A tendency to rigidly follow hierarchical structures and orders does not moderate the intention to adopt or the actual use of LLMs in software engineering, suggesting that the disruptive nature of LLMs, which significantly enhances perceived productivity, diminishes the impact of hierarchical dynamics on adoption. Moreover, the early stage of generative AI technology may explain the weak influence of hierarchical structures on its adoption.
H13: Masculinity moderates PE → BI and BI → UB	Not supported. The relationships are not significant.	The lack of significant moderation by masculinity on the relationships suggests that cultural orientations emphasizing competitiveness (high MAS) or collaboration (low MAS) do not play a decisive role in the adoption or use of LLMs for software engineering. This indicates that LLM adoption is likely driven by universal practical benefits, such as productivity gains, that transcend cultural predispositions toward assertiveness or cooperation. These findings highlight that, in this context, the utility of LLMs may override cultural values related to competitiveness or collaboration.

findings, efforts to promote LLM adoption should focus on demonstrating tangible benefits and improving individual user experiences rather than relying on SI.

Similarly to the previous factor, SI, *HM* was found to be non-significant in its relationship with *BI* for LLMs in software engineering. This result suggests that the pleasure or enjoyment derived from using LLMs does not strongly influence engineers' intentions to adopt these tools. This aligns with related research noting that HB and performance perception matter most [3, 23, 46]. One possible explanation is that software engineers prioritize practical and performance-related aspects of tools over the enjoyment factor. Given the professional context and focus on efficiency and productivity, engineers may value tools that enhance job performance and streamline workflows rather than those that provide hedonic satisfaction. Therefore, promoting LLM adoption in this field should emphasize practical advantages and performance enhancements rather than enjoyment.

6.1.2 On the Moderating Effect of Cultural Values. In this section, we focus on the significance of the cultural values constructs treated as moderators in the UTAUT2 model. Table 8 summarizes our findings.

Power Distance was hypothesized to moderate the relationship between SI and BI, as well as the relationship between BI and UB. However, this moderating effect was not significant. One

primary reason is that the impact of SI on the intention to adopt LLMs is negligible, thus rendering any potential moderating effect insignificant. Additionally, the disruptive nature of LLMs, which significantly enhances perceived productivity [3, 46], diminishes the impact of hierarchical dynamics on adoption. Moreover, the nascent stage of generative AI technology places it in the early phase of the Gartner Hype Cycle, as indicated in the 2023 Gartner report on the state of Artificial Intelligence [62]. During this phase, the technology experiences a peak of inflated expectations, which may explain the weak influence of social factors, such as SI and Power Distance, on its adoption.

LTO was hypothesized to moderate the relationships between PE and BI, HM and BI, and BI and UB. Nevertheless, the moderating effects were not significant. This outcome can be attributed to the immediate and practical focus of software engineering, where decisions are driven more by present benefits and usability rather than long-term cultural values. Engineers prioritize tools that offer immediate performance enhancements and practical benefits, such as LLMs, over considerations influenced by LTO. Furthermore, the rapid evolution of software development, with engineers at the forefront of adopting new technology, further justifies the absence of any moderating effect. The non-significant influence of HM also supports this result. Additionally, focusing on the PE construct reveals its crucial impact on the BI to use LLMs. As LLMs are disruptive and currently perceived as proving themselves as outstanding working companions, the long-term cultural perspectives of practitioners take a secondary role, unable to compete with the immediate benefits provided by such technology [3, 46].

COL was hypothesized to moderate the relationships between three independent variables (PE, EE, and SI) with BI, as well as between BI and UB. However, the moderating effects were not significant. This can largely be explained by the early stage and disruptive nature of LLMs [62], which are perceived as highly useful [3, 46]. In such a context, whether individuals lean towards COL or individualism may not significantly impact adoption decisions. This is further supported by the non-significance of SI, a variable closely related to collectivist values. In a field like software engineering, where personal expertise and the practical benefits of tools like LLMs are paramount, the collective social context has less impact on adoption decisions. Engineers are more likely to focus on how LLMs enhance their individual performance and efficiency, even if they value community and teamwork.

MAS was hypothesized to moderate the relationship between PE and BI, as well as between BI and UB. However, the moderating effects were not significant. This finding aligns with broader research suggesting that this cultural dimension does not play a concrete role in technology adoption [7, 53]. In environments characterized by high-performance expectations, both masculine (competitive, achievement-oriented) and feminine (collaborative, quality-of-life-oriented) values converge on the importance of effectiveness and efficiency in tool adoption. Furthermore, the disruptive and innovative nature of LLMs, which offer substantial improvements in productivity and problem-solving capabilities [3, 46], likely overshadows any cultural predispositions related to MAS or femininity. In summary, the lack of a significant moderating effect of MAS reinforces the idea that practical considerations and the inherent advantages of LLMs are the primary drivers of adoption in the field of software engineering [3, 23, 46, 71], rather than cultural values related to competitive preferences.

UAI was hypothesized to moderate the relationships between PE and BI, SI and BI, as well as between BI and UB. However, the moderating effects were not significant. This result can be explained by several factors specific to the context of software engineering and the nature of LLMs. In the fast-evolving field of software engineering, practitioners are often accustomed to dealing with high levels of uncertainty and rapid technological changes. This professional culture may reduce the impact of individual differences in Uncertainty Avoidance on technology adoption decisions. Moreover, the perceived significant improvements in efficiency and performance offered by LLMs

can mitigate concerns associated with uncertainty [3, 46]. When the advantages of adopting a disruptive technology are (perceived) evident, the predisposition to avoid uncertainty may have less influence on the decision-making process.

6.2 Implications for Practice and Researcher

The study aimed to explore the role of cultural values in the adoption of LLMs in software engineering, guided by the UTAUT2 and Hofstede's cultural dimensions. The results revealed two main points:

- *Cultural Values are not significant in moderating the adoption of LLMs.* The study found that cultural values do not significantly impact the intention to adopt or the actual adoption of LLMs among software engineers.
- *Traditional adoption factors are outstanding protagonists in explaining the adoption of LLMs.* Within the UTAUT2 framework, several factors were identified as significantly influencing the adoption of LLMs, such as HB (the tendency to use LLMs as part of regular work activities), PE (the belief that using LLMs will improve job performance), FC (the availability of resources and support to use LLMs), and BI (the intention to use LLMs in the future).

