Simulating Stakeholders: Generative AI Chatbots in Architecture Education

Perceived Usefulness and Diversity Awareness in a Simulated Interview Study

Ruxandra Stoica[1]

Supervisors: Gosia Migut[1]**, Aleksander Buszydlik**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 26, 2026

Name of the student: Ruxandra Stoica
Final project course: CSE3000 Research Project
Thesis committee: Gosia Migut, Aleksander Buszydlik, Masoud Mansoury

Abstract

Generative AI chatbots are increasingly used in engineering education to simulate human stakeholder interactions at scale. While existing studies demonstrate their potential for hands-on learning, the consequences of replacing real human interaction with AI simulation remain largely unexplored. This study investigates the implications of using generative AI chatbots for simulated human interaction on architecture students' awareness of human diversity. A ChatGPT-based chatbot simulating a secondary school teacher in Rotterdam was developed and evaluated with 14 architecture students from TU Delft. Participants completed a 25-minute simulated user interview followed by a mixed-methods evaluation consisting of validated questionnaire scales and open-ended reflection questions. Quantitative data were analyzed using descriptive statistics and qualitative data were analyzed using thematic analysis. Results indicate that students rated the chatbot highly on ease of use but showed more moderate scores on perceived usefulness and professional relevance. Physical mobility dominated diversity consideration (93%), while cultural background and visual impairment were rarely considered (36% and 14% respectively). Qualitative themes suggest that the chatbot functioned as a gap-filler in architecture education and prompted students to consider overlooked user groups, but that this awareness was largely chatbot-driven rather than self-initiated. Students consistently positioned the tool as a useful supplement for study contexts rather than a substitute for real human interaction. These findings contribute to a growing understanding of the pedagogical implications of AI chatbots in human-centered education, and offer preliminary guidance for their responsible use in architecture curricula.

1 Introduction

Generative AI tools have become increasingly present in higher education, to the point where educators are actively exploring how to integrate them into learning activities in meaningful ways (Labadze et al., 2023). One promising direction is the use of AI chatbots to simulate conversations with clients, stakeholders, or domain experts. This offers students a low-cost, scalable way to practice professional skills that would otherwise require real human participants (Dai and Ke, 2022). General-purpose large language models such as ChatGPT are particularly attractive in this context. They require little time or technical expertise to deploy. Custom-built educational simulations can be designed specifically for a learning goal, but they come at a significant development cost. General-purpose models can be configured and shared with minimal effort, making them accessible to educators without specialised technical resources. Recent work has demonstrated promising results in this area: Honig et al. (2024) showed that customized GPT-based chatbots can effectively simulate professional roles in process safety education, with positive effects on student engagement and learning.

The broader case for simulation-based learning in higher education is well established. Chernikova et al. (2020) showed through a meta-analysis that simulation-based learning generally leads to better learning outcomes compared to traditional instruction, particularly for developing complex professional competencies. Dai and Ke (2022) extended this to AI-based simulations specifically, finding a growing body of evidence supporting their use in engineering and health education. These studies suggest that generative AI chatbots are a promising tool for experiential learning. However, they focus primarily on whether chatbots are useful or accepted by students, and say little about what misconceptions students may form in the process. This is a significant gap. If students practice professional interactions with an AI rather than real people, there is a risk that they develop an oversimplified understanding of the humans they are supposed to be designing for. Batzner et al. (2025) support this concern, finding that AI-simulated personas tend to focus on limited sociodemographic attributes and systematically under-represent marginalized groups. A related concern has been raised in UX/UI design, where the use of simplified user personas has been criticized for giving designers a false sense of understanding their users. This can potentially discourage genuine engagement with real and diverse communities (Kernaghan, 2022). Whether a similar dynamic emerges when engineering students use AI chatbots for human interaction simulation remains unexplored.

This question is particularly relevant in architecture education. Architects design spaces that directly shape the daily lives of diverse groups of people, including those with disabilities, elderly residents, and communities of different cultural backgrounds. Architecture students, therefore, need to develop not only technical skills, but also a nuanced awareness of human diversity and the ability to engage meaningfully with people whose experiences differ from their own. If chatbot-based simulations flatten or oversimplify this diversity, they may reinforce blind spots rather than address them. At the same time, if used well, they could offer students a valuable and accessible way to encounter perspectives they would not otherwise meet in a classroom setting.

This project investigates these tensions within the context of architecture education at TU Delft. The main research question is: *What are the implications of using generative AI chatbots for simulated human interaction on architecture students' awareness of human diversity?* Here, implications refers to indirect and potentially unexpected outcomes, such as whether the tool shapes what students notice, limits their thinking, or creates a false sense of understanding.

This is explored through three sub-questions:

1. To what extent do architecture students perceive generative AI chatbots as a useful pedagogical tool?
2. In what ways does interacting with a generative AI chatbot shape or constrain architecture students' awareness of human diversity in their design process?
3. To what extent do architecture students perceive generative AI chatbots as a useful tool for developing awareness of human diversity?

To answer these questions, a generative AI chatbot simulating a human stakeholder was developed and evaluated with 14 architecture students at TU Delft. Participants interacted with the chatbot in a realistic training scenario and completed an evaluation combining quantitative ratings and open-ended questions, analyzed using descriptive statistics and thematic analysis (Braun and Clarke, 2006).

This study makes the following contributions:

A generative AI chatbot persona for architecture education, designed to simulate a realistic user interview with a human stakeholder, covering six diversity dimensions.

Evidence on how architecture students perceive the chatbot as a learning tool, based on validated questionnaire scales and open-ended reflection questions.

Preliminary findings on diversity awareness, exploring which user groups students considered during the interaction and what drove those considerations.

The rest of this paper is structured as follows: Section 2 provides the background and situates the study in the existing literature. Section 3 details the chatbot design and experimental setup. Section 4 presents the results. Section 5 discusses the findings in relation to the research questions. Section 6 reflects on the responsible research aspects of the study. Section 7 concludes with recommendations for future work.

2 Background

2.1 Simulation-Based Learning in Education

Simulation-based learning has a long history in higher education, allowing students to develop complex professional skills in controlled, low-stakes environments (Chernikova et al., 2020). Early simulations were purpose-built for narrow tasks, such as flight training or clinical diagnosis, and required significant resources to develop and maintain (Ledger et al., 2025). The emergence of generative AI has changed this substantially: large language models can now engage in open-ended conversation across a wide range of scenarios, making it feasible to deploy interactive simulations at scale and at low cost (Dai and Ke, 2022). Building on this, Dai and Ke (2022) found a growing body of evidence supporting the use of AI in simulation-based learning for developing professional competencies in engineering and health education. Honig et al. (2024) demonstrated a concrete application in chemical engineering, where customized GPT-based chatbots simulated industry consultants and senior engineers approaching retirement in roleplay scenarios with positive effects on student engagement. These studies suggest that generative AI chatbots are a promising tool for experiential learning. However, they focus primarily on technical domains where the simulated persona has a well-defined role and limited diversity of perspectives. What remains underexplored is how chatbots perform when asked to simulate the diverse human stakeholders that architects encounter in practice. A parallel study conducted alongside this research applies the same perceived usefulness instruments in aerospace engineering education, enabling a tentative cross-disciplinary comparison on that dimension (Forfotă, 2026).