The implications of these findings are twofold. Firstly, they provide actionable insights for organizations seeking to implement LLMs effectively by highlighting the importance of fostering habitual use, demonstrating performance benefits, ensuring adequate resources and support, and encouraging positive use intentions among practitioners. Secondly, they challenge the assumption that cultural values significantly influence technology adoption in this context, shifting the focus towards other influential factors identified within the UTAUT2 framework.

6.2.1 Implications for Practitioners. The factors within the UTAUT2 framework that showed significant impact highlight practical considerations for organizations aiming to encourage LLM adoption.

The primary and most significant implication of our study is the central role that performance expectations play in the intention to adopt LLMs in software engineering. This finding underscores the necessity for organizations aiming to incorporate LLMs into their workflows to demonstrate and monitor the tangible benefits of LLMs in enhancing job performance. Based on our results, we recommend the following actionable insights:

- *Identify, Implement, and Monitor Performance Metrics*: Managers should identify and establish clear performance metrics to measure the impact of LLMs on software engineering tasks. These metrics should be monitored over time to regularly review and communicate ongoing improvements and areas needing optimization to the team. Furthermore, as suggested by the literature [46, 54], the importance of specific tasks varies among developers. Therefore, organizations should link the metrics to specific tasks to enhance communication with the team.
- *Adapt and Support Training*: In line with the previous point, organizations should regularly gather feedback from users to understand their experiences and identify any challenges or additional needs. Based on this feedback, organizations should adapt training materials for newcomers to support the adoption of the tool from the onboarding phase.
- *Highlight Tangible Benefits*: Related to the previous two points, organizations should create comprehensive reports and presentations that detail improvements in efficiency, accuracy, and overall productivity resulting from LLM adoption. Additionally, as suggested by related literature [46, 54] and mentioned before, organizations can categorize these improvements

by specific tasks (e.g., requirements elicitation, code completion, and decision support) to attract developers' interest based on the areas they find most important.

- *Promote Success Stories*: Organizations should publicize successful projects and individual achievements resulting from the use of LLMs to build enthusiasm and confidence across the organization.

Our investigation also highlights the importance of habitual use of LLMs in supporting both the intention to use and the actual use of these tools in the software engineering domain. This result aligns with previous work in the field [3, 23]. Therefore, we recommend the following actionable insights:

- *Encourage Regular Usage*: Organizations and managers should integrate LLM usage into daily workflows to help employees develop an HB. This can be done by facilitating access and integration of these tools in the everyday working activity. Nevertheless, organizations should take into consideration ethical concerns. For instance, the Copenhagen Manifesto on the adoption of generative AI in software engineering [72] claims particular attention to the ethical implications and the centric role of human behaviors in such adoption. Organizations are suggested to consider the manifesto principles while integrating LLMs into their working environment.
- *Provide Continuous Training*: Related to the previous discussion on the PE, organizations should offer continuous training sessions that reinforce the usage of LLMs over time. Moreover, managers should schedule periodic refresher courses to keep employees updated on new features and best practices for using LLMs. All these activities should take into account the feedback and metrics collected during the development lifecycle on the technology usage.
- *Support a Community of Exchange*: Organizations should establish user communities where employees can share tips, experiences, and support each other in using LLMs. Managers should encourage participation in these communities to foster a culture of collaboration and habitual use.

As an ulterior result, facilitating access to the technology arose as an important factor in the actual adoption of LLMs for software engineering. All the insights proposed so far contribute to a broader effort to facilitate technology adoption by companies. As an ulterior actionable insight, we recommend the following:

- *Invest in Building In-House Teams*: The findings of this study highlight the significance of HB and PE as key drivers of LLM adoption. To foster regular and effective use of LLMs, organizations must establish a dedicated in-house team responsible for implementing, managing, and optimizing the integration of LLMs into daily workflows. This team can serve as a central hub for identifying opportunities to enhance performance through LLMs, providing targeted training to cultivate habitual use, and ensuring that metrics for productivity and efficiency improvements are tracked and reported systematically. By assembling such a team, organizations can mitigate challenges associated with ad hoc implementation, maximize the benefits of LLMs, and align adoption strategies with their specific operational contexts.

To sum up, organizations aiming to implement LLMs effectively should start by integrating them into everyday tasks to develop habitual use. Moreover, it is crucial to highlight the performance benefits by clearly demonstrating how LLMs can enhance productivity and job performance. Furthermore, providing comprehensive support and resources, such as training and technical assistance, will facilitate their use. Additionally, encouraging positive intentions through motivational strategies, incentives, and continuous engagement can foster a positive attitude toward adopting LLMs.

6.2.2 Implications for Researchers. This study contributes to the understanding of technology adoption by highlighting that cultural values may not play a significant role in moderating the adoption of LLMs in software engineering. Instead, individual and organizational factors, as outlined in the UTAUT2 framework, are more influential. Researchers could:

- *Deepening the Role of Cultural Values in Technology Adoption*: This study challenges the assumption that cultural values play a significant role in the adoption of disruptive technologies like LLMs. Nevertheless, these results can be influenced by a plethora of aspects. Other than conducting longitudinal studies to see if culture starts to be significant (maybe when LLMs will reach the next phases of the Gartner lifecycle), researchers could try to use different scales to measure culture. In this study, we relied on the well-known scale of Yoo because it is largely recognized in the community. Still, another scale like the one Sharma [75] exists that measures and represents cultural values with a broader range of constructs [81].
- *Focus on UTAUT2 Constructs*: The findings highlight the dominance of UTAUT2 constructs such as HB, PE, FC, and BI in explaining the adoption of LLMs. Future research should continue to explore and refine these constructs, considering their practical implications and how they can be leveraged to enhance technology adoption.
- *Conduct Longitudinal Studies on HB Formation*: The significant role of HB in LLM adoption indicates a need for longitudinal studies to understand how habitual use of new technologies develops over time. Researchers should explore the processes and interventions that can effectively transform initial use into sustained, routine behavior.
- *Develop Frameworks and Tools*: Researchers should focus on developing practical frameworks and guidelines based on UTAUT2 constructs to assist organizations in implementing new technologies. These frameworks should provide actionable strategies for fostering habitual use, demonstrating performance benefits, ensuring adequate resources, and encouraging positive use intentions. Moreover, researchers could take the challenge of developing a recommendation system to support managerial figures in the process of integrating LLMs into the development workflow without damaging the work environment.
- *Study Trust-Related Factors*: Khojah et al. [46] found that professional reliance on LLMs for guidance and problem-solving depends heavily on the trustworthiness of these tools. Future research could extend the findings of this study by incorporating trust-related factors, such as source trustworthiness, privacy, security, and risks like hallucination and misinformation. These aspects are especially relevant in the context of LLM adoption, where user trust in the system's outputs can significantly influence adoption behaviors. While the current study was limited to the core constructs of UTAUT2 and Hofstede's cultural dimensions to maintain model parsimony and address the challenges associated with overly complex models, future research could build on this foundation to explore the interplay between trust and cultural values in technology adoption.

By addressing these implications, the research community can advance the understanding of technology adoption in specialized fields and contribute to more effective implementation strategies for emerging technologies.

6.3 Threats to Validity and Limitations

As previously mentioned, our study primarily involved quantitative investigations supported by statistical analysis. Consequently, we discuss the threats to validity by examining statistical conclusion validity, internal validity, construct validity, and external validity as defined by Wohlin et al. [91]. Moreover, since we involved human participants, we also discussed reliability as a potential threat.

6.3.1 Conclusion Validity. Conclusion Validity pertains to limitations that affect the ability to draw conclusions about the relationships between the independent and dependent variables [91]. Threats in this category are mainly associated with the tests employed for the analysis. To mitigate this threat, we heavily relied on PLS-SEM statistics, known for their robustness across various scenarios. We meticulously followed the process described by Hair et al. [31] in their comprehensive work on using PLS-SEM. Additionally, we used SmartPLS, a state-of-the-art software utilized in over 1,000 peer-reviewed articles [31].

6.3.2 Internal Validity. Internal validity concerns the risk that the investigation did not account for factors influencing the dependent variable [91]. To address this, we relied on solid and well-validated theories. Given our focus on technology adoption, we utilized the UTAUT2, which is well-suited for investigating similar phenomena [86]. For studying cultural values, we employed the Hofstede cultural framework [37, 38], a widely adopted instrument operationalized for this purpose. Furthermore, we applied a set of filters to our participant sample to ensure it represented the target population while maintaining heterogeneity.

As an ulterior threat, the integration of UTAUT2 and Hofstede's cultural dimensions required limiting the inclusion of additional confounding factors—such as software engineering experience and trust—to maintain model parsimony. While this approach ensured clarity in addressing the primary research question, it represents a limitation in internal validity. Future work should incorporate these factors to validate the findings further.

One limitation of this study is the potential for self-selection bias. Participants with a particular interest in or skepticism about LLMs may have been more inclined to participate, potentially resulting in a sample that is not fully representative of the broader software engineering population. While the survey design aimed to mitigate such biases, this limitation should be considered when interpreting the results.

6.3.3 Construct Validity. Construct validity addresses the measurement and tools used to represent the study variables [91]. All variables in our theoretical model were measured using validated instruments [86, 93]. Specifically, for measuring culture, we avoided the common error of using country of origin combined with Hofstede's values. Instead, we relied on scales designed to measure culture at the individual level [93]. We designed the questionnaires following the most recent guidelines in the field and employed various strategies, such as question randomization and attention-check questions, to enhance the reliability of the results [19, 47, 65].

The decision to adopt LLMs and the moderating factors influencing this adoption may vary significantly across specific work contexts and tasks (e.g., requirements elicitation vs. code understanding). This study provides a theoretical starting point for understanding LLM adoption in software development but does not account for the potential contextual variations. Future research should explore these differences in depth to refine and contextualize the findings presented here.

6.3.4 External Validity. External validity concerns the generalizability of the findings to the entire population [91]. We filtered the entire Prolific population to ensure that participants met the characteristics ideal for our investigation. Additionally, we collected a sufficient number of data according to G*Power recommendations [28]. However, generalizability is a complex issue. Given that the technology is still in its early stages [62], we anticipate that different results might emerge in the future.

6.3.5 Reliability of Participants. Ensuring the reliability of participants is a critical aspect of survey-based research, as unreliable respondents may introduce noise, compromising the validity of the results. In this study, multiple measures were implemented to mitigate this threat.

First, participants were recruited through Prolific, which allows researchers to apply strict selection criteria, including a 100% prior approval rate in other studies, reducing the likelihood of unreliable responses [22, 24]. Moreover, a pre-screening survey was conducted to verify participants' background, ensuring they met the required criteria of software engineering expertise and prior experience with LLMs in a professional context. To further control for data quality and participant reliability, we incorporated validated programming knowledge questions [19], ensuring that participants had sufficient technical expertise.

Second, attention-check questions were (iteratively) embedded within the surveys to identify inattentive or random responses. These checks are a well-established method for improving data reliability by filtering out inattentive participants. Responses failing these checks were discarded.

Third, we leveraged Qualtrics' built-in detection mechanisms to identify and filter out bot-generated responses, response inconsistencies, and suspicious completion times. Finally, an iterative pilot study was conducted prior to deployment to assess comprehension and survey duration [47], ensuring that the instrument was appropriately designed.

☞ UTAUT2 and Cultural Values for LLMs in Software Engineering: Summary of the discussion.