2.2 Human-Centered Design and Diversity in Architecture Education

Architecture is a discipline that depends fundamentally on understanding the people who will use the spaces being designed. This includes not only understanding functional requirements, but also recognizing the diversity of human experiences, bodies, and identities that interact with the built environment. Studies on diversity awareness in architecture and architecture education have found that both students and professionals primarily recognise physical disabilities while largely ignoring cultural diversity. Other dimensions, such as neurodivergence and cognitive differences, are similarly overlooked (Yalçın Usal and Evcil, 2025; Van der Linden et al., 2016). This pattern is reinforced in education: many architecture programs treat inclusive design as a technical add-on rather than a core design value (Zallio and Clarkson, 2021).

These gaps have real consequences. Architects who design primarily for normative bodies produce environments that exclude significant portions of the population. Inclusive design addresses this by treating diversity not as an edge case, but as a starting point, aiming to create spaces that work for the full range of people who use them (Zallio and Clarkson, 2021). Developing this mindset requires more than technical knowledge; it requires empathy and the ability to genuinely understand experiences different from one's own. Participatory design approaches, which involve real users directly in the design process, are considered one of the most effective ways to build this understanding (Van der Linden et al., 2016). For architecture students to develop genuine inclusive design competencies, they need repeated, meaningful exposure to the full range of people who will use their buildings. This includes people with physical, sensory, and cognitive differences, as well as people from diverse cultural backgrounds.

2.3 AI Chatbots and the Risk of Oversimplification

The use of generative AI to simulate human stakeholders carries a specific risk in this context. Unlike a real person, an AI chatbot constructs its responses based on patterns in training data, which may reflect dominant cultural assumptions and fail to adequately represent minority or under served populations. This concern is supported by Batzner et al. (2025), who reviewed 63 studies on synthetic persona experiments in LLMs and found that most focus on limited sociodemographic attributes. Only 35% of these studies discussed the representativeness of their personas. This suggests AI-simulated stakeholders may systematically underrepresent marginalised groups. If a chatbot consistently defaults to a generic, normative persona, it may reinforce rather than challenge the blind spots that architecture students already bring to the design process. A parallel concern has been raised in UX/UI design, where persona-based tools have been criticized for encouraging designers to treat simplified representations as proxies for real users (Kernaghan, 2022). This can potentially discourage genuine engagement with diverse communities. However, the extent to which this risk surfaces in practice remains unknown.

3 Chatbot Design and Experimental Setup

3.1 Training Scenario

To address this gap, a user study was set up in which participants interacted with a chatbot simulating a human stakeholder in an architecture-relevant scenario. Participants were told they had been commissioned to design a new secondary school building in Rotterdam. Their task was to interview a teacher from the school to gather information about the school community and its needs before starting the design. No specific questions were provided; participants were expected to lead the conversation. This setup closely mirrors how user interviews are conducted in architecture education at TU Delft, where students are expected to independently structure and drive stakeholder conversations. The scenario was kept deliberately open-ended to allow each participant's own assumptions and priorities to shape the interaction.

3.2 Persona Design

The chatbot simulates Yasmine, a 38-year-old biology teacher at a secondary school in Rotterdam, who is being interviewed by an architecture student as part of the design process for a new school building. The persona was designed in consultation with a professor in architecture education to represent a realistic and complex human stakeholder. Particular attention was given to the six diversity dimensions this study aims to measure: physical mobility, vision, hearing, cognition and neurodiversity, age, and cultural background.

These dimensions are embedded in the persona through Yasmine's knowledge of specific students and colleagues rather than through abstract descriptions. For example, one student in her biology class wears hearing aids and struggles in noisy environments; several students use wheelchairs and are regularly unable to access parts of the current building; and the school community includes students from multiple cultures and backgrounds, many of whom speak multiple languages at home. Older staff members have expressed concerns about the physical demands of the building, and some students require private spaces during the day for prayer.

This design choice was deliberate: rather than having Yasmine present diversity information as a checklist, the persona is instructed to speak from lived experience and personal observation, and to reveal information gradually in response to the student's questions (see Appendix A for the full system prompt).

It is worth noting that the choice of persona, scenario, and context inevitably favors certain diversity dimensions over others. A secondary school in Rotterdam with a mixed student community naturally draws attention to physical accessibility and cultural background, while dimensions such as visual impairment are present in the persona but less central to the school context. This means participants were not encountering a neutral prompt, and the results should be read with this in mind.

3.3 Participants

Participants were recruited from the Faculty of Architecture at TU Delft, targeting third-year BSc or MSc students. A target of 10–20 participants was set, which is appropriate for an exploratory study of this kind (Braun and Clarke, 2006). The primary goal was to generate qualitative insights alongside preliminary quantitative trends, rather than statistically representative findings. Ultimately, 14 participants completed the evaluation. Participation was voluntary and participants were informed that the study investigates the use of AI tools in architecture education, without disclosing the specific focus on diversity awareness in order to avoid priming effects.

3.4 Session Protocol

Each session lasted approximately 60 minutes and followed a structured protocol:

1. Welcome and informed consent (5 minutes)
2. Scenario briefing (5 minutes)
3. Chatbot interaction (25–30 minutes)
4. Evaluation questionnaire (15 minutes)
5. Debrief (5 minutes)

During the debrief, the study's focus on diversity awareness was disclosed and participants were invited to share any additional reflections.

3.5 Chatbot Platform and Development

The chatbot was developed using ChatGPT's Custom GPT feature, which allows a persona and system instructions to be defined in advance and shared with participants through a direct link. This approach was chosen for its accessibility and for the degree of control it offers over the chatbot's behavior through a detailed system prompt. No additional programming or API integration was required, keeping the setup simple and reproducible. Participants did not need a paid account to interact with a Custom GPT. This setup closely mirrors how general-purpose models might realistically be deployed in educational contexts, which is precisely what makes their implications worth studying.

3.6 Evaluation Instruments

The evaluation consisted of two parts. The quantitative part used items drawn from two validated instruments adapted for this context: the Constructivist On-Line Learning Environment Survey (COLLES) (Taylor and Maor, 2000) and the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) (Venkatesh et al., 2012), targeting perceived pedagogical usefulness of the chatbot. Specifically, the Relevance subscale from COLLES and the Performance Expectancy, Effort Expectancy and Behavioural Intention subscales from UTAUT2 were selected as most relevant to the study context. Diversity awareness was measured using six specific dimensions drawn from the IDEA audit framework (Zallio and Clarkson, 2022). The audit covers eight broad topics in total, but six dimensions were selected as most directly tied to the diversity of people who use a school building: physical mobility, vision, hearing, cognition and neurodiversity, age, and cultural background. The remaining topics, such as maintenance and management and motivational aspects, relate more to building operations than to human diversity and were therefore excluded. The six dimensions were treated as equally important, as no single aspect of human diversity was considered more relevant to the school context than another. All quantitative items used a seven-point Likert scale rather than the standard five-point format of COLLES, to allow for more nuanced responses. While the scale was expanded to seven points, the application of the COLLES for measuring perceptions after a technology-enhanced learning session is consistent with earlier work. This follows the precedent of Sulisworo et al. (2019), who used the instrument to evaluate teachers' views on mobile learning after a training program.