This study investigated the impact of cultural values as moderator in the adoption of LLMs in software engineering using the UTAUT2 framework. The research revealed that cultural values did not significantly moderate the intention to adopt or the actual adoption of LLMs. Instead, the UTAUT2 constructs—HB, PE, FC, and BI—were the key factors affecting adoption. PE was the most influential, emphasizing the importance of perceived benefits in adopting new technologies. HB also played a crucial role, indicating that once LLMs become part of routine work, their continued use is likely. FC, such as resources and support, were critical for actual use, while BI directly influenced adoption. For practitioners, the study suggests focusing on integrating LLMs into daily workflows, showcasing their performance benefits, ensuring robust support and resources, and fostering positive use intentions through training and awareness programs. For researchers, the findings highlight the need to explore UTAUT2 constructs further, particularly HB formation, and to develop practical frameworks for technology adoption. Additionally, the study challenges the assumption that cultural values significantly impact technology adoption in this context, suggesting a shift in focus toward individual and organizational factors.

7 Conclusions

The objective of this study was to explore the adoption of LLMs in software development, with a focus on the potential moderating effects of practitioners' cultural values. We concentrated first on typical factors influencing technology adoption, as derived from the UTAUT2, and second on cultural values as defined by Hofstede.

Our statistical analysis, conducted using PLS-SEM, revealed that HB and perceived performance are the primary drivers of integrating LLMs into daily routines. In contrast, the moderating effect of cultural values was not significant. This outcome can be attributed to the complexity of software development, which is heavily influenced by evolving socio-technical factors and the increasing demand for rapid release cycles. Furthermore, the disruptive and highly beneficial nature of LLMs likely overshadows cultural influences, providing a clearer justification for our findings. In addition, our results align with what was identified by the HACAF [71].

For practitioners, our findings have direct and practical implications. They suggest integrating LLMs into daily workflows, clearly demonstrating their performance benefits, providing comprehensive support and resources, and fostering positive use intentions through motivational strategies. The findings also underscore the need for further research to explore cultural values, conduct longitudinal studies on HB formation, and develop practical frameworks and tools to assist organizations in effectively implementing new technologies like LLMs.

For our future research agenda, we plan to explore additional ways to integrate cultural values into the technology adoption of LLMs in software engineering, beyond using the extended UTAUT2 framework, which our research has validated. Additionally, we aim to investigate alternative scales to the CVSCALE proposed by Yoo et al. [93] to measure cultural values, as differences, though minimal, can exist among various representations [81]. Finally, we advocate for longitudinal studies to understand how the BI of software engineers evolves as the technology matures and moves beyond the early adoption stage [62].

Data Availability

All the data and materials used in this research are publicly available in our online appendix on Figshare [50].

Acknowledgements

The authors thank the questionnaire participants for their valuable responses and for providing the essential data required to conduct this research. The authors thank the reviewers and the editor for their valuable comments and suggestions, which have significantly contributed to enhancing the quality of our article.

References

[1] Muhammad Sharif Abbasi, Ali Tarhini, Tariq Elyas, and Farwa Shah. 2015. Impact of individualism and collectivism over the individual's technology acceptance behaviour: A multi-group analysis between Pakistan and Turkey. *Journal of Enterprise Information Management* 28, 6 (2015), 747–768.

[2] Sameer Abufardeh and Kenneth Magel. 2010. The impact of global software cultural and linguistic aspects on Global Software Development process (GSD): Issues and challenges. In *Proceedings of the 4th International Conference on New Trends in Information Science and Service Science*. IEEE, 133–138.

[3] Alexandre Agossah, Frédérique Krupa, Matthieu Perreira Da Silva, and Patrick Le Callet. 2023. LLM-based interaction for content generation: A case study on the perception of employees in an IT department. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*, 237–241.

[4] Hisham Alhirz and A. S. M. Sajeev. 2015. Do cultural dimensions differentiate ERP acceptance? A study in the context of Saudi Arabia. *Information Technology & People* 28, 1 (2015), 163–194.

[5] Dorine Andrews, Blair Nonnecke, and Jennifer Preece. 2007. Conducting research on the Internet: Online survey design, development and implementation guidelines (2007).

[6] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, H. Michalewski, David Dohan, Ellen Jiang, Carrie J. Cai, Michael Terry, Quoc V. Le, et al. 2021. Program synthesis with large language models. arXiv:2108.07732. Retrieved from https://arxiv.org/abs/2108.07732

[7] Gonçalo Baptista and Tiago Oliveira. 2015. Understanding mobile banking: The unified theory of acceptance and use of technology combined with cultural moderators. *Computers in Human Behavior* 50 (2015), 418–430.

[8] Shraddha Barke, Michael B. James, and Nadia Polikarpova. 2023. Grounded copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages* 7, OOPSLA1 (2023), 85–111.

[9] Paheli Bhattacharya, Manojit Chakraborty, Kartheek N. S. N. Palepu, Vikas Pandey, Ishan Dindorkar, Rakesh Rajpurohit, and Rishabh Gupta. 2023. Exploring large language models for code explanation. arXiv:2310.16673. Retrieved from https://arxiv.org/abs/2310.16673

[10] Greg Borchers. 2003. The software engineering impacts of cultural factors on multi-cultural software development teams. In *Proceedings of the 25th International Conference on Software Engineering*. IEEE, 540–545.

[11] Frederick P. Brooks Jr. 1995. *The Mythical Man-Month: Essays on Software Engineering*. Pearson Education.

[12] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 33, 1877–1901.
[13] Valentine Casey. 2011. Imparting the importance of culture to global software development. *ACM Inroads* 1, 3 (2011), 51–57.
[14] Marcelo Cataldo, James D. Herbsleb, and Kathleen M. Carley. 2008. Socio-technical congruence: A framework for assessing the impact of technical and work dependencies on software development productivity. In *Proceedings of the 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, 2–11.
[15] Jagdeep S. Chhokar, Felix C. Brodbeck, and Robert J. House. 2007. *Culture and Leadership Across the World: The GLOBE Book of In-Depth Studies of 25 Societies*. Routledge.
[16] Terry L. Childers, Christopher L. Carr, Joann Peck, and Stephen Carson. 2001. Hedonic and utilitarian motivations for online retail shopping behavior. *Journal of Retailing* 77, 4 (2001), 511–535.
[17] Wynne W. Chin. 1998. The partial least squares approach to structural equation modeling. *Modern Methods for Business Research* 295, 2 (1998), 295–336.
[18] Deborah R. Compeau and Christopher A. Higgins. 1995. Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly* 19 (1995), 189–211.
[19] Anastasia Danilova, Alena Naiakshina, Stefan Horstmann, and Matthew Smith. 2021. Do you really code? Designing and evaluating screening questions for online surveys with programmers. In *Proceedings of the IEEE/ACM 43rd International Conference on Software Engineering (ICSE '21)*. IEEE, 537–548.
[20] Fred D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13 (1989), 319–340.
[21] Fred D. Davis, R. P. Bagozzi, and P. R. Warshaw. 1989. Technology acceptance model. *Journal of Management Sciences* 35, 8 (1989), 982–1003.
[22] Benjamin D. Douglas, Patrick J. Ewell, and Markus Brauer. 2023. Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS One* 18, 3 (2023), e0279720.
[23] Fiona Draxler, Daniel Buschek, Mikke Tavast, Perttu Hämäläinen, Albrecht Schmidt, Juhi Kulshrestha, and Robin Welsch. 2023. Gender, age, and technology education influence the adoption and appropriation of LLMs. arXiv:2310.06556. Retrieved from https://arxiv.org/abs/2310.06556
[24] PeerEyal, RothschildDavid, GordonAndrew, EverndenZak, and DamerEkaterina. 2021. Data quality of platforms and panels for online behavioral research. *Behavior Research Methods (*2021*), 1–20*.
[25] Tony Fang. 2012. Yin Yang: A new perspective on culture. *Management and Organization Review* 8, 1 (2012), 25–50.
[26] Yanping Fang and S. Gopinathan. 2009. Teachers and teaching in Eastern and Western schools: A critical review of cross-cultural comparative studies. In *International Handbook of Research on Teachers and Teaching*. L. J. Saha and A. G. Dworkin (Eds.), Springer, 557–572.
[27] Khaled M. S. Faqih and Mohammed-Issa Riad Mousa Jaradat. 2015. Assessing the moderating effect of gender differences and individualism-collectivism at individual-level on the adoption of mobile commerce technology: TAM3 perspective. *Journal of Retailing and Consumer Services* 22 (2015), 37–52.
[28] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (2009), 1149–1160.
[29] Leandro Carísio Fernandes. 2024. Programming computational electromagnetic applications assisted by large language models [Em Programmer's Notebook]. *IEEE Antennas and Propagation Magazine* 66, 1 (2024), 63–71.
[30] Dianne P. Ford, Catherine E. Connelly, and Darren B. Meister. 2003. Information systems research and Hofstede's culture's consequences: An uneasy and incomplete partnership. *IEEE Transactions on Engineering Management* 50, 1 (2003), 8–25.
[31] Joseph F. Hair Junior, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt. 2014. *A Primer on Partial Least Squares Structural Equation Modeling*.
[32] M. A. Hameed, S. Counsell, and S. Swift. 2012. A meta-analysis of relationships between organizational characteristics and IT innovation adoption in organizations. *Information and Management* 49 (2012), 218–232. DOI: https://doi.org/10.1016/j.im.2012.05.002
[33] Jörg Henseler, Theo K. Dijkstra, Marko Sarstedt, Christian M. Ringle, Adamantios Diamantopoulos, Detmar W. Straub, David J. Ketchen Jr, Joseph F. Hair, G. Tomas, M. Hult, and Roger J. Calantone. 2014. Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods* 17, 2 (2014), 182–209.
[34] Jörg Henseler, Christian M. Ringle, and Marko Sarstedt. 2015. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science* 43 (2015), 115–135.
[35] Rashina Hoda. 2021. Socio-technical grounded theory for software engineering. *IEEE Transactions on Software Engineering* 48, 10 (2021), 3808–3832.

[36] Hartmut Hoehle, Xiaojun Zhang, and Viswanath Venkatesh. 2015. An espoused cultural perspective to understand continued intention to use mobile applications: A four-country study of mobile social media application usability. *European Journal of Information Systems* 24, 3 (2015), 337–359.

[37] Geert Hofstede. 1980. Culture and organizations. *International Studies of Management & Organization* 10, 4 (1980), 15–41.

[38] Geert Hofstede. 2011. Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture* 2, 1 (2011), 8.

[39] Robert J. House, Peter W. Dorfman, Mansour Javidan, Paul J. Hanges, and Mary F. Sully De Luque. 2013. *Strategic Leadership Across Cultures: GLOBE Study of CEO Leadership Behavior and Effectiveness in 24 Countries*. Sage Publications.

[40] Robert J. House, Paul J. Hanges, Mansour Javidan, Peter W. Dorfman, and Vipin Gupta. 2004. *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Sage Publications.

[41] Yujong Hwang and Kun Chang Lee. 2012. Investigating the moderating role of uncertainty avoidance cultural values on multidimensional online trust. *Information & Management* 49, 3–4 (2012), 171–176.

[42] Il Im, Seongtae Hong, and Myung Soo Kang. 2011. An international comparison of technology adoption: Testing the UTAUT model. *Information & Management* 48, 1 (2011), 1–8.

[43] Juyong Jiang, Fan Wang, Jiasi Shen, Sungju Kim, and Sunghun Kim. 2024. A survey on large language models for code generation. arXiv:2406.00515. Retrieved from https://arxiv.org/abs/2406.00515

[44] Xue Jiang, Yihong Dong, Lecheng Wang, Qiwei Shang, and Ge Li. 2024. Self-planning code generation with large language models. *ACM Transactions on Software Engineering and Methodology* 33, 7 (2024), 1–30. DOI: https://doi.org/10.1145/3672456

[45] Junfeng Jiao, Saleh Afroogh, Yiming Xu, and Connor Phillips. 2024. Navigating LLM ethics: Advancements, challenges, and future directions. arXiv:2406.18841. Retrieved from https://arxiv.org/abs/2406.18841

[46] Ranim Khojah, Mazen Mohamad, Philipp Leitner, and Francisco Gomes de Oliveira Neto. 2024. Beyond code generation: An observational study of ChatGPT usage in software engineering practice. *Proceedings of the ACM on Software Engineering* 1, FSE (2024), 1819–1840.