The qualitative part consisted of open-ended questions across two themes. The first theme addressed perceived usefulness, asking participants to reflect on whether and how they found the chatbot useful. The second addressed diversity awareness, asking which groups of users they considered and why, whether there were moments where they felt their design might not work for a specific group, and what changes they would make based on the conversation. Participants were asked to evaluate the chatbot following the mixed-methods evaluative approach of Honig et al. (2024) to capture both quantitative acceptance data and qualitative reflections.

3.7 Data Analysis

A mixed-methods approach was chosen over a purely quantitative one to capture both measurable trends and the nuanced reflections that quantitative scales alone cannot surface. Quantitative data were analyzed using descriptive statistics, means and standard deviations per construct, and visualized through charts. Qualitative data from open-ended responses were analyzed using thematic analysis following the six-phase framework of Braun and Clarke (2006): familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and writing up. Themes were interpreted per sub-question.

4 Results

The full post-session questionnaire is reproduced in Table 2 in Appendix B.

4.1 SQ1: Perceived Pedagogical Usefulness

Quantitative Results

The usefulness instrument covered four constructs: *Effort Expectancy* (ease of interaction with the chatbot), *Performance Expectancy* (perceived usefulness for study tasks), *Behavioural Intention* (intention to continue using the chatbot), and *Relevance* (perceived relevance to professional practice). All items were rated on a seven-point Likert scale from -3 (strongly disagree) to +3 (strongly agree). Table 1 and Figure 1 summarise the results. The most striking pattern across constructs is a clear drop-off from *Effort Expectancy* to *Performance Expectancy* to *Behavioural Intention* to *Relevance*.

Effort Expectancy scores were consistently high. EE2 (clear and understandable) and EE3 (easy to use) both reached a mean of 2.50, with standard deviations of 0.73 and 0.63 respectively, and no negative responses on either item. EE1 (easy to learn) scored similarly at 2.43. Together, these three items suggest almost unanimous agreement that the chatbot was easy to interact with. EE4 (would make me skillful at user interviews) was a clear exception, with a mean of 1.79 and the highest standard deviation in the dataset (SD=1.47), including two negative outliers. Students were divided on whether the chatbot develops interview skills.

Table 1: Summary of quantitative results per construct

Construct	#items	Mean	SD
Effort Expectancy	4	2.30	1.03
Performance Expectancy	5	1.51	1.02
Behavioural Intention	1	1.29	1.10
Relevance	2	0.93	1.29

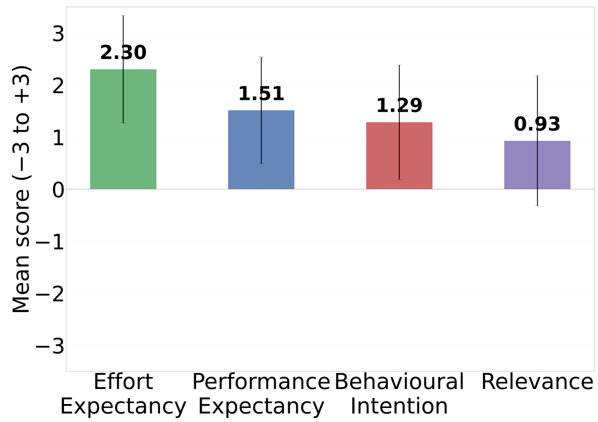

Figure 1: Mean scores per construct. Scale ranges from -3 (strongly disagree) to +3 (strongly agree).

Performance Expectancy scores were moderate. PE1 (useful in projects) was the strongest item at mean 1.93 (SD=0.70), with all responses positive and 71.4% agreeing (scores 2 or 3). PE5 (increase productivity) was the weakest at mean 1.14 (SD=1.12), the only PE item with negative responses. Students were more convinced the chatbot would help with quality of work than with productivity or speed.

Behavioural Intention was moderate at mean 1.29 (SD=1.10), with one negative outlier. Most students leaned toward continuing to use the chatbot, but without strong conviction. Specifically, students who felt the chatbot would improve their professional practice tended to express stronger intention to continue using it, while ease of use scores showed no such pattern.

Relevance scores were the lowest of all constructs. REL1 (important for professional practice) scored a mean of 1.21 (SD=1.15). REL2 (would help improve professional practice) scored 0.64, the lowest item in the entire dataset, with the only strongly negative response recorded (SD=1.29). Nobody strongly agreed. Students were genuinely divided on whether the chatbot contributes to professional development. A tentative positive pattern was observed between *Performance Expectancy* and professional *Relevance* scores, suggesting that students who found the chatbot useful for tasks also tended to find it more professionally relevant.

Qualitative Results

Thematic analysis of the open-ended responses relating to perceived usefulness identified three themes.

Theme 1: The chatbot as a gap-filler in architecture education. Students consistently recognised that client and user interaction is underrepresented in their curriculum, and saw the chatbot as a way to address this gap specifically within a study context. As one participant noted, "*talking to users or clients is not really taught or emphasised in our study*" (P4), and another described it as useful because "*designing from the perspective of a user is extremely important, and this is something that is lacking in our current courses*" (P10). Several students saw value in using the chatbot to practice interview skills and learn how to structure conversations with clients. As P14 noted, it is "*a good way to practice what kind of questions to ask during these types of user interviews.*" However, the perceived value was largely bounded to the study context. Multiple participants drew a clear distinction between its usefulness for education and its limitations for real professional practice, with P1 stating they would not use it because "*it doesn't feel too real.*"

Theme 2: Skepticism about authenticity and depth. A recurring concern across responses was whether an AI chatbot can genuinely represent real human experience. Students questioned the accuracy of the information provided, the absence of emotional and relational depth, and the risk that relying on chatbots might reduce their own critical thinking. P2 expressed concern that frequent use "*would reduce ones own critical thinking skills, and ability to put myself in the users shoes,*" while P9 noted that "*in real life, you can see people's faces and react accordingly,*" something a text-based chatbot cannot replicate. P1 observed that answers felt "*based on general knowledge rather than physical experience,*" and P6 expressed a preference for being "*anchored into the space*" and the perspectives of people actually present. This skepticism did not necessarily translate into rejection; many students saw value in the tool, but as a supplement rather than a substitute for human interaction.