[47] Barbara A. Kitchenham and Shari L. Pfleeger. 2008. Personal opinion surveys. In *Guide to Advanced Empirical Software Engineering*. Forrest Shull, Janice Singer, and Dag I. K. Sjøberg (Eds.), Springer, 63–92.

[48] Jahnavi Kumar and Sridhar Chimalakonda. 2024. Code summarization without direct access to code-towards exploring federated LLMs for software engineering. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*, 100–109.

[49] Chun Lai, Qiu Wang, Xiaoshi Li, and Xiao Hu. 2016. The influence of individual espoused cultural values on self-directed use of technology for language learning beyond the classroom. *Computers in Human Behavior* 62 (2016), 676–688.

[50] Stefano Lambiase, Gemma Catolino, Fabio Palomba, Filomena Ferrucci, and Daniel Russo. 2024. Investigating the Role of Cultural Values in Adopting Large Language Models for Software Engineering—Online Appendix. DOI: https://doi.org/10.6084/m9.figshare.26376391

[51] Hyunae Lee, Namho Chung, and Timothy Jung. 2015. Examining the cultural differences in acceptance of mobile augmented reality: Comparison of South Korea and Ireland. In *Proceedings of the International Conference in Lugano on Information and Communication Technologies in Tourism 2015*. Springer, 477–491.

[52] Sang-Gun Lee, Silvana Trimi, and Changsoo Kim. 2013. Innovation and imitation effects' dynamics in technology adoption. *Industrial Management & Data Systems* 113, 6 (2013), 772–799.

[53] Hsien-Cheng Lin. 2014. An investigation of the effects of cultural differences on physicians' perceptions of information technology acceptance as they relate to knowledge management systems. *Computers in Human Behavior* 38 (2014), 368–380.

[54] Mingxing Liu, Junfeng Wang, Tao Lin, Quan Ma, Zhiyang Fang, and Yanqun Wu. 2024. An empirical study of the code generation of safety-critical software using LLMs. *Applied Sciences* 14, 3 (2024), 1046.

[55] Hazel RoseMarkus and Shinobu Kitayama. 2014. Culture and the self: Implications for cognition, emotion, and motivation. In *College Student Development and Academic Life*. Routledge, 264–293.

[56] Deep Mehta, Kartik Rawool, Subodh Gujar, and Bowen Xu. 2023. Automated DevOps pipeline generation for code repositories using large language models. arXiv:2312.13225. Retrieved from https://arxiv.org/abs/2312.13225

[57] Christian Munley, Aaron Jarmusch, and Sunita Chandrasekaran. 2024. LLM4VV: Developing LLM-driven testsuite for compiler validation. *Future Generation Computer Systems* 160 (2024), 1–13.

[58] Syed Asad Ali Naqvi, Markus Philipp Zimmer, Rehan Syed, and Paul Drews. 2023. Understanding the socio-technical aspects of low-code adoption for software development. In *Proceedings of the European Conference on Information Systems (ECIS)*.

[59] Michael Neumann, Klaus Schmid, and Lars Baumann. 2023. Characterizing the impact of culture on agile methods: The moCA model. In *Proceedings of the IEEE/ACM International Conference on Software and System Processes (ICSSP '23)*. IEEE, 81–85.

[60] Michael Neumann, Klaus Schmid, and Lars Baumann. 2024. What you use is what you get: Unforced errors in studying cultural aspects in agile software development. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering (EASE '24)*. ACM, New York, NY, 405–410. DOI: https://doi.org/10.1145/3661167.3661229

[61] Nicolae Nistor, Thomas Lerche, Armin Weinberger, Ciprian Ceobanu, and Oliver Heymann. 2014. Towards the integration of culture into the unified theory of acceptance and use of technology. *British Journal of Educational Technology* 45, 1 (2014), 36–55.

[62] Lori Perri. 2023. What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle. Retrieved from https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle

[63] Andrés Piñeiro-Martín, Carmen García-Mateo, Laura Docío-Fernández, and Maria Del Carmen Lopez-Perez. 2023. Ethical challenges in the development of virtual assistants powered by large language models. *Electronics* 12, 14 (2023), 3170.

[64] Sascha Raithel, Marko Sarstedt, Sebastian Scharf, and Manfred Schwaiger. 2012. On the value relevance of customer satisfaction. Multiple drivers and multiple markets. *Journal of the Academy of Marketing Science* 40 (2012), 509–525.

[65] Paul Ralph, Nauman bin Ali, Sebastian Baltes, Domenico Bianculli, Jessica Diaz, Yvonne Dittrich, Neil Ernst, Michael Felderer, Robert Feldt, Antonio Filieri, et al. 2020. Empirical standards for software engineering research. arXiv:2010.03525. Retrieved from https://arxiv.org/abs/2010.03525

[66] Paul Ralph, Mike Chiasson, and Helen Kelley. 2016. Social theory for software engineering research. In *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE '16)*. ACM, New York, NY, Article 44, 11 pages. DOI: https://doi.org/10.1145/2915970.2915998

[67] Sanka Rasnayaka, Guanlin Wang, Ridwan Shariffdeen, and Ganesh Neelakanta Iyer. 2024. An empirical study on usage and perceptions of LLMs in a software engineering project. In *Proceedings of the 1st International Workshop on Large Language Models for Code*, 111–118.

[68] Christian M. Ringle, Sven Wende, and Jan-Michael Becker. 2024. SmartPLS 4. Retrieved from https://www.smartpls.com/

[69] Steven I. Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D. Weisz. 2023. The programmer's assistant: Conversational interaction with a large language model for software development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 491–514.