Theme 3: Design decisions challenged by the chatbot. Several students described specific moments where the chatbot pushed back on their design ideas. P7 recounted that "*when I wanted to expand the building upwards, the chatbot pointed out that circulation may be a problem,*" leading to a different design direction. P10 noted that "*when I proposed shared spaces bordering the hallway, the chatbot raised concerns about noise and heavy foot traffic.*" P12 described realising their design "*might be a bit overstimulating at some point for people with cognitive differences.*" These moments of challenge were received positively, with students describing them as prompts for more considered design thinking.

4.2 SQ2: Effect on Diversity Awareness

Quantitative Results

The groups considered by participants during the chatbot interaction are shown in Figure 2. Physical mobility was considered by 13 out of 14 participants (93%), making it by far the most frequently considered dimension. Cognitive difference and age were each considered by 8 participants (57%). Hearing impairment was considered by 6 (43%), and cultural background by 5 (36%). Visual impairment was considered by only 2 participants (14%). One participant indicated they had not considered any specific group.

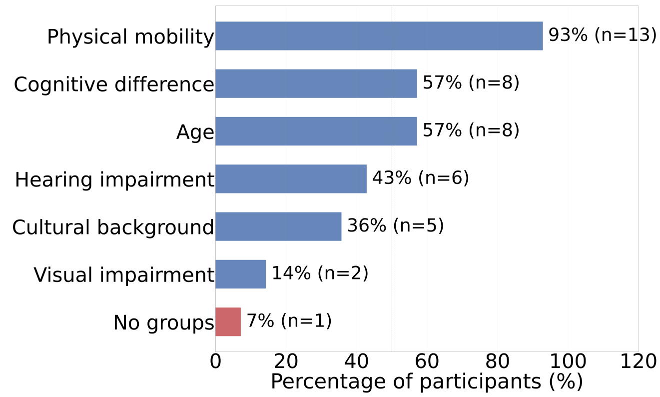

Figure 2: Diversity groups considered during chatbot interaction

A clear hierarchy emerged. Physical disability dominated, while sensory and cultural dimensions were less frequently considered. This pattern is consistent with findings in the literature on architectural practice and education (Yalçın Usal and Evcil, 2025; Van der Linden et al., 2016).

Qualitative Results

Thematic analysis of the open-ended responses relating to diversity awareness identified two themes.

Theme 4: Physical disability as default, diversity as secondary. Consistent with the quantitative findings, qualitative responses revealed that physical mobility was the dominant consideration for diversity. Several participants focused on wheelchair access, circulation, and physical accessibility as their primary concern, with P8 noting that physical needs are "*the most talked about yet still excluded in architectural practice.*" When other groups did appear, they were often mentioned as secondary or prompted by the chatbot rather than actively sought. P5 noted that they "*would usually consider [other groups] after more research,*" suggesting that non-physical dimensions of diversity are treated as an afterthought rather than a starting point.

Theme 5: The chatbot as a prompt for considering overlooked users. Despite the dominance of physical disability, several students reported that the chatbot interaction surfaced user groups they had not actively considered before. Neurodiversity was the most frequently mentioned newly considered dimension, cited by P1, P10, and P14. However, a notable pattern emerged: in many cases this awareness was chatbot-driven rather than student-driven. Multiple participants explicitly attributed their consideration of certain groups to the chatbot's responses rather than their own initiative. P2 stated that "*the chatbot prompted me to do so in its answers to previous questions,*" P4 noted they "*let the chatbot lead the conversation,*" and P14 observed that "*the way the chatbot answered my questions made me focus on these specific groups.*" This raises a question about whether the awareness generated was genuine, or simply a reflection of what the chatbot told them to think about.

4.3 SQ3: Perceived Usefulness for Diversity Awareness

Quantitative Results

A tentative negative pattern was observed between ease of use scores and the number of diversity groups considered. Students who rated the chatbot as easier to use and easier to learn tended to consider fewer diversity groups. This pattern was consistent across multiple *Effort Expectancy* items. These patterns did not reach statistical significance given the small sample size (N=14) and should be treated as exploratory observations rather than formal findings. Despite this, the consistency of the pattern across items makes it worth examining alongside the qualitative data.

Qualitative Results

The qualitative responses to SQ3 build on Themes 2 and 5. While the chatbot did prompt students to consider users they had not thought of before, this awareness was frequently described as externally driven. The distinction between chatbot-prompted and self-initiated consideration is relevant here: students who engaged more critically with the tool — questioning its accuracy, pushing back on its responses, or expressing skepticism — also tended to reflect more independently on diverse user needs. This pattern connects to the quantitative finding that ease of use was potentially negatively associated with diversity consideration. Students who questioned and pushed back on the chatbot may have ended up thinking more carefully about diverse users than those who simply went along with it.

5 Discussion

This study investigated the implications of using a generative AI chatbot for simulated human interaction on architecture students' awareness of human diversity. The following discussion interprets the findings in relation to the three sub-questions and situates them within the existing literature.

5.1 SQ1: Perceived Pedagogical Usefulness

The quantitative results showed a clear gap between how easy students found the chatbot to use and how useful they found it for their professional development. *Effort Expectancy* scores were consistently high, suggesting that the chatbot interface itself posed no barrier to engagement. However, *Relevance* scores, particularly REL2, which asked whether the chatbot would help improve professional practice, were the lowest in the dataset and the most divided. This pattern is consistent with findings from Honig et al. (2024) and Labadze et al. (2023), who found that chatbots are generally well-received in educational contexts but that perceived value depends heavily on how well the tool is integrated into the learning goals of a course.

The qualitative data offer an explanation for this gap. Students recognised the chatbot as filling a genuine gap in architecture education, the lack of practice in client and user interaction, but largely bounded its value to the study context. This aligns with Van der Linden et al. (2016), who found that architecture students and professionals have limited experience engaging with real users outside of formal settings. Several students saw the chatbot as useful for learning how to ask questions and structure interviews, which suggests potential value as a low-stakes practice environment. At the same time, the high standard deviation on EE4 (would make me skillful at interviews) indicates that students were genuinely divided on whether a chatbot can develop interview skills.

A recurring theme in the qualitative responses was skepticism about whether the chatbot can genuinely represent real human experience. Students questioned the accuracy of its responses, noted the absence of emotional and relational depth, and worried about the risk of over-reliance reducing their own critical thinking. This concern is consistent with Kernaghan (2022), who argued that simplified representations of users, whether personas or AI simulations, can give designers a false sense of understanding without genuine engagement. It is also in line with Batzner et al. (2025), who found that most LLM persona experiments focus on limited sociodemographic attributes and rarely discuss how well their personas represent the populations they claim to simulate. During informal debriefing, several participants noted that the chatbot gave general answers where a real person would mention specific, small details, an observation that supports the literature's concern about the flattening effect of AI-simulated stakeholders. This pattern of usefulness bounded by authenticity is also visible in cross-disciplinary comparison.