[70] Daniel Russo. 2022. Recruiting software engineers on prolific. arXiv:2203.14695. Retrieved from https://arxiv.org/abs/2203.14695

[71] Daniel Russo. 2024. Navigating the complexity of generative AI adoption in software engineering. *ACM Transactions on Software Engineering and Methodology* 33, 5 (2024), 135:1–135:50.

[72] Daniel Russo, Sebastian Baltes, Niels van Berkel, Paris Avgeriou, Fabio Calefato, Beatriz Cabrero-Daniel, Gemma Catolino, Jurgen Cito, Neil Ernst, Thomas Fritz, et al. 2024. Generative AI in software engineering must be human-centered: The Copenhagen manifesto. *Journal of Systems and Software* 216 (2024), 112115.

[73] Daniel Russo and Klaas-JanStol. 2021. PLS-SEM for software engineering research: An introduction and survey. *ACM Computing Surveys* 54, 4 (2021), 1–38.

[74] Francesca Sanguineti and Thomas Maran. 2024. Navigating the clouds: Cultural influences on technology acceptance in the workplace. *Technology in Society* 77 (2024), 102553.

[75] Piyush Sharma. 2010. Measuring personal cultural orientations: Scale development and validation. *Journal of the Academy of Marketing Science* 38 (2010), 787–806.

[76] Galit Shmueli, Soumya Ray, Juan Manuel Velasquez Estrada, and Suneel Babu Chatla. 2016. The elephant in the room: Predictive performance of PLS models. *Journal of Business Research* 69, 10 (2016), 4552–4564.

[77] Galit Shmueli, Marko Sarstedt, Joseph F. Hair, Jun-Hwa Cheah, Hiram Ting, Santha Vaithilingam, and Christian M. Ringle. 2019. Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing* 53, 11 (2019), 2322–2347.

[78] Mark Srite and Elena Karahanna. 2006. The role of espoused national cultural values in technology acceptance. *MIS Quarterly* (2006), 679–704.

[79] Detmar Straub, Mark Keil, and Walter Brenner. 1997. Testing the technology acceptance model across cultures: A three country study. *Information & Management* 33, 1 (1997), 1–11.

[80] Kuttimani Tamilmani, Nripendra P. Rana, Samuel Fosso Wamba, and Rohita Dwivedi. 2021. The extended Unified Theory of Acceptance and Use of Technology (UTAUT2): A systematic literature review and theory evaluation. *International Journal of Information Management* 57 (2021), 102269.

[81] Vas Taras, Piers Steel, and Madelynn Stackhouse. 2023. A comparative evaluation of seven instruments for measuring values comprising Hofstede's model of culture. *Journal of World Business* 58, 1 (2023), 101386.

[82] Ronald L. Thompson, Christopher A. Higgins, and Jane M. Howell. 1991. Personal computing: Toward a conceptual model of utilization. *MIS Quarterly* 15 (1991), 125–143.

[83] Hans Van der Heijden. 2004. User acceptance of hedonic information systems. *MIS Quarterly* 28 (2004), 695–704.
[84] Viswanath Venkatesh and Michael G. Morris. 2000. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly* 24 (2000), 115–139.
[85] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly* 27, 3 (2003), 425–478. Retrieved from http://www.jstor.org/stable/30036540
[86] ViswanathVenkatesh, James Y. L. Thong, and Xin Xu. 2012. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly* 36, 1 (2012), 157–178. Retrieved from http://www.jstor.org/stable/41410412
[87] Viswanath Venkatesh and Xiaojun Zhang. 2010. Unified theory of acceptance and use of technology: US vs. China. *Journal of Global Information Technology Management* 13, 1 (2010), 5–27.
[88] Thomas L. Webb and Paschal Sheeran. 2006. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin* 132, 2 (2006), 249.
[89] I. D. Weerd, Ivonne Sartika Mangula, and S. Brinkkemper. 2016. Adoption of software as a service in Indonesia: Examining the influence of organizational factors. *Information and Management* 53 (2016), 915–928. DOI: https://doi.org/10.1016/j.im.2016.05.008
[90] Bingyang Wei. 2024. Requirements are all you need: From requirements to code with LLMs. In *Proceedings of the IEEE 32nd International Requirements Engineering Conference (RE '24)*, 416–422. DOI: https://doi.org/10.1109/RE59067.2024.00049
[91] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Experimentation in Software Engineering*, Vol. 236. Springer.
[92] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Zhibo Sun, and Yue Zhang. 2024. A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing* 4 (2024), 100211.
[93] Boonghee Yoo, Naveen Donthu, and Tomasz Lenartowicz. 2011. Measuring Hofstede's five dimensions of cultural values at the individual level: Development and validation of CVSCALE. *Journal of International Consumer Marketing* 23, 3–4 (2011), 193–210.
[94] Zhen Zeng, William Watson, Nicole Cho, Saba Rahimi, Shayleen Reynolds, Tucker Balch, and Manuela Veloso. 2023. FlowMind: Automatic workflow generation with LLMs. In *Proceedings of the Fourth ACM International Conference on AI in Finance*, 73–81.
[95] Jianyi Zhang, XuJi, Zhangchi Zhao, Xiali Hei, and Kim-Kwang Raymond Choo. 2023. Ethical considerations and policy implications for large language models: Guiding responsible development and deployment. arXiv:2308.02678. Retrieved from https://arxiv.org/abs/2308.02678
[96] Zhongyun Zhou, Xiao-Ling Jin, Yulin Fang, and Doug Vogel. 2015. Toward a theory of perceived benefits, affective commitment, and continuance intention in social virtual worlds: Cultural values (indulgence and individualism) matter. *European Journal of Information Systems* 24, 3 (2015), 247–261.

Appendices

A Questionnaires Items

Table A1. Items Description

Constructs	Item ID	Questions	Ref.
Power Distance Index	PDI_1	People in higher positions should make most decisions without consulting people in lower positions.	[93]
	PDI_2	People in higher positions should not ask the opinions of people in lower positions too frequently.	[93]
	PDI_3	(*) People in higher positions should avoid social interaction with people in lower positions.	[93]
	PDI_4	People in lower positions should not disagree with decisions by people in higher positions.	[93]
	PDI_5	(*) People in higher positions should not delegate important tasks to people in lower positions.	[93]
Uncertainty Avoidance	UAI_1	(*) It is important to have instructions spelled out in detail so that I always know what I'm expected to do.	[93]
	UAI_2	It is important to closely follow instructions and procedures.	[93]
	UAI_3	(*) Rules and regulations are important because they inform me of what is expected of me.	[93]
	UAI_4	(*) Standardized work procedures are helpful.	[93]
	UAI_5	Instructions for operations are important.	[93]
Collectivism	COL_1	(*) Individuals should sacrifice self-interest for the group.	[93]
	COL_2	Individuals should stick with the group even through difficulties.	[93]
	COL_3	Group welfare is more important than individual rewards.	[93]
	COL_4	Group success is more important than individual success.	[93]
	COL_5	(*) Individuals should only pursue their goals after considering the welfare of the group.	[93]
	COL_6	Group loyalty should be encouraged even if individual goals suffer.	[93]
Masculinity	MAS_1	It is more important for men to have a professional career than it is for women.	[93]
	MAS_2	Men usually solve problems with logical analysis; women usually solve problems with intuition.	[93]
	MAS_3	Solving difficult problems usually requires an active, forcible approach, which is typical of men.	[93]
	MAS_4	(*) There are some jobs that a man can always do better than a woman.	[93]
Long Term Orientation	LTO_1	(*) Careful management of money (Thrift)	[93]
	LTO_2	Going on resolutely in spite of opposition (Persistence)	[93]
	LTO_3	(*) Personal steadiness and stability	[93]
	LTO_4	Long-term planning	[93]
	LTO_5	(*) Giving up today's fun for success in the future	[93]
	LTO_6	(*) Working hard for success in the future	[93]
Performance Expectancy	PE_1	I find the use of LLM useful in supporting my job	[86]
	PE_2	Using LLMs to support my job increases my productivity	[86]
	PE_3	Using LLMs to support my job helps me accomplish things more quickly	[86]
	PE_4	Using LLMs to support my job increases my chances of achieving things that are important to me	[86]
	PE_5	Using LLMs would improve the quality of work I do	[86]

(Continued)

Table A1. Continued

Constructs	Item ID	Questions	Ref.
Effort Expectancy	EE_1	Learning how to use LLMs to support my job is easy for me	[86]
	EE_2	My interaction with LLMs to support my job is clear and understandable	[86]
	EE_3	I find LLMs easy to use in supporting my job	[86]
	EE_4	It is easy for me to become skillful at using LLMs to support my job	[86]
	EE_5	I would find LLMs flexible to interact with during my job	[86]
	EE_6	Using LLMs would make it easier to do my job	[86]
Social Influence	SI_1	People who are important to me think that I should use LLMs to support my job	[86]
	SI_2	People who influence my behavior think that I should use LLMs to support my job	[86]
	SI_3	(*) LLMs use is a status symbol in my job environment	[86]
	SI_4	(*) I use LLMs because of the proportion of coworkers who use the system	[86]
	SI_5	(*) People in my organization who use LLMs have more prestige than those who do not	[86]
Facilitating Condition	FC_1	I have the resources necessary to use LLMs to support my job	[86]
	FC_2	I have the knowledge necessary to use LLMs to support my job	[86]
	FC_3	LLMs are compatible with other technologies I use in my job	[86]
	FC_4	(*) I can get help from others when I have difficulties using LLMs to support my job	[86]
Hedonic Motivation	HM_1	Using LLMs to support my job is fun	[86]
	HM_2	Using LLMs to support my job is enjoyable	[86]
	HM_3	Using LLMs to support my job is entertaining	[86]
Habit	HB_1	The use of LLMs to support my job has become a habit for me	[86]
	HB_2	I am addicted to using LLMs to support my job	[86]
	HB_3	I must use LLMs to support my job	[86]
	HB_4	Using LLMs to support my job has become natural to me	[86]
Price Value	PV_1	LLMs that I use to support my job are reasonably priced	[86]
	PV_2	LLMs that I use to support my job are a good value for the money	[86]
	PV_3	At the current price, LLMs provide a good value	[86]
Behavioral Intention	BI_1	I intend to continue using LLMs to support my job in the future	[86]
	BI_2	I will always try to use LLMs to support my job	[86]
	BI_3	I plan to continue to use LLMs to support my job frequently	[86]

The items prefixed with an asterisk (*) were dropped because of their insufficient loading onto their latent variable.

B Heterotrait–Monotrait Ratio of Correlations (HTMT) for Discriminant Validity

Table B1. HTMT for Discriminant Validity

	COL	EE	FC	HB	HM	LTO	MAS	PE	PDI	SI	UAI	UB	BI
COL													
EE	0.089												
FC	0.092	0.803											
HB	0.090	0.572	0.392										
HM	0.084	0.536	0.282	0.533									
LTO	0.079	0.253	0.166	0.155	0.327								
MAS	0.141	0.079	0.047	0.132	0.069	0.164							
PE	0.097	0.791	0.544	0.753	0.607	0.201	0.083						
PD	0.136	0.223	0.149	0.117	0.307	0.072	0.363	0.258					
SI	0.069	0.424	0.360	0.564	0.362	0.147	0.054	0.579	0.136				
UA	0.344	0.117	0.124	0.063	0.052	0.557	0.097	0.049	0.148	0.072			
UB	0.132	0.499	0.424	0.587	0.305	0.057	0.057	0.530	0.204	0.365	0.025		
BI	0.070	0.671	0.453	0.730	0.591	0.189	0.083	0.817	0.234	0.439	0.058	0.554	

Received 24 August 2024; revised 17 March 2025; accepted 17 March 2025