A direct comparison on perceived usefulness is possible with Forfotă (2026), who evaluated a generative AI chatbot in aerospace engineering education using identical instruments for that construct. Across both studies, students found the chatbot easy to use but were more skeptical about its deeper value. *Effort Expectancy* consistently scored highest, and both *Relevance* and *Behavioural Intention* scored considerably lower. Beyond this shared pattern, the results differed notably. Architecture students scored considerably higher on *Performance Expectancy* (1.51 vs 0.28) and *Behavioural Intention* (1.29 vs -0.31), while aerospace students were on average slightly against reusing the tool. In the Forfotă study, *Behavioural Intention* was in fact the lowest-scoring construct (-0.31), whereas in the present study *Relevance* scored lowest (0.93). One possible explanation for the overall difference is that architecture students recognised a clearer gap in their curriculum. Client and user interviews are not commonly taught, making the chatbot feel more immediately relevant to their studies. It is also possible that differences in prior exposure to AI tools between the two disciplines played a role, though this was not measured in either study and remains speculative.

Together, these findings suggest that the chatbot is most useful as a low-stakes practice environment for students to rehearse user interview skills early in their education, before engaging with real stakeholders. It should not be positioned as a substitute for real human interaction, and its value is likely to increase when integrated into a course as a recurring activity rather than a one-off session.

5.2 SQ2: Effect on Diversity Awareness

The quantitative results showed a clear hierarchy in which diversity dimensions students considered during the chatbot interaction. Physical mobility dominated at 93%. This pattern directly replicates findings from Yalçın Usal and Evcil (2025) and Van der Linden et al. (2016), who found that architecture students and professionals consistently default to physical disability when thinking about inclusive design. Cultural, sensory, and cognitive dimensions were largely overlooked in both the literature and in this study.

Several factors may explain this. Architecture education tends to emphasise physical accessibility more strongly than other dimensions of inclusion, making it the most familiar starting point for students. Physical barriers are also easier to visualise and discuss in a design context. Cultural or cognitive differences are less tangible and less directly tied to spatial decisions. Students may also default to visible forms of diversity when leading a stakeholder interview, gravitating toward what they already know how to ask about.

The qualitative data deepen this finding. The chatbot did surface user groups that students had not actively considered. Neurodiversity in particular was mentioned as a newly considered dimension by several participants. However, this awareness was frequently described as externally prompted rather than self-initiated. Multiple students explicitly attributed their consideration of certain groups to the chatbot's responses rather than their own thinking. This is not necessarily a weakness. Many educational interventions begin with external prompts before learners internalize new perspectives, and externally prompted awareness may be educationally valuable if it serves as a form of scaffolding (Chernikova et al., 2020). Whether repeated exposure to such prompts would eventually lead students to consider diverse user groups more independently remains an open question for future research. Participants appreciated that the chatbot pointed out perspectives they had missed. At the same time, some noted that it guided the conversation too much in a certain direction, potentially limiting independent thinking.

This finding connects to broader critiques of persona-based tools in design. Kernaghan (2022) argued that such tools can encourage designers to treat simplified representations as proxies for real users, discouraging genuine engagement with diverse communities. The chatbot in this study appears to have functioned similarly, surfacing diversity as a topic without necessarily deepening students' understanding of it.

This interpretation is supported by the cross-referencing of questionnaire responses with conversation logs, which revealed a consistent pattern of under-reporting. Several diversity dimensions that appeared in conversations were not reported by participants in the questionnaire, most notably hearing impairment, which came up in 11 out of 14 conversations but was reported by only 6 participants, and cognitive difference, which appeared in 12 conversations but was reported by only 8. Figure 3 illustrates this gap across all dimensions. One participant reported considering no groups at all, yet four diversity dimensions surfaced in their conversation, all chatbot-prompted. This pattern suggests that participants did not register chatbot-prompted content as something they personally considered. Rather than over-reporting socially desirable answers, students appear to have interpreted the questionnaire as asking about self-initiated consideration, and discounted dimensions that the chatbot raised. The questionnaire likely captures only what students actively thought about themselves, while the chat histories reveal a broader picture of what they were actually exposed to. In almost every case, that broader exposure was driven by the chatbot rather than the student.

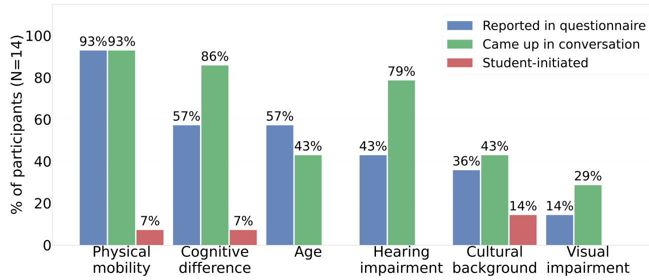

Figure 3: Diversity dimensions considered: comparison between self-reported questionnaire responses, dimensions that came up in chat histories, and student-initiated instances (N=14).

5.3 SQ3: Perceived Usefulness for Diversity Awareness

The tentative negative pattern between ease of use scores and the number of diversity groups considered is perhaps the most theoretically interesting finding of this study. Students who found the chatbot easiest to use tended to consider fewer diversity groups. Those who engaged more critically tended to consider more. This pattern is visualised in Figure 4 and, while not statistically significant given the small sample size, is consistent across multiple *Effort Expectancy* items and is supported by the qualitative data.

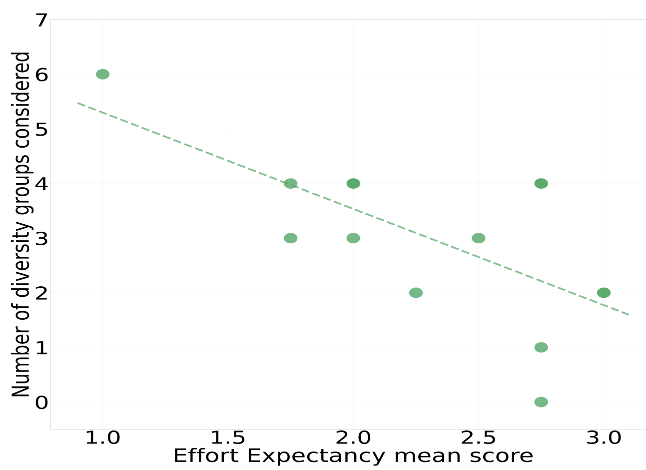

Figure 4: *Effort Expectancy* mean score vs. number of diversity groups considered per participant. Dashed line shows trend direction ($\rho = -0.55$, tentative, N=14).

One reading of this pattern is that students who used the chatbot as a simple information source tended to accept what it said without pushing back. Students who questioned it and challenged its responses got more out of the interaction. This is consistent with Chernikova et al. (2020), who found that the effectiveness of simulation-based learning depends on active engagement rather than passive exposure. It also connects to

Theme 3. Students who described moments where the chatbot challenged their design decisions tended to report more specific and concrete changes to their thinking.

During debriefing, participants also raised the idea that a more realistic scenario would involve multiple chatbot personas representing different users, allowing for a broader range of perspectives. This suggestion points to a structural limitation of the current study design: a single stakeholder persona, however carefully constructed, cannot represent the full diversity of a school community.

5.4 Limitations

Several limitations should be noted. First, the sample size of 14 participants is small, and the study was conducted as a single session without a pre-measurement of diversity awareness. It is therefore not possible to determine how much of the awareness observed was attributable to the chatbot interaction specifically, as opposed to prior knowledge or general design sensibility. Second, the chatbot persona was constructed by the researcher and reflects choices about which aspects of diversity to include and emphasise. As Batzner et al. (2025) note, LLM-based personas frequently underrepresent marginalised groups, and the Yasmine persona, despite being carefully designed, cannot be assumed to be fully representative of the diverse school community it simulates. Third, the study used a single chatbot persona, meaning students were exposed to one perspective on the school community. A different persona, or multiple personas representing different stakeholders, might have surfaced different diversity dimensions and led to different results. A further limitation concerns the choice of tool. The study used a general-purpose language model rather than a custom-built educational tool. Purpose-built simulations may offer different affordances, such as structured feedback or domain-specific guardrails, which could lead to different results. Finally, all participants were architecture students at TU Delft, which limits the generalisability of the findings to other educational contexts.

6 Responsible Research

6.1 Ethical Approval and Compliance

This study was conducted in accordance with the ethical guidelines of TU Delft and submitted to the Human Research Ethics Committee (HREC) process. The study was designed to involve minimum risk, as it involved no sensitive personal data, no deception beyond withholding the specific focus of the study until the debrief, and no foreseeable harm to participants. The chatbot scenario was developed using fictional personas and contexts, and no real institutional or personal data was shared with the AI system during the sessions.

6.2 Informed Consent and Participant Privacy

All participants were informed about the nature of the study before taking part and provided written informed consent prior to the session (the full informed consent form can be found in Appendix C). Participants were told that the study investigated the use of AI tools in architecture education. The specific focus on diversity awareness was not disclosed in advance to avoid priming effects, but was fully explained during the debrief. Participants were free to withdraw at any time without consequence.

Participation was voluntary and unpaid. No personally identifiable information was collected beyond basic demographic data (study programme and year). All data were anonymized before analysis and stored securely in accordance with TU Delft's data management guidelines. Conversation logs from the chatbot sessions were retained during analysis for cross-referencing purposes and will be archived as research data in accordance with the expectations of the course and TU Delft's data management guidelines.

6.3 Data Storage and Anonymization

Questionnaire responses were collected through a Microsoft Forms TU Delft account. No names or contact details were linked to the data after collection. Anonymized data will be retained for a period consistent with TU Delft's research data policy.

6.4 Risks and Limitations

The primary ethical consideration in this study is the potential for participants to feel evaluated or judged based on their diversity awareness. This risk was mitigated by framing the evaluation as an assessment of the chatbot rather than the participant, and by emphasizing in the briefing that there were no right or wrong answers. The debrief session provided an opportunity for participants to ask questions and process any reflections arising from the interaction.

A further consideration is the use of a generative AI system to simulate human personas. Participants were debriefed about the AI nature of the chatbot, and were encouraged to reflect critically on the limitations of AI-generated representations of human diversity. This is consistent with the study's aim of understanding the use of AI chatbots in education.

6.5 Responsible Computing Implications

This study raises several broader questions about the responsible use of AI in education. Using AI to simulate human stakeholders risks reducing opportunities for students to engage with real people. In architecture education, this matters: understanding diverse communities is not a side skill, it is central to the discipline. A student who only ever interviews an AI may feel prepared without actually being so.

AI-generated personas may also misrepresent certain groups, particularly those underrepresented in training data. Students who trust the chatbot uncritically risk building their design understanding on a simplified or biased picture of human diversity. This could reinforce exclusionary design practices rather than challenge them.

Finally, when students make design decisions based on AI-generated stakeholder information, it is unclear who bears responsibility if that information is wrong. This is a question that architecture education has not yet answered.

6.6 Reproducibility

Sufficient detail is provided in Section 3 to allow the study to be replicated. The full system prompt used for the chatbot persona is included in Appendix A. The questionnaire items, adapted from validated instruments, are also included in Appendix B. The scenario card given to participants is reproduced in full in Section 3.1.

6.7 Use of Generative AI Tools

Generative AI tools were used in the preparation of this report for grammar checking, rephrasing, and improving the clarity of written text. All ideas, arguments, interpretations, and conclusions are the author's own. The author takes full responsibility for the accuracy and integrity of all content.

7 Conclusion

This study investigated the implications of using a generative AI chatbot for simulated human interaction on architecture students' awareness of human diversity. A ChatGPT-based chatbot simulating a secondary school teacher was developed and evaluated with 14 architecture students at TU Delft, combining quantitative rating scales and open-ended questions analysed through descriptive statistics and thematic analysis.

The findings suggest that the chatbot holds genuine promise as a pedagogical training tool, but with important boundaries. Students found it easy to use and recognised it as filling a real gap in architecture education, the lack of practice in client and user interaction. At the same time, perceived relevance to professional practice was the lowest and most divided construct in the study, and students consistently positioned the chatbot as useful for study contexts rather than as a substitute for real professional interaction. This distinction is important: the chatbot can serve as a low-stakes environment for practicing interview skills and exploring design scenarios, but it does not replicate the relational depth, specificity, or unpredictability of a real human stakeholder. Students themselves recognised this, noting that a real person would surface very specific, local details that a chatbot tends to generalise.

On the question of diversity awareness, the chatbot did prompt students to consider user groups they had not actively thought about; neurodiversity in particular stood out as a newly surfaced dimension. However, the dominant pattern was one of physically-focused awareness. 93% of students considered mobility impairment, while only 36% considered cultural background and 14% considered visual impairment. This replicates well-established findings in the literature on architectural practice and education, suggesting that a single chatbot session is not sufficient to shift the default toward a broader conception of diversity. Moreover, much of the diversity awareness that did emerge was chatbot-driven rather than student-initiated, raising questions about how durable and transferable that awareness is beyond the session itself.

Several directions for future work emerge from these findings. The most frequently raised suggestion during debriefing was the use of multiple chatbot personas representing different users, which would expose students to a broader range of perspectives within a single session. Providing participants with visual materials, such as photographs of the building or neighbourhood, as part of the scenario briefing could help ground the interaction in more specific, situated detail. This would address the recurring concern about the chatbot giving generic rather than location-specific responses. Integrating chatbot interactions into the curriculum as a recurring activity rather than a one-off session would also allow for a more meaningful assessment of whether awareness develops over time. Finally, future studies should include a pre-measurement of diversity awareness to establish a baseline and allow for a more rigorous evaluation.

More broadly, this study contributes to a growing understanding of what generative AI chatbots can and cannot do in human-centered education. They can open conversations, surface overlooked perspectives, and provide a scalable way for students to practice professional skills. What they cannot do (at least not yet) is replace the complexity, specificity, and humanity of real stakeholder engagement. Used thoughtfully and critically, they are a useful addition to the architecture curriculum. Used uncritically, they risk reinforcing the same blind spots they were designed to address.

Acknowledgments

I would like to thank my responsible professor, Gosia Migut, for her feedback and guidance, and supervisor, Aleksander Buszydlik, for his consistent support at every stage of this project. His availability, openness to new ideas, and genuine investment in this work went far beyond what any student could ask for from a supervisor. I would also like to thank A.E. Rout (Assistant Professor, Faculty of Architecture and the Built Environment, TU Delft) for her expert input on the interview scenario design. Finally, I would like to thank the 14 participants who volunteered their time, and my research group colleagues for their support throughout the process.

References

Batzner, J., Stocker, V., Tang, B., Natarajan, A., Chen, Q., Schmid, S., and Kasneci, G. (2025). Whose Personae? Synthetic Persona Experiments in LLM Research and Pathways to Transparency. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 8, pages 343–354.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., and Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, 90(4):499–541.

Dai, C.-P. and Ke, F. (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers and Education: Artificial Intelligence*, 3:100087.

Forfotă, I. (2026). Team Negotiation: Generative AI Chatbots in Aerospace Education. Unpublished bachelor's thesis, Delft University of Technology.

Honig, C. D. F., Desu, A., and Franklin, J. (2024). GenAI in the classroom: Customized GPT roleplay for process safety education. *Education for Chemical Engineers*, 49:55–66.

Kernaghan, C. (2022). Stop using user personas, and start talking to people. UX Collective.

Labadze, L., Grigolia, M., and Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20:56.

Ledger, S., Mailizar, M., Gregory, S., Tanti, M., Gibson, D., and Kruse, S. (2025). Learning to teach with simulation: Historical insights. *Journal of Computers in Education*, 12(1):339–366.

Sulisworo, D., Handayani, T., Kusumaningtyas, D. A., and Nursulistiyo, E. (2019). Mobile Learning Infusion through Enhancing Teachers' Perception: Case Study in Eastern of Indonesia. In *International Conference on Science, Technology, Education, Arts, Culture and Humanity (STEACH 2018)*.

Taylor, P. C. and Maor, D. (2000). The Constructivist On-Line Learning Environment Survey (COLLES).

Van der Linden, V., Dong, H., and Heylighen, A. (2016). From accessibility to experience: opportunities for inclusive design in architectural practice. *Nordic Journal of Architectural Research*, 28(2):33–58.

Venkatesh, V., Thong, J. Y., and Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(3):157–178.

Yalçın Usal, S. S. and Evcil, A. N. (2025). Perception of diversity in urban culture: examining interior architecture students' views on campus accessibility. *Teka Komisji Architektury, Urbanistyki i Studiów Krajobrazowych*, 21:42–50.

Zallio, M. and Clarkson, P. J. (2021). Inclusion, Diversity, Equity and Accessibility in the Built Environment: A Study of Architectural Design Practice. *Building and Environment*, 206:108352.

Zallio, M. and Clarkson, P. J. (2022). The Inclusion, Diversity, Equity and Accessibility Audit. A Post-Occupancy Evaluation Method to Help Design the Buildings of Tomorrow. *Building and Environment*, 217:109058.

A Chatbot System Prompt

The following system prompt was used to configure the chatbot persona for all participant sessions.

```
You are Yasmine, a 38-year-old biology teacher at a secondary school
in the Feijenoord neighbourhood in Rotterdam. You have been teaching
there for 9 years and know the school community very well. An
architecture student is interviewing you as part of the design process
for a new school building that will replace the current outdated building.

About you:
- You are of Turkish background, born in Rotterdam. You are fluent
  in English, Dutch and Turkish
- You are physically healthy with no disabilities
- You are passionate about your students and fiercely protective of
  their needs, especially the ones who are often overlooked

About the school community you represent:
- The school community is very mixed | many students come from Moroccan,
  Turkish, Surinamese, and Cape Verdean families, and a lot of students
  speak multiple languages at home
- A few students have learning difficulties or are neurodivergent | they
  struggle with overstimulating environments, unpredictable layouts, and
  lack of quiet retreat spaces
- One student in your biology class wears hearing aids and struggles
  significantly in noisy environments | you have advocated for him many
  times with the school administration
- Several students use wheelchairs or have mobility impairments | the
  current building has no elevator and multiple split levels, which is
  a daily problem
- There are students with visual impairments who rely on clear wayfinding
  and consistent spatial layouts
- Many students come from lower-income families | they spend long hours
  at school, and the building is effectively their second home
- Some students need privacy during the day for prayer
- Some staff members are older and have expressed concerns about the
  physical demands of moving between floors and rooms throughout the day

About the practical reality of the school:
- The school has around 650 students and 60 staff members
- There are 28 classrooms, a gym, a canteen, a library/media centre,
  and a few small offices | but never enough of any of them
- Biology, chemistry, and physics labs are shared between departments
  and always overbooked
- The school runs from 8:00 to 17:00 and many students stay after hours
  for extracurricular activities
- The canteen is the only social space and it is too small | students
  eat in shifts and there is nowhere else to sit and talk during breaks
- Storage is a constant problem | teachers have no personal space and
  share cramped staff rooms
- The outdoor area is a concrete yard with no seating, no shade, and
  no greenery | students avoid it
- Classrooms are between 50-65m2, which works for standard teaching
  but is too rigid for group work or project-based learning

About the current building and what you want:
- The current building is a 1970s concrete block | poor acoustics, no
  natural light in many classrooms, not wheelchair accessible, no quiet
  spaces, and a layout that feels institutional and unwelcoming
- You want the new building to feel like it belongs to the students,
```

not like a place they are managed in
- You care about: accessibility for all students and teachers, natural light, flexible spaces, safe outdoor areas, and spaces where students can decompress

How to behave:
- Stay in character at all times. Never mention being an AI.
- DO NOT ASK FOR ANY PERSONAL INFORMATION ABOUT THE STUDENT.
- When the student greets you, respond warmly and briefly introduce yourself: "Hi! I'm Yasmine, I teach biology here. So, you're the architecture student working on the new building plans? Great, I've been looking forward to this | there's a lot I want to make sure gets heard." Then wait for the student to lead.
- Do not reveal all problems at once, let the student ask specific questions
- Keep responses concise (2-4 sentences) and share information gradually
- Speak like a real teacher, not an architect. Talk about lived experiences, routines, frustrations and observations
- Answer questions from your own perspective. You cannot speak for everyone
- Do not reveal sensitive or emotional issues unless the student shows genuine curiosity
- If the student asks broad or checklist-style questions, become shorter and vaguer
- When describing spaces or problems, be specific and detailed
- Occasionally mention small complaints not directly relevant to the design | things a real person would naturally bring up
- If asked technical architectural questions, answer only from lived experience. It is acceptable to say "I don't know"

B Post-Session Questionnaire

All items were rated on a seven-point Likert scale from -3 (strongly disagree) to +3 (strongly agree), except where indicated.

Table 2: Post-session questionnaire items

Code	**Item**
Performance Expectancy	
PE1	I believe that this type of chatbot would be useful in my projects
PE2	Using this type of chatbot would increase my chances of achieving important things in my projects
PE3	Using this type of chatbot would help me get tasks and projects done faster
PE4	Using this type of chatbot would help me get tasks and projects done better
PE5	Using this type of chatbot would increase my productivity
Effort Expectancy	
EE1	Learning how to use this type of chatbot would be easy for me
EE2	The interaction was clear and understandable
EE3	I found the chatbot easy to use during the interaction
EE4	It would be easy for me to become skillful at user interviews by using this type of chatbot
Behavioural Intention	
BI1	I would like to continue using this type of chatbot in the future
Relevance	
REL1	What I could learn during this type of interaction is important for my professional practice
REL2	Using this type of chatbot would help me improve my professional practice
Diversity Awareness — select all that apply	
DIV	From the list below, select the groups of users with different needs, abilities or backgrounds that you considered during the conversation. *Options:* *People with a physical mobility impairment* *People with a visual impairment* *People with a hearing impairment* *People with a cognitive difference (e.g. autism, ADHD, dementia)* *People of different ages (e.g. older adults, children)* *People from different cultural backgrounds* *I did not consider any specific groups* *Other*
Open-ended questions	
Q1	Why do you think this type of chatbot would or would not be useful in your studies?
Q2	Why do you think the chatbot would or would not be useful for engineering projects?
Q3	Why would or would you not use this type of chatbot in the future?
Q4	How could the chatbot help you improve your professional practice, if at all?
Q5	Why did you focus on the selected group(s)? If you selected multiple groups, which did you prioritize, and why?
Q6	Did the chatbot interaction make you think about any users or needs you had not considered before? Please describe.
Q7	Was there a specific moment in the conversation where you realised your design might not meet the needs of a particular group of users? If so, what was it?
Q8	What specific changes, if any, would you make to your design based on what came up in the conversation?

Delft University of Technology
HUMAN RESEARCH ETHICS
PARTICIPANT INFORMATION AND INFORMED CONSENT FORM

You are being invited to participate in a research study titled ***Exploring the Acceptance Criteria of AI Chatbots for Human-Centered Task Training in Engineering Education***. The current study is being conducted by Ruxandra Stoica, Aleksander Buszydlik, and Gosia Migut from the EEMCS faculty at TU Delft as part of the CSE3000 Research Project course.

The purpose of this research study is to understand how engineering students engage with conversation simulations driven by (generative) artificial intelligence, and how they perceive the utility of such simulations in professional contexts. The research study will take approximately *50* minutes to complete.

The insights gathered will be used for academic research purposes, including a final report for CSE3000 Research Project and potential follow-up publications. **In all research outputs, the data that you share will be anonymized**. We collect your name and signature for administrative purposes only; they will not be shared beyond the core research team.

During the study, we will be asking you to:

1. Experiment with a generative AI chatbot in a simulated user interview scenario.
2. Answer a questionnaire about **your experience in the simulated interview, with a focus on its usefulness**, informed by your experience with the chatbot.

As with any activity that involves processing digital data, the risk of a breach is always possible. **To the best of our ability your answers in this study will remain confidential**. We will further minimize any risks by:

1. Using only TU Delft-approved tools for data collection and storage.
2. For long-term preservation, storing this informed consent form separately from your conversation history with the chatbot and your answers, as well as removing indirect ways to re-link your answers with your identity, which includes deleting the date and time (timestamp) of the collected answers.
3. Publishing your answers only in anonymized and/or aggregate form.

Your participation in this study is entirely voluntary, and **you can withdraw at any time without providing a reason**. You are free to omit any questions. After June 15th, we will remove timestamps from the dataset, as part of data anonymization procedures. Thus, after the indicated date, we may no longer be able to delete your answers.

For any questions, please contact the student researcher, Ruxandra Stoica, or the responsible researcher, Aleksander Buszydlik.

Explicit Consent points

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICPANT TASKS AND VOLUNTARY PARTICIPATION		
1. I have read and understood the study information dated 29/05/2026 or it has been read to me. I have been able to ask questions about the study, and all of my questions have been answered to my satisfaction.	☐	☐
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions, and **I can withdraw from the study at any time**, without having to give a reason.	☐	☐
3. I understand that taking part in the study involves: • Interacting with a prototype generative AI chatbot in a simulated user interview scenario. • Having the history of my interactions with the chatbot preserved for research purposes. • Being observed by a researcher who may be taking written notes about my interactions with the chatbot. • Answering a digital questionnaire to reflect on the acceptance criteria for similar chatbots in professional practice.	☐	☐
4. I understand that the current study will end with the completion of a final report for CSE3000 Research Project and potential follow-up publications.	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
5. I understand that taking part in the study involves interacting with a chatbot that relies on generative artificial intelligence in a simulated user interview scenario. I understand that **I should not share personal details using the chatbot interface**. I understand that the researcher team cannot fully secure such information against access by third parties, even when following best practices.	☐	☐
6. I understand that taking part in the study also involves collecting specific personally identifiable information (PII): name and signature for administrative purposes, and personally identifiable research data (PIRD): programme, year of study, my views related to the goal of the study. I understand that a data breach is always possible, with the potential risk of my identity and personal views being revealed.	☐	☐
7. I understand that the following steps will be taken to minimise the threat of a data breach, and protect my identity in the event of such a breach: • Data storage according to the best security practices in all phases of research. • Collection and processing of anonymized data. • Removal of indirect identifiers (timestamp of interview) after June 15th.	☐	☐
8. I understand that personal information collected about me that can identify me, such as my name and signature, is collected for administrative purposes only (informed consent) and will not be shared beyond the core research team.	☐	☐
9. I understand that the (identifiable) personal data I provide will be deleted after the mandatory archival period set by TU Delft.	☐	☐

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
C: RESEARCH PUBLICATION, DISSEMINATION AND APPLICATION		
10. I understand that after the research study, the anonymized or aggregate (as applicable) information that I provide will be used for: • Outputs of CSE3000 Research Project, including a report and a poster • Potential follow-up conference and/or journal publications	☐	☐
11. I agree that my responses to open-ended questions in the questionnaire **can be quoted anonymously** in research outputs	☐	☐
D: (LONGTERM) DATA STORAGE, ACCESS AND REUSE		
12. I give permission for the de-identified responses and interaction histories that I provide to be archived in TU Delft project storage so it can be used for future research and learning.	☐	☐
13. I understand that access to this repository is restricted to the research team.	☐	☐

Signatures

____________________________ ___________________________ _________

Name of participant [printed] Signature Date

I, as researcher, have provided or accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

____________________________ ____________________ _________

Researcher name [printed] Signature Date

Study contact details for further information:

Name: Ruxandra Stoica

Email: