

Highlights

How do software practitioners perceive human-centric defects?

Vedant Chauhan,Chetan Arora,Hourieh Khalajzadeh,John Grundy

- We describe the phenomenon of “human-centric defects” in software caused by end-user human differences.
- We describe results from a survey of 50 practitioners and follow-up detailed interviews of 10 practitioners on human-centric defect definition, occurrence, reporting and fixing issues in software engineering.
- We provide key recommendations for SE practitioners to improve human-centric defect reporting and fixing. We recommend organisations and teams utilise diverse users for formulating solutions and create diverse environments where human-centric training and customer-centric processes are employed. Additionally, defect reporting tools should include automation, HCD-specific templates and diverse user information to manage HCDs efficiently.
- For researchers, we suggest research is needed on how to improve the human-centric defect reporting and fixing process: the creation of a consistent HCD taxonomy, HCD-supporting tools, concrete checklists or templates for HCD reporting, and training mechanisms to improve the HCD management process.

How do software practitioners perceive human-centric defects?

Vedant Chauhan^a, Chetan Arora^a, Hourieh Khalajzadeh^b and John Grundy^a

^aFaculty of Information Technology, Monash University, Wellington Road, Clayton, 3800, Victoria, Australia

^bSchool of Information Technology, Deakin University, 221 Burwood Highway, Burwood, 3125, Victoria, Australia

ARTICLE INFO

Keywords:

Human-centric defects
Software Engineering
Information Technology
Defect reporting
Software Development Lifecycle

ABSTRACT

Context: Human-centric software design and development prioritises the way users prefer to complete their jobs, rather than expecting users to adapt to the software. Software users can have different genders, ages, cultures, languages, disabilities, socioeconomic statuses, and educational backgrounds, among many other differences. Due to the inherently varied nature of these differences and their impact on software usage, preferences and issues of users can vary, resulting in user-specific defects that we term as ‘human-centric defects’ (HCDs).

Objective: This research aims to understand the perception and current management practices of such HCDs by software practitioners, identify key challenges in reporting, understanding and fixing them, and provide recommendations to improve HCDs management in software engineering.

Method: We conducted a survey and interviews with software engineering practitioners to gauge their knowledge and experience on HCDs and the defect tracking process.

Results: We analysed fifty (50) survey- and ten (10) interview- responses from SE practitioners and identified that there are multiple gaps in the current management of HCDs in software engineering practice. There is a lack of awareness regarding human-centric aspects, causing them to be lost or under-appreciated during software development. Our results revealed that handling HCDs could be improved by following a better feedback process with end-users, a more descriptive taxonomy, and suitable automation.

Conclusion: HCDs, given their diverse end-user base, present a major challenge to software practitioners. In the software engineering domain, research on HCDs has been limited and requires effort from research and practice communities to create awareness and support for human-centric aspects.

1. Introduction

Software bugs, or *defects*, (Strate and Laplante, 2013; Yusop et al., 2016a) are a well-known phenomenon in software engineering (SE). Such defects can appear as a result of mistakes by developers, misunderstood requirements, technical problems with designs and code, or the varied nature of software end users themselves i.e., some defects are in the eye of the beholder.

Human-centric design and development focus on creating software by placing end-user requirements at the forefront of developers’ focus and effort (Winograd and Woods, 1997). The design and development activities are centred around how end-users perceive a particular task rather than making the users adapt their behaviours to the developed software. Ideally, software requirements need to consider many differences in end users, for example, gender, age groups, cultures, languages, disabilities, socioeconomic statuses, and educational backgrounds (Grundy, 2020; Grundy et al., 2020a; Fazzini et al., 2022). The subjective nature of these differences means that problems can vary for individual users (Khalajzadeh et al., 2022), resulting in defects in the application that we term ‘human-centric defects’ (HCDs). Consequently, it is essential to recognise, appreciate and handle such HCDs responsibly by developers, teams, and organisations for the overall success of a product.

What we term as conventional “technical defects” – defects in software not dependent on end user human differences – typically arise from missing functionality, developer mistakes, or unclear business requirements. Defects are often classified into (i) functional and (ii) non-functional defects. Functional defects violate the working of the system (i.e., functional requirements) while non-functional defects occur when the quality of the system is violated (i.e., non-functional requirements) (Glinz, 2007), e.g., usability, security, performance, and compatibility defects. HCDs may fall under both these defect categories. For instance, while booking an Uber¹, the option to include the child seat is very confusing and requires extra effort from the user side to get it included in their ride. It is further not available in all locations and the option to include a car seat varies within a city. For a user travelling without a child under a certain age, this is not an issue. However, this is an HCD for a parent. This can be classified as a functional defect for the application e.g., no option to book a child seat. An example of a non-functional defect is when colour blindness of the parent impacts the accessibility of booking this feature.

Software defects are usually reported to developers via a *defect report*. Reporters, i.e., testers or end-users, provide these reports to developers. These reports generally have several feedback fields, for example, description, expected and actual behaviour, impact, severity, and supporting supplements (Bettenburg et al., 2008; Zimmermann et al., 2010; Yusop et al., 2016b). If the report feedback is presented to the developers in a tailored way, it can help solve the defects



vedant.chauhan@monash.edu (V. Chauhan);

chetan.arora@monash.edu (C. Arora); hkhalajzadeh@deakin.edu.au (H.

Khalajzadeh); john.grundy@monash.edu (J. Grundy)

ORCID(s):

¹<https://www.uber.com>

faster (Strate and Laplante, 2013; Yusop et al., 2016a). However, research shows that defect reports are frequently incomplete and ambiguous, and the quality varies widely due to inexperience in defect reporting, communication issues, and incomplete defect form fields (Bettenburg et al., 2008; Yusop et al., 2016a,b, 2017). The quality of the information in defect reports can be affected due to a lack of knowledge of the bug reporting process, tools and their experience in reporting issues (Ko and Chilana, 2010; Yusop et al., 2015). The information provided in defect reports often lacks clarity, or detailed information (Davies and Roper, 2014). Additionally, the information presented in many defect reports has a shortage of recommendations about defect solutions (Hornbæk and Frøkjær, 2006) and an absence of logical and consistent textual information and stack traces (Strate and Laplante, 2013). The mismatch in terms of what is required by the developers and what is presented to them further impacts the report quality (Bettenburg et al., 2008; Yusop et al., 2016a,b).

HCDs are very nuanced and are often very subjective defects that occur when different end-users interact with the application. These defects occur due to end-user perception or differences, such as their educational background, age, gender, physical or mental disability, language, experience, and ethnic background. These defects can be considered outliers or edge conditions missed by the development teams while developing or testing their products. Figure 1 illustrates a real-world example of an HCD reported in a healthcare open-source project called 'OpenEMR'². This issue highlights the problem faced by an end-user where the existing videos or materials of the application are outdated. Hence, causing user problems in utilising the features of the application in the best possible sense. For that purpose, the reporter is asking to improve the videos and explanations in the application. Overall, this could also be an issue for a wider audience of the application. The issue report contains only the basic description of the problem and a solution required by the user. There are missing defect report fields such as expected behaviour, screenshots, and additional context regarding the issue. Triaging the issue can be difficult due to specific missing information regarding videos or materials. Additionally, this issue could be a false positive for more experienced users of the application. Hence, it becomes an HCD in the usability defect category.

Existing SE approaches either do not consider human-centric aspects or address them in an ad-hoc manner (Huynh et al., 2021; Khalajzadeh et al., 2022; Prikładnicki et al., 2013; Curumsing et al., 2019). As a result, many critical human-centric aspects are generally not collected, planned for, or addressed while creating or evaluating software (Huynh et al., 2021; Yusop et al., 2016a). Furthermore, some reporters do not effectively use defect reporting tools to capture defect information because of the lack of functionality provided by the tools to capture human-centric defect information (Yusop et al., 2016c; Huynh et al., 2021). Hence, fixing human-centric defects is impacted due to the quality

²<https://github.com/openemr/openemr/issues/5771>

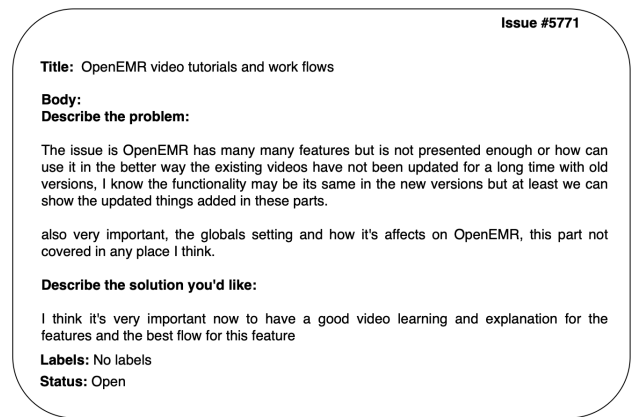


Figure 1: An example HCD reported in the OpenEMR GitHub Repository

of the resultant defect reports. There is a long-standing need for a defect taxonomy due to a lack of distinguishability and incompleteness (Keenan et al., 1999; Yusop et al., 2016a, 2020b; Zhang et al., 2023). However, human-centric aspects still require responsiveness and specificity to be handled by development teams. Teams should be able to handle HCDs responsively and tend to the specific human aspects related to the defect (Huynh et al., 2021). Our research focuses on the subjectivity of defects formed due to user perception or underlying conditions, and extends to HCDs across functional and non-functional defect categories. To the best of our knowledge, this is the first research dealing with human aspects across functional and non-functional defect categories as well as the perception of software practitioners regarding such HCDs.

Due to the challenging characteristics of the human-centric defect reporting process and a lack of awareness about human-centric defects, the goal of this work is to understand the perception of human-centric defects in the software industry. To achieve this, we conducted a survey with 50 respondents and interviews with 10 participants. Based on our analysis of these responses, the key contributions of this paper are:

- an overview of how HCDs are currently perceived by software developers and reporters during the development, testing, and maintenance phase of the software development life cycle (SDLC);
- exploration of current defect reporting tools, their challenges and enhancements required to better support HCDs; and
- identification of improvements needed in managing HCDs, i.e., reporting and fixing HCDs.

Structure of the Paper. Section 2 positions our work against the related work. Section 3 discusses the research questions of this study. Section 4 provides an insight into our research methodology, i.e., our survey design and the interview process. Results and discussion from our survey and interviews

are provided in Sections 5 and 6. Section 7 presents threats to the validity, and Section 8 concludes the paper.

2. Related Work

Numerous studies have been conducted on software defects, defect triaging, remediation, reporting, quality of defect reports, automatic detection and fixing. Strate and Laplante (2013) conducted a systematic review of software defect reporting targeting five potential spaces of defect tracking: triaging, quality, metrics of defect reports, automatic defect fixing and defect detection. Triaging is usually a time-consuming activity, and the SLR identifies automatic assignments, detects duplicate reports, and determines the validity of reports. The quality and metrics of defect reports are also a driving factor in the swiftness of fixing the issue, and SLR manages to identify the gaps in the area. Automatic defect detection and fixing focus on the manual effort in the process and what advancements can be conducted in the domain for faster fixing of defects.

Bettenburg et al. (2008) and Zimmermann et al. (2010) received responses from developers and reporters, revealing that bug reports suffer from poor quality because of missing or incorrect information. According to developers, steps to reproduce, stack traces and test cases are the most helpful information in defect reports, which are usually complex for the users to provide. The research developed a prototype called CUEZILLA to measure and improve the quality of bug reports by recommending elements to include in reports. Consequently, the study suggests improving four areas for bug reports: tool-centric, information-centric, process-centric, and user-centric by engaging the user and enhancing the tool support. If the quality of reporting structure is improved, it can lead to better defect management. Software defect reporting space is a widely explored domain. However, HCDs are still a domain that requires detailed investigation in relation to the whole defect-tracking process.

Several studies have been conducted to capture age, gender, personality, behaviour, culture, and physical & cognitive disabilities. Cruz et al. (2015) conducted a systematic mapping study to gauge the effects of personality on team performance. The study found considerable differences in team performance due to personality and suggested the field is not mature and requires more research. The SLR presented by Soomro et al. (2016) explored developers' personalities and team climate impact on software performance. The study found limited research in the area and identified no agreed team climate composition in the domain of SE. Burnett et al. (2016) and Mendez et al. (2019) introduced techniques to evaluate gender and diversity in software development practices. They developed a magnifier to generate gender and diversity insights into the product evaluation and increase the usability and overall success of the product. Developers' emotions also play a significant role in the success of a product's growth; positive emotions can improve a developer's productivity, and negative emotions can lead to frustration (Müller and Fritz, 2015). The study observed

that training and code review could enhance the results. Lee et al. (2021) analysed accessibility issues under the context of home appliance usage for the target user groups: visually, hearing, spinal cord impaired, and elderly users.

Human aspects and human-aspect-related defect studies (Grundy et al., 2018; Grundy, 2020; Grundy et al., 2020b; Huynh et al., 2021; Hidellaarachchi et al., 2021; Grundy et al., 2021; Haggag et al., 2022; Ahmad et al., 2023; Shamsujjoha et al., 2024) indicate the challenges in getting access to diverse users, capture and description of human-centric issues, absence of adequate feedback to developers, lack of adoption strategies, effort, and sustainability. More research is required to create better awareness regarding human aspects in SE (Grundy et al., 2020b).

Several relevant studies have been carried out in the usability and accessibility domains. For instance, some studies (Bigham et al., 2010; Mbipom and Harper, 2011; Puzis, 2012; Feiner et al., 2016; Moreno et al., 2020) focus on accessibility problems in the website design elements for individuals with low vision, colour-blindness, visually impaired, and cognitive issues. Web developers who create web applications with accessible features usually fail to include usability for disabled individuals. Bigham et al. (2010) provided an Accessibility By Demonstration (ABD) solution, which lets visually-impaired users record accessibility problems. Mbipom and Harper (2011) addressed the design elements in the web context and their interplay with accessibility. The study analysed what types of web pages are aesthetically pleasing or displeasing and examined the barriers to accessibility, such as low vision, colour-blindness, and visual impairment. Puzis (2012) presented the automation tools used in recording and replaying web interfaces based on the history of browsing actions. The system enhances the understandability and readability of textual content for cognitive, language and learning disabilities.

Yusop et al. (2015, 2016b,c, 2017, 2020b) explored usability defects, the defect reporting process and the tools used for such purposes. Usability defects arise when a user experiences inadvertent behaviour while using an application. These defects are reported by expert or non-expert users. The research evaluates reporting and challenges of usability defects and analyses usability defect information in reports. Their work identified the reporting mechanisms used for reporting and tracking issues, such as tool-based, model-based, and end-user-based. Additionally, the format and common attributes such as description, severity, steps to reproduce, stack traces, and expected behaviour, are helpful to solve defects faster. Recommendations to improve the reporting and fixing of usability defects include new taxonomy, guidelines, training for non-expert users, automation in reporting and fixing, and capturing the right attributes to provide the most applicable information regarding the defects. However, there needs to be more relevant research such as these to capture all forms of human-centric software defects. Hence, our emphasis is on improving the current reporting and fixing of HCDs in the SE domain.

3. Research Questions

Our main objective in this study is to understand the occurrence and management of HCDs in current software defect reporting and fixing processes. We formulated the following key research questions (RQs):

RQ1. What is the state of practice for managing human-centric defects in software engineering? We wanted to investigate the current practices around reporting and handling of HCDs by software teams. We aim to understand the differences between handling technical defects and HCDs.

RQ2. To what extent do software practitioners' human aspects – such as experience, culture, emotions, language, ethnicity, gender, and age – affect reporting and fixing HCDs? RQ2 gauges the significance of an individual's traits in identifying and fixing HCDs. For instance, we posit that a practitioner's experience can impact the effectiveness of identifying and fixing HCDs.

RQ3. What is the role of organisations and teams in addressing HCDs? An essential factor in understanding the HCDs management process is the organisation and team's impact on identifying and fixing HCDs. For instance, how much influence do organisations and teams play in representing a human-centric approach, and do they create awareness and necessary discussions regarding human-centric aspects?

RQ4. What role do defect tracking tools play in reporting and fixing HCDs?

RQ4a. What are the key challenges for defect tracking tools for reporting and resolving HCDs? Do defect reporting tools have any challenges in capturing information about reported HCDs.

RQ4b. What characteristics make defect tracking tools user-friendly? What are practitioner recommendations for improving such defect tracking tools.

RQ4c. What are the most prominent defect tracking fields required in a defect reporting form to manage HCDs better? To improve the HCDs reporting, we want to understand what fields are required to describe a HCD so that it can be fixed by a developer quickly.

RQ4d. Does experience with defect tracking tools influence the quality of HCD reports?

RQ5. What improvements can be applied to effectively assist in reporting and resolving of HCDs? Finally, RQ5 investigates the improvements in resolving and reporting HCDs. This RQ provides suggestions for better management of HCDs in the SE domain.

4. Research Methodology

We conducted a mixed-method study. We started with a survey and followed this with detailed online interviews with SE practitioners to understand their perspectives on defining, reporting and fixing HCDs. Sections 4.1 and 4.2 explain the

design process of our survey and interviews. Section 4.3 defines our data analysis process.

4.1. Survey Process: Design and Questionnaire

Figure 2 illustrates our survey design process using the Qualtrics³ platform. We constructed our questions based on the Kitchenham et al. principles (Kitchenham et al., 2002; Kitchenham and Pflieger, 2002), authors' experience in the IT industry as well as from multiple studies on defect reporting, industry practitioners' and human-centric aspects (Bettenburg et al., 2008; Yusop et al., 2016a,b,c; Votipka et al., 2018; Yusop et al., 2020b; Huynh et al., 2021; Hidellaarachchi et al., 2021). The total number of questions in the survey was 51, divided into five sections, and the time to complete the study was around 15-20 mins on average. We grouped our survey into five sections: Demographics, Selection, Developer, Reporter, and Conclusion:

- **Demographics.** Collects basic demographic information of the participants. The survey was anonymous, and we did not record any identifying information of participants.
- **Selection.** Provided participants with a choice based on their previous or current experiences. If the participants had experience developing and working on fixing the defects, they could select the developer section, and the participants were presented with only developer-related questions. Similarly, if the participants had experience in reporting defects, then they could select the reporter section, and the participants were presented with only reporter-related questions. If the participants had experience in both, they could select both.
- **Developer and Reporter.** The third and fourth sections ask about how developers and reporters perceive HCDs. A first part asked about perceptions of HCDs, and a second part regarding HCD management and defect tracking tools.
- **Conclusion.** Asked for feedback regarding the survey and inclination towards a follow-on interview.

Our survey questionnaires are **available on Zenodo⁴**, please see Vedant et al. (2022) for more details.

After designing the questionnaire, we conducted a pilot study with industry and academic participants to validate our survey clarity, suggestions on improving the survey, and time to complete the survey (Baltes and Ralph, 2022). We adopted a convenience sampling approach for both the pilot and final study by distributing the survey to our professional contacts who satisfied the specified criteria for inclusion. Convenience sampling at pilot stages allowed us to test the survey with individuals we knew had relevant experience and were easily accessible. The target criteria consisted of

³<https://www.qualtrics.com/>

⁴<https://zenodo.org/>

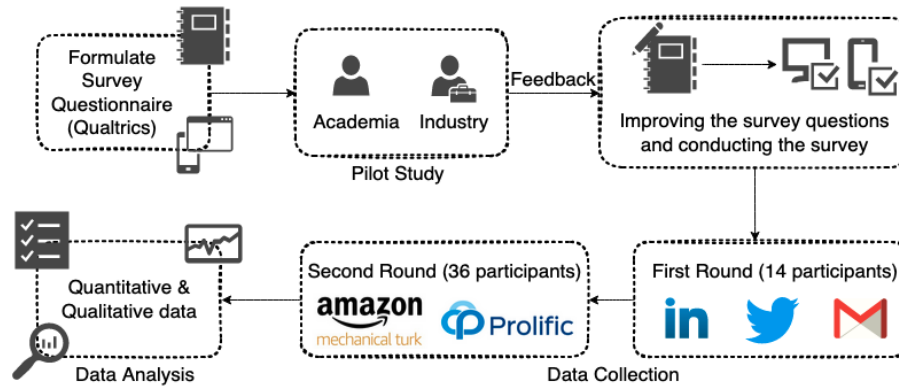


Figure 2: Survey process

software developers, reporters, testers, requirements engineers, customer-centric engineers, data engineers, data scientists, UI/UX professionals, and project managers who have experience in developing or managing software applications. We presented the study to four industry practitioners and one academic with prior experience in software development and reporting. We added some examples and definitions to the questions based on the feedback to provide more transparency. For instance, in one of the HCD examples, we have mentioned “missing keyboard shortcuts in Web applications causes problems with accessibility”. One of the pilot study participants suggested improving the HCD examples to something concrete that can justify the impact. Hence, we provided the Uber sample as an example in the surveys. For the final study, we focused only on participants who met the experience criterion in the roles noted above. In order to mitigate selection bias, we employed social media networks such as LinkedIn and Twitter in the final study.

We ran our survey in two rounds. Each round was conducted over two months. First-round was conducted from Dec 2021 to Jan 2022 and was advertised on social media platforms such as LinkedIn⁵ and Twitter⁶, through the authors’ colleagues, industry contacts, and via email. We received 40 responses, of which we selected 14 valid responses for further analysis after a preliminary check. Unfortunately, not-selected responses consisted of empty or incomplete responses for closed-ended questions, responses with a standard repeated answer for multiple questions, or the users who did not provide consent. In the second round, we decided to use crowdsourcing platforms for survey collection, such as AMT⁷ and Prolific⁸, where we paid participants. We paid 6.40 USD/hour on AMT and, on average, 7.12 GBP/hour on Prolific for each response. This round was conducted from Feb 2022 to Mar 2022. We received 64 responses, out of which we selected 36 valid responses for data analysis. Similar to the first round, the not-selected responses consisted of

empty or incomplete responses for closed-ended questions, standard repeated answers for multiple questions, and the participants who did not provide consent or did not have relevant experience in the select roles.

4.2. Interview Process: Design and Questionnaire

Figure 3 illustrates our interview design process. We constructed our questions based on the Kitchenham et al. principles (Kitchenham et al., 2002; Kitchenham and Pfleeger, 2002), author’s experience in the IT industry as well as from multiple studies on defect reporting, industry practitioners’ and human-centric aspects (Bettenburg et al., 2008; Yusop et al., 2016a,b,c; Votipka et al., 2018; Yusop et al., 2020b; Huynh et al., 2021; Hidellaarachchi et al., 2021). The total number of questions in the interview was 18, divided into two sections, and the time to complete the interview was around 30 mins. We grouped our interview study in two areas: demographics and discussion of HCDs. An overview of the interview questions is as follows:

- **Demographics.** The first section contained participants’ demographic information. The interview was anonymous, and we did not record any personally identifiable information of participants. After recording the session, we transcribed the interview and deleted the recordings.
- **HCDs Discussion.** The second section contained a semi-structured set of questions designed to keep the conversation going when discussing the themes such as HCDs examples, the influence of emotions and culture, defect reporting, improvements, taxonomy, and automation related to human aspects.

Our interview questions are **available on Zenodo, please see Vedant et al. (2022)** for more details.

After designing the interview questionnaire, we again piloted the study with industry-experienced academic researchers to validate our interview questions. We employed convenience sampling for the pilot study, selecting our professional contacts who satisfied the inclusion criteria, i.e., had relevant experience in specific SE roles. For the final

⁵<https://www.linkedin.com/>

⁶<https://twitter.com/>

⁷<https://www.mturk.com/>

⁸<https://www.prolific.co/>

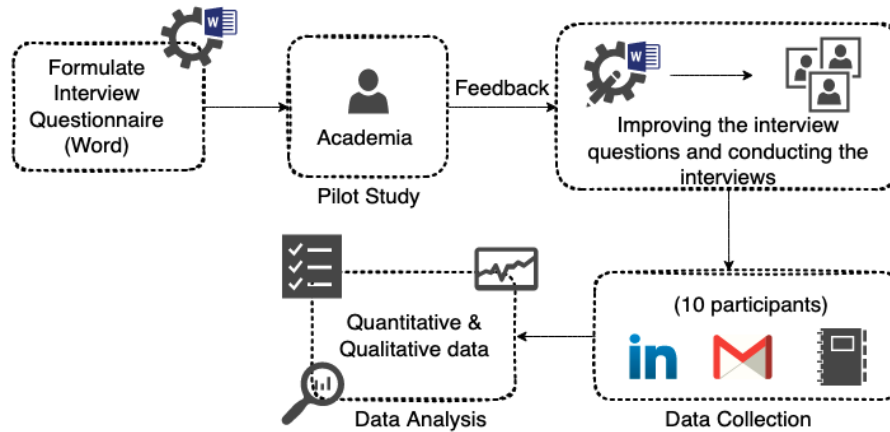


Figure 3: Interview process

study, we utilised the participants who voluntarily indicated their willingness in surveys to participate in the interviews. The target criteria consisted of software developers, reporters, testers, requirements engineers, customer-centric engineers, data engineers, data scientists, UI/UX professionals, and project managers who have experience in developing or managing software applications. The participants who did not belong in the selected roles were excluded from the study. In order to mitigate selection bias, we employed social media networks such as LinkedIn in the final study. We presented the interview study to three academics with prior experience in software development and reporting. As a result, we divided some questions, modified the wording of questions, and delivered some examples based on the feedback to provide more transparency. With the improvements, we finalised the interview questionnaire. For instance, in the pilot study, we received feedback to add questions on practitioners' experience in dealing with HCDs in their respective projects. This proposal was beneficial in offering practitioners clarification on HCDs. If someone lacks comprehension of the notion, we, as interviewers, can provide a detailed explanation. We then conducted our detailed interview study after completing the second round of the survey. We had two motivations to conduct an interview study. First, the survey responses are limited in depth, as typical with surveys. Second, surveys are often restrictive in providing sufficiently detailed responses to key questions (Couper, 2000). In our case, we wanted to understand more about the current industry practices impacting HCDs. We did not have dedicated sections on the developer and reporter profiles, but the second segment of our interviews focused on covering HCDs both from the developers' and reporters' perspectives.

We conducted the interviews from Apr 2022 to May 2022, all using Zoom and recorded, due to the interviewees' geographic distribution. Of the ten professionals interviewed, only four had not previously completed the survey and were industry contacts of the authors. Post interviews, we provided our interview participants with a choice of a

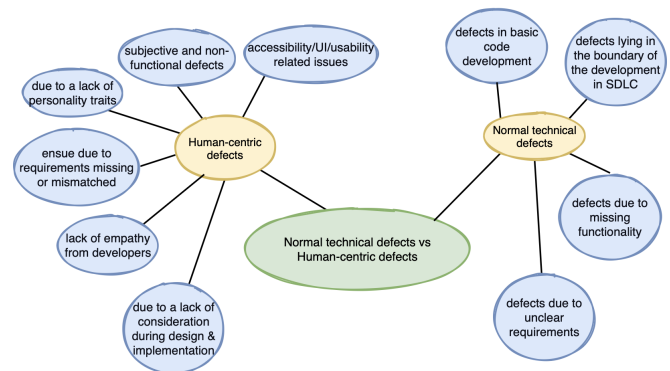


Figure 4: Theme map for RQ1

\$20 gift voucher in their currency from Myers⁹, Coles¹⁰, iTunes¹¹, and Amazon¹² to thank them for their time and participation in our interview study.

4.3. Data Analysis

We collected both qualitative and quantitative data using surveys and interviews. First, we analysed survey responses. Based on the analysis we identified general themes which were spread across interviews. Since the interviews were a deeper extension of the survey answers tackling more prominent and detailed themes, we decided to combine the analysis of both and generate combined themes from them. We utilised descriptive statistical analysis on the quantitative data using the Qualtrics platform, as shown in Section 5.1.

For the qualitative data, we used thematic synthesis (Cruzes and Dyba, 2011). The thematic analysis identifies, analyses, and reports patterns or themes within qualitative data (Braun and Clarke, 2006). However, if the thematic analysis is not applied inside an underlying conceptual framework, it has minimal explanatory ability beyond basic

⁹<https://www.myer.com.au/>

¹⁰<https://www.coles.com.au/>

¹¹<https://www.apple.com/au/itunes/>

¹²<https://www.amazon.com/>

description (Braun and Clarke, 2006). Hence, we decided to use thematic synthesis. Thematic synthesis builds on the principles of thematic analysis and identifies, explains, and summarises the themes captured in the process. We utilised this concept to determine the recurring patterns in the qualitative data and establish the purpose and summary of those patterns for our overall research questions. After analysing the qualitative data, we identified specific text segments in the data. We then labelled the text using an open coding technique to build ideas, concepts, and overall themes (Khandkar, 2009). For instance, we coded the survey responses to ‘What do you think human-centric defects mean?’ and in the interview, ‘According to you, what is the difference between normal defects and human-centric defects?’, using similar and consistent codes, e.g., “Human-Centric Defects Meaning”. Figure 4 provides a sample theme map for RQ1 (see Section 5.2). It shows all the codes for the theme of HCD and normal technical defects. These codes then create lower-order themes such as normal technical and human-centric defects. Finally, these themes are further categorised into higher-order themes, such as normal technical defects vs human-centric defects. Similarly, we used the open-coding technique for open-ended questions by constant comparison, creating lower-order themes and merging them into higher-order ones.

5. Results

5.1. Demographics

Survey respondents are referred to as *SR-x* and interview participants *IP-x*. When referring to both respondents and participants in this paper, we refer to them generally as practitioners.

5.1.1. Survey

Table 1 shows the demographics of the 50 survey respondents, which were considered in our final analysis. Our target participants were developers and reporters. We thus wanted to gauge the technical skills and roles of the respondents in software development and testing activities. Figure 5 shows the skills measured on a Likert scale of 1 to 5, with 1 being a beginner or no experience and 5 being the expert. The scale ranges in the order of novice/no experience, advanced beginner, competent, proficient, and expert (HI and Dreyfus, 1986). Figure 5 provides a mean plot of the development, testing, database, usability, administration, and security skills with the respondents’ experience in these skills (Florea and Stray, 2019). The respondents had the most experience in programming, and mobile development was rated as the one with the least experience. The most common skills were web development, testing, and user interface design. We received 22 responses from only developers, 16 responses from only reporters, and 12 from respondents who were experienced, as both.

5.1.2. Interview

Table 2 shows the demographics of the 10 interview participants. We received most responses from Australia

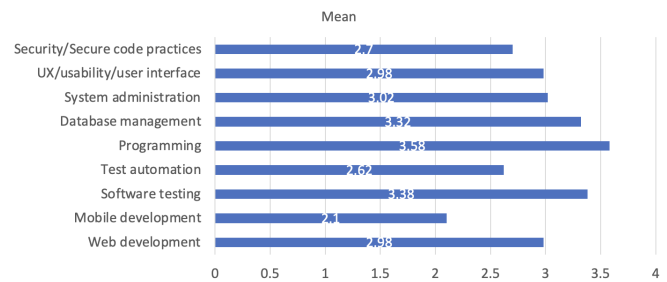


Figure 5: Mean plot of technical skills of the respondents (in Surveys)

(50%), while the remaining participants were from Canada, the USA, Germany, and Romania. All ten participants had a university degree in Computer Science, Information Technology, Data Science, and Software Engineering. We interviewed participants with varied job roles: developers, data scientists, testers, consultants, and customer-centric innovation engineers. All the participants had previous or current experience in software development or reporting issues.

5.2. RQ1 Results - Current state of the practice for managing HCDs

We asked about practitioner’s views on the disparities between technical defects and HCDs, as well as defect categories being reported in the industry. To comprehend the management of HCDs in current industry practice, we identified the following key themes from the responses:

5.2.1. Frequently reported HCD categories

We presented respondents with a Likert scale of ‘never, sometimes, often, and always’ to highlight HCDs in defect categories, such as functional, performance, security, usability, and compatibility. Figure 6 shows a bar graph with the usability defect category comprising most HCDs. The graph shows 36 responses out of 50 survey responses for the usability defect category under ‘always’ and ‘often’ scale. The security defect category contains the least HCDs according to the responses. Overall, the usability category is followed by functional and compatibility and concludes with the performance and security defect categories, respectively.

Some interview participants provided examples of HCDs in functional and non-functional defect categories based on their personal experiences. For instance, an interview participant explained a usability issue in Microsoft Teams¹³ and reported: “This is what I experienced daily, in my meetings (on MS Teams) - when we blur the background, it instantly recognises the face of one of my colleagues. I’m a coloured person, it does not recognise my face, and it is just blank. It takes some time to recognise my face.” - [IP-4].

¹³<https://www.microsoft.com/en-au/microsoft-teams/group-chat-software>

Table 1
Survey participants' demographics information

Demographic Information of the participants	Overall
Countries of the participants	
Australia, the UK, Mexico, Poland, and South Africa	50%
Chile, Portugal, USA, and Italy	24%
Spain, Romania, Canada, Netherlands, Iran, Greece, New Zealand, Austria, and Denmark	18%
Sweden and India	8%
Age ranges of the participants	
25-29	32%
30-34	20%
40-50	18%
20-24	14%
35-39	6%
51-60	6%
60+	4%
Gender of the participants	
Male	60%
Female	40%
Highest Education	
Bachelor's degree	44%
Master's degree	32%
High school graduate, diploma, or the equivalent and Some college credit, no degree	12%
Doctorate degree	4%
Trade/technical/vocational training	4%
Associate or Professional degree	4%
Education information of the participants	
Computer Science, Information technology, Computer Engineering, Data Science, Programming, and Information Systems	86%
Arts and Humanities, Business, Education, Engineering Physics, and Civil Engineering	14%
Job roles of the participants	
Software Engineer/Software Developer/Programmer	48%
Operations, Consultant, Database Engineer, UI/UX, and Security Engineer	20%
Project Management, Business Analyst, and Partner Manager	14%
Test Engineer/Tester/Quality Assurance	10%
Delivery Engineer/Lead	8%
Total Experience of participants in IT industry	
3-5 years	34%
9+ years	28%
0-2 years	22%
6-8 years	16%
Total Experience in defect reporting	
0-2 years	64%
3-5 years	16%
9+ years	14%
6-8 years	6%
Total Experience in defect reporting tools	
0-2 years	67%
3-5 years	20%
9+ years	13%
6-8 years	0%
Field of Work	
Information Technology (IT)	24%
IT-Finance	16%
IT-Healthcare	16%
IT-Manufacturing	14%
IT-Spatial	10%
IT-Gaming	8%
IT-Government	6%
IT-Telecommunications	6%

Table 2

Interview participants' demographics information

Demographic Information of the participants	Overall
Countries of the participants	
Australia	50%
USA, Germany, and Romania	30%
Canada	20%
Age ranges of the participants	
25-29	50%
30-34	30%
40-50	20%
Gender of the participants	
Male	90%
Female	10%
Highest Education	
Master's degree	70%
Doctorate degree	20%
Bachelor's degree	10%
Education information of the participants	
Computer Science, Information technology, Computer Engineering, Data Science, and Software Engineering	100%
Job roles of the participants	
Software Engineer/Software Developer/Programmer	40%
Data Scientist	20%
Consultant and Customer-Centric Innovation Engineer	20%
Test Engineer/Tester/Quality Assurance	10%
Data Engineer	10%
Total Experience of participants in IT industry	
9+ years	40%
3-5 years	30%
6-8 years	30%
Field of Work	
IT-Finance	30%
IT-Healthcare	20%
IT-Government	20%
IT-Spatial	20%
IT-Manufacturing	10%

5.2.2. Practitioners' perception of the differences between HCDs and technical defects

From the gathered responses, normal technical defects arise for everyone during code development, due to missing functionality, or unclear business requirements. For example, IP-10 claimed: *“Normal technical defects are the basic development defects, which come to developers when the code is not working as per the functionality or as per the business rules that have been defined”*. However, HCDs are subjective defects that occur due to the lack of consideration of human aspects, empathy, and missing requirements (Arora et al., 2019; Grundy, 2020; Gunatilake et al., 2023). For instance, a survey respondent commented that *“HCDs are biased, uncoordinated, inaccurate, or imperfect results during project development.”* - [SR-48].

Overall, 96.7% of our survey and interview responses state there is a similarity in technical defects and HCDs as they both arise in SDLC due to missing or unclear requirements. However, there is a difference as HCDs are perceived as more subjective and technical defects are more definitive. HCDs arise from the necessity of an individual, the philosophy of individuals, teams and organisations, and the human perspective. They are unique and hence, considered differently from normal technical defects. For instance, a survey respondent on HCDs mentioned: *“Since each person is unique, these human-centric defects appear out of necessity, either because of the culture, type of learning, and experience of each person.”* - [SR-47].

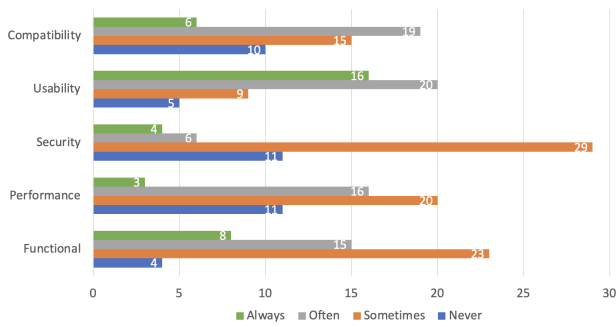


Figure 6: Defect categories provided by developers and reporters (in Surveys).

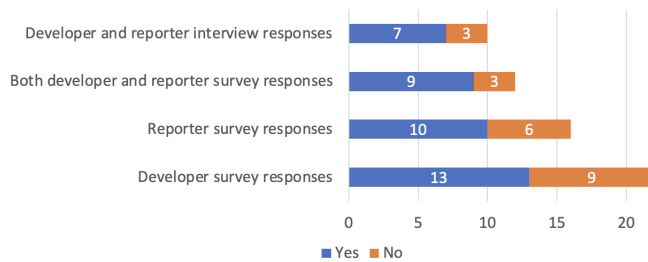


Figure 7: Bar chart of HCDs experienced by developers and reporters (in Surveys and Interviews).

5.2.3. Approaches for identification of technical defects and their application to HCDs

Practitioners provided their experiences in defect identification, triaging, and resolution to understand the nature of software defects and their identification. They also showcased some examples and methodologies which can be utilised for HCDs. As shown in Figure 7, over 65% respondents experienced what they defined as HCDs in previous or current projects. Our respondents and participants work in different application domains, such as health, finance, manufacturing, government, telecommunications, and information technology, as shown in Tables 1 and 2.

Practitioners reported that they identify defects by following one or more of the following steps: (i) reviewing bug reports, log files, invalid logical flows, requirement specifications, expected results vs. actual results, and similar errors detected by the specialists; (ii) trial and error; (iii) categorising issues into functionality, performance and usability; (iv) considering how a user will perceive the problem based on needs, ease of use, accessibility and expected behaviour; (v) performing peer reviews; (vi) creating models, e.g., data models and UX flows; (vii) positive and negative testing of data input possibilities; and (viii) recreating the user path.

On defect identification, SR-3 commented that *“While categorising or segregating issues, I try to divide the problems into functional, performance and usability. Many human-centric issues can be addressed while writing the functional requirements. However, just addressing functional requirements might not resolve the problem once the application is developed. We might have to dive as deep*

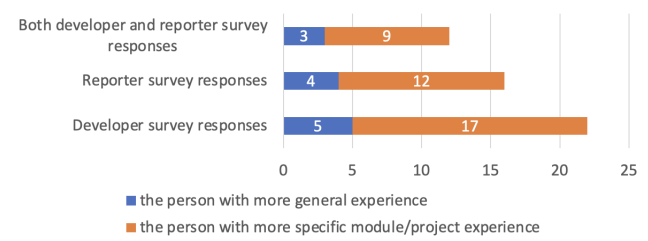


Figure 8: Bar chart of defect assignment (in Surveys).

as changing the fundamental architecture or database at a granular level, like entity-relationship mapping.”

Most respondents claimed that HCDs could benefit from a similar identification process, such as developers understanding user needs and resolving issues. For example, some survey respondents shared their perspectives on the identification process: *“Who, what, and why are the questions I ask while fixing an issue. Who raised the issue, what caused the issue and why the issue was present in the first place. Yes, the same questions can be applied to human-centric defects as well.”* - [SR-2].

“Identification is a process of trial and error, but I always try to think on how to break the design and the same could be applied to human-centric defects since you could always put yourself in someone else’s shoes.” - [SR-44].

5.2.4. Software defect assignment to the practitioners and their application to HCDs

Defect assignment depends on the specific software module and general experience of practitioners. We define general experience as the total experience in the industry and the module experience as the total experience in a specific module of the project or the project itself. How particular HCDs are assigned to practitioners is an interesting question. 76% of our respondents observe that a defect could be assigned to someone with more specific project experience, as shown in Figure 8. Respondents claimed that it is more practical to allocate defects to the individual with specific module/project expertise where the defect occurs. For example, a respondent claimed, *“Usually developers who had previously worked on a module, can resolve any errors within the module faster given their understanding of the module. Similarly, human-centric defects can be mitigated easily if developers understand the entire paradigm of those defects, their causes beforehand and formulate a solution accordingly.”* - [SR-4].

In our interviews, we asked for the rationale behind allocation of HCDs based on expertise. We found that organisations and teams utilise expertise to help provide the best solution in a shorter time due to the methodologies used, such as Agile. Project and resource pressure is usually the time pressure that occurs due to meeting release deadlines and is one of the drivers in assigning defects to a developer. For instance, any developer who has worked on a given module will also be responsible for resolving the issues in the module within a sprint. An interview participant suggested,

“Since the work is in an Agile methodology, we will first check the availability of a person. Therefore, if a defect is encountered and a person who has previous experience with the defect, but working on a different assignment should not be encumbered with the defect. However, if the defect priority is high, the person has to prioritise it above other tasks. Generally speaking, preference will be given to the person who has already worked on the defect. For example, XYZ defect has been encountered, and there are two people A and B. Person A has previous experience with the defect and the module. Hence, Person A will be given the precedence to work on the defect.” - [IP-10].

Practitioners further noted that the defects are assigned to other developers if the responsible developer is unavailable. For instance, an interviewee stated on defect assignment, “If someone is involved in feature development and they are not available to look at the defects, then generally whoever else is available is assigned the defect.” - [IP-9]. Sometimes, the assignment of defects also depends on the nature of the defect. If the defect falls under human aspects, the practitioner with more experience in HCDs should be assigned the defect. For instance, some survey respondents indicated: “In general, the person who has the most experience with defect reporting process will work on the defect. For human-centric defects, it might make sense to have someone with a more general experience to take a second look.” - [SR-30]

“For sure someone who has more experience on human-centric defects can be faster and better at resolving them” - [SR-33].

An argument was that since HCDs are often subjective, any developer can understand the defects. For instance, a survey respondent claimed, “I suppose that human-centric defects can be assigned to any developer because any developer can understand the principles of human-centric defects, while normal defects require specific module experience to be solved” - [SR-22].

RQ1 Key findings.

- HCDs are often reported under the usability category, followed by functional, compatibility, performance, and security defects.
- Normal technical defects are more definitive and arise from usage, design and development. In contrast, HCDs are subjective and derive from individual differences based on different human aspects, e.g., age, culture, language proficiency, emotions, and experience.
- Software practitioners identify technical defects by looking at bug reports, log files, trial and error, categorising issues, checking requirement specifications, and checking the application’s functionality, performance, and usability.
- Software defects are assigned to developers with specific module expertise rather than the general overall experience.

5.3. RQ2 Results - Extent to which the human aspects of software practitioners affect reporting and fixing HCDs

We define human aspects as the set of differences between individuals, e.g., which culture or ethnic group they belong to, their experience, emotions, educational background, gender, age, and the languages they speak (Khalajzadeh et al., 2022; Grundy, 2020). Based on practitioners’ feedback, we identified that human aspects are an influential factor in how developers or reporters perceive human-centric defects in general. The key human aspects identified by our survey and interviews include: (i) knowledge and experience of an individual; (ii) communication and language; (iii) emotional background (individuals experiencing emotions which reflect their mental state (James, 1922) is termed as ‘emotional background’ of an individual); (iv) diversity, ethnicity, culture, and age.

Different human aspects of developers and reporters need to be identified and respected by leadership in organisations and teams. Individuals and teams should be flexible enough to look beyond just their own personal and organisational values and environment to identify HCDs. For example, an interview participant claimed, “If the developers are not flexible to look beyond their cultural values, the environment they work in or adapt to the conditions, they will not be able to discover human-centric defects which are global in nature” - [IP-1].

94% of respondents claim that the **experience of a practitioner plays a crucial role**. The experience level is directly proportional to the level of understanding of a defect and the efficiency in resolving the issues. For instance, SR-27 highlighted, “Experienced workers can quickly identify complicated issues. They have the experience to notice things which others ignore”. Experience provides the knowledge and learning to solve a variety of problems. Experienced developers and reporters can anticipate issues and provide the best possible mitigation. For instance, some survey respondents claimed:

“Experience is easily the best mitigation. If you work in the same field for many years, the same problems (human-centric or other) arise and you’re better equipped to solve them.” - [SR-1]

“While the defect context and developers context have common points, for example, same age, same language, etc. Experiences can be useful for understanding problem statement” - [SR-7].

The remaining 6% participants disagreed that experience helps identify and fix HCDs. According to them, if HCDs are related to inclusiveness and diversity, experience may not have much of a bearing on understanding human concerns. For example, SR-3 posited:

“Experience may or may not help in fixing human-centric defects. For example, if the experience has been in an environment full of people from diverse regions/countries/cultures, we can learn about human-centric issues and fix them. At the same time, experience with people from a limited

regional area might not provide good insights regarding human-centric defects.” - [SR-3].

A vast majority of practitioners highlighted **communication as an important factor** in the understanding of HCDs from impacted end users. Developers and reporters do not directly handle client feedback in many organisations. They are often managed by business analysts or product owners and communicated down the line to developers. In some cases, there are scenarios that lead to mishandling of user requirements as people who are building software are different from those who are testing or managing. For instance, IP-1 suggested, “A big issue is communication. In my experience, you usually have a team of people with one person as the contact point to the client, and other people who are going to be using the software. The issues need to be communicated to a range of people i.e., developers who are not on the front line, they’re not sort of dealing with the client directly”.

Emotions are diverse phenomena that represent distinct body expressions, e.g., happy, sad, fear, disgust, and so on (Frijda et al., 1986). Individuals experiencing these emotions reflect their mental state (James, 1922) and hence we term it as the **‘emotional background’ of an individual**. Sometimes emotions of a developer can impact the solution to an issue. Developers’ mental state can be sometimes stressful while working on issues under pressure situations. For example, IP-6 claimed, “Developers usually work in pressure situations due to deadlines and constant issues with projects and working environment can lead to stress”. Some individuals claim that knowledge and experience play a more significant role than the ‘emotional background’ of a developer. For example, an interviewee claimed, “I think emotions play a very significantly less factor than the background experience. For example, experience and knowledge of your field” - [IP-3]. As a result, this helps in providing a much better end-user experience in their software.

Including **diversity in the development teams** can provide different perspectives and experiences to aid in understanding and appreciating HCDs. In addition, different perspectives bring new ideas and opinions that can help design, implement, and test solutions. For instance, IP-10 suggested, “Diversity in your team is always a plus point. Because it brings different points of views and thinking which when collaborated can help in achieving the best solutions to the tough questions as well”. Exploration of different perspectives and values can be experienced when you reside in a particular community where you explore different perspectives and values which shape your knowledge on the working of the society. This can be beneficial in providing the best possible solution. For instance, IP-3 claimed, “...cultural diversity can help in elevating some of the issues and lead to best possible solutions”.

RQ2 Key findings.

- Most software practitioners’ experience plays a crucial role in identifying and fixing HCDs.
- Communication and translation of issues can lead to proper handling/mishandling of user requirements, particularly those that result in HCDs.
- Developer’s emotions can impact the resolution of HCDs. However, the knowledge and experience of a developer are deemed more crucial in identifying HCDs than emotions.
- Diversity in the teams can provide enriching experiences and unique perspectives to design, test, and develop solutions for managing HCDs.

5.4. RQ3 Results - Role of organisations and teams in addressing HCDs

Organisations and teams want to deliver high-quality products (Mockus, 2010; Bass, 2015), thus it is essential to understand the role organisations and teams play in handling HCDs. Organisations and teams do not seem to handle these HCDs comprehensively. We identified the following key themes from our analysis.

5.4.1. HCD Awareness (or lack thereof)

First, there is a lack of awareness regarding HCDs within teams and organisations, which creates a problem when communicating these defects from project managers or business analysts to the developers. Two interview participants reported, “I’m an experienced software developer having experience of eight to nine years. I have not encountered the topic of human-centred defects before two or three years ago. So it came quite late into my information that these issues also exist” - [IP-1].

and “I think most people in organisations will not be looking into human-centric defects because first of all, the organisation will have to provide awareness regarding this” - [IP-9].

Secondly, due to the extensive workload and the structure of delivery and planning, there could be a possibility that organisations’ and teams’ prioritisation lies in satisfying the product functional deadlines and not fixing HCDs. For instance, IP-3 reported, “I think planning can solve these issues. However, sometimes on short deadlines, the time for planning is less and then human-centric problems can arise frequently. Therefore, it becomes an ongoing cycle of rectifying these problems”. There should be some change management from the organisations to highlight HCDs; otherwise, developers will solely focus on the code and release of the product before the deadline. A customer-centric process for finding defects should be undertaken to highlight HCDs and resolve them faster. IP-6 highlighted this by saying “...there should be some change management from the organisation to highlight human-centric defects better otherwise, a developer will only focus on releasing the code. There should be a customer-centric thought process for finding these defects better”.

Similar to our findings in RQ2, practitioners highlighted the importance of diversity in teams and organisations. They

emphasised that “Diverse teams will come up with different scenarios and ways in which people will use the product because of their upbringing and environment” - [IP-5].

“In case the organisation is an IT consulting firm or develops multiple software products for clients, the company can set up a community of reviewers from a diverse group. Human-centric defects should be listed and shared with all so that similar issues do not occur repeatedly” - [SR-3].

5.4.2. Organisation’s size impacting HCDs

A large majority of respondents (84%) believe that an organisation’s size impacts the handling of HCDs. Of these, there was a balance in terms of the responses on the advantages and disadvantages of a small and large organisation handling HCDs. For instance, SR-1 claimed, “Plus and minus for both: bigger organisation has more resources to identify/address the problem, but requires more communication (chain of fixers are often further away from end-user experiencing the issue). I think these issues balance each other out”.

Large organisations can be more bureaucratic, take more time for addressing changes, and the communication chain is longer. However, large organisations have more resources and budget to manage HCDs and likely have more diverse team members who may appreciate some HCDs than others. For example, survey respondents reported, “Larger organisations would have different cultures involved and may make escalating to a relevant person harder, but they would have a wider base of experience” - [SR-18].

“If the organisation is bigger there is more people who might be able to fix the problems faster” - [SR-21]. While another survey respondent suggested, “This question again does not have one fixed answer. On one side bigger organisation with a huge number of diverse users can help us quickly identify different kinds of issues and can lead to fixing human-centric issues faster. Whereas a big organisation may also have a bigger chain of commands, getting approval on a problem or a ticket might take longer than usual due to a bigger hierarchy” - [SR-3].

Smaller organisations may lack the resources to handle or test some HCDs. They may also want to release a product faster to the market to gain a competitive advantage and, in turn, may lose focus on HCDs or de-prioritise some of them. However, the HCDs can be reported faster due to less chain of command and can sometimes be resolved more quickly. Some survey respondents claimed:

“I do not think human-centric defects are considered by smaller organisations at all. Its not that the smaller organisations do not want to solve them; they just do not have the resources to test the products extensively.” - [SR-13].

“The smaller the team, the quicker the bugs and dysfunctioning reach the developers” - [SR-20].

“It depends on the response time of the organisation. The smaller the organisation, the faster bug may be reported to the developers and they might fix the issue faster. But sometimes developers may be working on some other project which is why they do not have time to fix the bug. This is

where bigger organisations are better. Since they have more resources, it is easier to fix these bugs, but the downside is, it is difficult to report and get the bug to a stage where someone actually notices it and works on it” - [SR-2].

Sixteen percent of responses suggested that organisation size does not matter. They claimed if the teams are diverse, inclusive, and experienced enough, HCDs can be handled better. Additionally, if the developer follows good practices while developing software, it could be critical in handling HCDs. For instance, a survey respondent insinuated, “Size of the organisation does not matter but by looking at the innovation and inclusive team in a company we can determine which organisations lead this case” - [SR-10]. Another survey respondent claimed, “Based on my experience, I do not think it is the size of the organisation which causes these defects. But it may be due to the practices followed by the developers when developing the software. When the developers have different experience levels and work ethics, there is a chance different development practices can lead to human-centric defects” - [SR-5].

5.4.3. Team climate impact on identifying and fixing HCDs

Employees’ collective impressions of organisational policies and processes are referred to as team climate (Anderson and West, 1998). Based on the responses received in the survey, most respondents claimed team climate is essential in identifying and fixing of HCDs. For example, SR-10 mentioned, “Team climate plays a major role to allocate time and resources to human-centric defects and push for more fixing of defects”. Team philosophy, diversity, and culture create an impact on the overall success of the product. If the team culture promotes a human-centric process, then HCDs can be fixed faster; otherwise, developers will always ignore blind spots and edge cases. For example, some survey respondents proposed:

“The different cultures of the testers, developers and end-users has an impact on fixing and reporting of the human-centric issues.” - [SR-18].

“Depending on the defect - a cultural blind-spot would affect the identification of the issue.” - [SR-1].

“If the team culture is to prioritise human-centric issues, then these issues will get fixed faster” - [SR-2].

Team climate impacts the motivation, creativity, and productivity of a team delivering a product. The more diverse a group, the better the quality of the product. Additionally, it reduces communication delays and fastens the process of resolving issues. For example, SR-30 claimed that “A better team climate will generate more ideas, solutions and positive feedback. This will help fixing defects, but also identifying them more easily”.

A shared perception can help focus the efforts, the issues to work on, and the issues to solve for better product quality. An organisation’s cultural and sensitivity training could be a big part of the team’s thinking while building software. For example, a survey respondent suggested, “Mainly the attitude of the team, whether they are motivated

or dissatisfied, is reflected in the possibility of finding defects” - [SR-47].

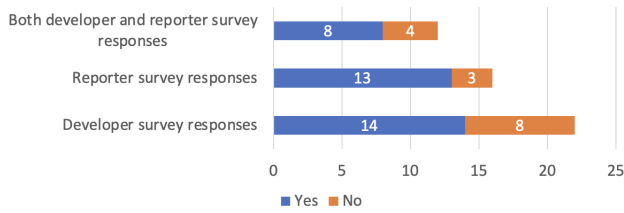


Figure 9: Bar chart of geographic location impact in finding and fixing HCDs (in Surveys).

5.4.4. Geographic location’s role in finding and fixing HCDs

Based on the answers we received from the survey respondents as shown in Figure 9, 70% of respondents asserted that geographic location may add some delays in fixing and reporting of HCDs (or, defects in general). Distributed teams can be difficult to handle when the goal is to build a single standard software. There will be more discrepancies if the developers are located in different geographic locations because fixing a simple issue can take more hours due to varied working hours. For example, a survey respondent stated, “A defect identified in the USA time zone, has to wait for the India team to login and witness the issue. This can lead to wastage of hours to fix an issue” - [SR-11]. In addition, communication across time zones restricts the progress and reporting of the problems. For instance, some survey respondents indicated:

“Access to resources and time play a huge role in finding defects. These factors can delay finding defects at different locations.” - [SR-26]

“Offshore and timezone differences may lead to delay in resolving issues if there is no proper task management methodology or lack of communication.” - [SR-3]

Due to cultural and environmental differences, issues arising from different geographic areas can have translation problems. Native slangs could affect the bug terminology, making the HCDs longer to interpret. For example, a survey respondent claimed, “Translation issues are the biggest “human centric” defect as it can lead to delay in fixing issues” - [SR-13]. The quality of reports can suffer due to spelling and grammar and even lead to the escalation of issues. Due to cultural differences finding a common solution for an HCD could also be a concern. For instance, a survey respondent stated, “Different locations have different ways of living, the culture might not be the same. So having a homogenous solution to a problem might serve one community while it might be detrimental to another. Those nuances in finding solutions definitely can delay a fix” - [SR-37].

RQ3 Key findings.

- More awareness about HCDs in organisations is required.
- Diverse teams can expedite the process and help identify HCDs at earlier stages of the SDLC.
- Organisation size could impact reporting and fixing of HCDs. Larger organisations can have time-consuming processes but may have more resources to manage HCDs. Smaller organisations often lack such resources and may focus on quick-release cycles, but HCDs might be resolved swiftly due to the flexibility in such organisations.
- Majority of participants suggested that HCDs from one location and fixed in another could lead to significant delays. Issues from different geographic areas can have translation problems due to cultural and environmental differences.

5.5. RQ4 Results - Role of defect tracking tools in reporting and fixing HCDs

We identified that most of our participants have significant experience in using defect reporting tools. Table 3 shows Jira and Trello¹⁴ are actively used closed-source tools, and GitHub¹⁵ is the most actively used open-source tool by developers and reporters. The following sub-questions explore the existing defect-tracking tool handling of HCDs.

5.5.1. RQ4a Results - Key challenges for defect tracking tools in reporting and resolving HCDs

81.7% of our practitioners claimed that defect tracking tools have a generic structure that does not provide many options to manage HCDs. One stated, “There is no option to handle human-centric issues in these tools. We usually create a regular Jira ticket without customisation” - [IP-1]. Participants said that tools do not generally include (i) categories and tags for HCDs; (ii) no feature to associate end-user complaints with code commit history; (iii) the priority and severity of HCDs are inconsistent; (iv) lack of details in the reports such as steps to reproduce a problem and tool support; (v) time-consuming; (vi) task alignment and progression of the bugs lead to delays in fixing the defects. SR-10 claimed, “Defect reports provides limited tagging options on the defects, lack of categories and lineage for different types of defects”. An interview participant stated, “If you leave it to people to explain the bug detail, everyone would explain it differently” - [IP-8].

18.3% of participants consider these tools do not provide particular challenges because it depends on the reporter’s skills in documenting the issue. An interview participant claimed, “The skill of the reporter to accurately describe the issue is critical. The reporter that does not convey the problem is the major issue” - [IP-7]. While some survey respondents suggested: “It comes to the reporter’s skill in accurately capturing the defect. If an end user is unable/incapable of even submitting a complaint then this is a fundamental problem.” - [SR-15]

¹⁴<https://trello.com/>

¹⁵<https://github.com/>

Table 3
Popular defect tracking tools used by the participants

Defect tracking tools used by the participants	Overall
Survey responses	
JIRA and GitHub	24.6%
Trello	21%
Bugzilla	11.4%
Zendesk	9.6%
Service NOW	7.9%
Redmine, Mantis, Remedy, Asana, Backlog, Zoho, BugHerd, etc.	25.5%
Interview responses	
JIRA and GitHub	90%
In-house tool	10%

“Most of the defect tracking tools are matured and have all the bells and whistles to capture the defect details” - [SR-9].

5.5.2. RQ4b Results - Factors that make defect tracking tools user-friendly

We asked our practitioners an open-ended question regarding possible recommendations for improving defect reporting tools to handle HCDs better. They said while creating a new defect in the tool, there should be an option to create more categories, tags, and sub-tasks relating to HCDs. In addition, there should be an alternative to providing feedback from a human perspective. IP-1 claimed that “There is a limitation in Jira to accommodate human-centric issues. The tool does not provide option to include feedback in various contexts such as contexts from a cultural or language perspective are not tracked.”. The defect priority and severity should be adjusted based on the demographics impacted to reduce bias. Furthermore, the report should have the option to add a snapshot or video, a detailed and condensed description, and end-user surveys to make the defect precise for the developers. There should be defect report templates and targeted practices for more accessible learning and training to ensure human-centric requirements are not missed. Finally, they should include testing for edge conditions based on different perspectives to handle multiple scenarios while using a product. For instance, some survey respondents commented:

“By extending the workflow functionality and customisability to business teams, they can track work, manage projects, and stay on top of everything.” - [SR-26].

“By defining identified defects as tags, we can manage, and mitigate future human-centric issues.” - [SR-10].

5.5.3. RQ4c Results - Key fields required in a defect reporting form to manage HCDs better

We asked participants to comment on a generalised defect reporting structure common in defect tracking tools such as Jira, Bugzilla¹⁶, and GitHub. The commonly used fields in defect report forms are shown in Figure 10.

¹⁶<https://www.bugzilla.org/>

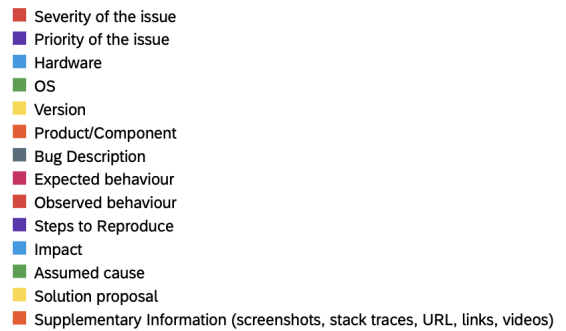


Figure 10: Defect reporting fields

We asked respondents to rate these defect report fields based on their significance to understand and resolve HCDs. We used a Likert scale of 1 to 5, with ‘1’ being the least important and ‘5’ being the most significant. Figure 11 shows the mean plot of responses from developers and reporters based on the importance when creating an HCD. We used weighted arithmetic mean (Cochran, 1977) of developer and reporter responses as follows:

$$x_c = \frac{m \cdot x_a + n \cdot x_b}{m + n} \quad (1)$$

where,

- x_a : the mean of the developer set,
- m : the number of developer responses (34),
- x_b : the mean of the reporter set,
- n : the number reporter responses (28),
- x_c : the weighted arithmetic mean.

The observed behaviour is highly rated, followed by the severity of the issue and bug description. This could be because HCDs are subjective and qualitative defects, and most of the tools provide an option for the observed behaviour, severity, and description, but a detailed report

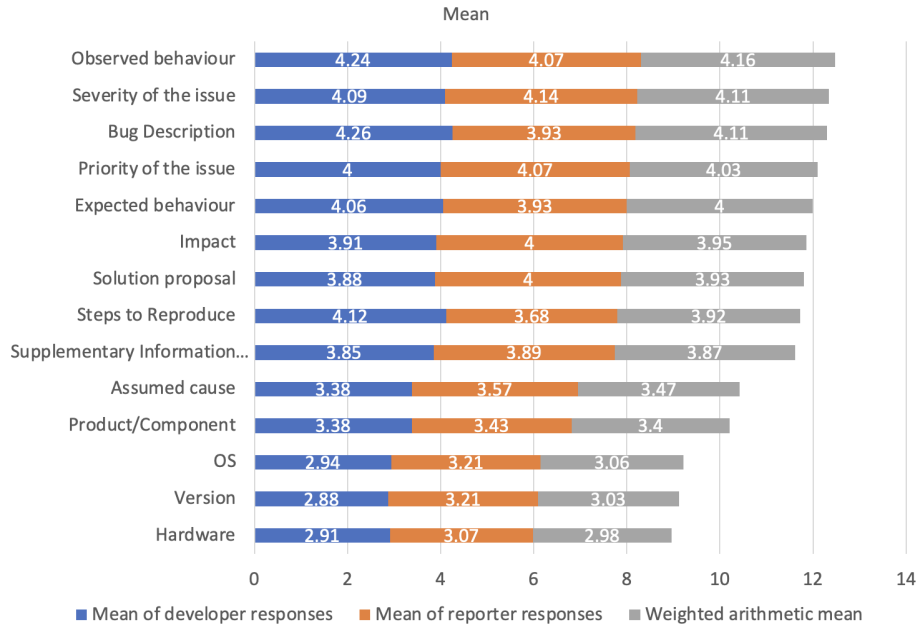


Figure 11: Mean plot of defect reporting fields based on developer and reporter responses (in Surveys)

is fundamental. The least significant fields are Operating System (OS), version and hardware.

5.5.4. RQ4d Results - Influence of experience on HCD defect quality

88% of participants highlighted the role of experience in delivering high-quality HCD reports. Respondents suggested, *“To deliver high-quality product, we need to improve the Return on Investment (ROI) by reducing the cost of development. Overall, the communication, teamwork and connectivity should be better to detect issues earlier and understand defect trends. In short, the better, the service, the better is customer satisfaction which usually comes with experience”* - [SR-26].

“Experience in using a defect reporting tool will influence the quality of a defect report. The deeper you understand a tool and use all its features, you can generate better reports” - [SR-3].

Some respondents claimed that the formulation of a HCD defect report comes with experience and should improve over time. The tools are always secondary; the organisation must provide the necessary support and training to develop employees’ knowledge base. SR-15 claimed, *“A particular defect reporting tool is not much important - what is important is that there is organisational support to use and update them based on feedback”*.

RQ4 Key findings.

- Jira and Trello are commonly used closed-source tools, and GitHub is the commonly used open-source tool.
- Most practitioners highlighted that most current defect reporting tools provide a generic structure focusing on quantitative defects and don’t provide options to handle many HCDs, which are usually more qualitative.
- Practitioners said that defect-tracking tools should provide snapshots or videos, a detailed and condensed description, end-user surveys, templates, and tags, to make the HCD reports more precise for developers.
- In a defect report for an HCD, fields such as observed behaviour, severity and bug description are fundamental. The least significant fields are OS, version, and hardware affected by the issue.
- A majority of practitioners claimed that experience in defect reporting tools could be influential in delivering high-quality defect reports. More experience can lead to better presentation and consistent reports.

5.6. RQ5 Results - Improvements required for more effective reporting and resolution of HCDs

Practitioners said that **more automation** could be beneficial if we can learn from comments, version control histories, guidelines, and existing or previous issues to provide use cases to manage HCDs. The underlying guidelines for successful automation would be to avoid duplicate defects, aggregate different types of HCDs, learn from prior HCD causes and solutions, and improve the HCDs reporting process. For example, some survey respondents suggested:

“Learning and applying experience from past defects means you can fix them faster using the skills you have acquired.”

The same reasoning can be applied to human-centric defects.” - [SR-16]

“I believe there is nothing more valuable than learning from old mistakes, report them, document them and use them in the future for faster work” - [SR-27]. Some interview participants added to this by claiming: *“Automation would be beneficial to a good extent for aggregating all these defects such as scraping web to find similar human-centric issues.” - [IP-5]*

“A recommender engine which could create a knowledge base and taxonomy that underpins human-centric defects would be very useful.” - [IP-2].

The provision of an HCD-oriented **template or checklist while reporting issues** would help. A checklist on what characteristics will be beneficial for a developer to understand, appreciate and fix the HCD can help a reporter effectively capture information, speed up the process, and reduce delays. For example, an interview participant recommended, *“In the case of the human-centric defects, a template should be there to guide the reporter on how to make a complete report” - [IP-7].*

Some practitioners claimed that **more awareness** regarding HCDs is needed. One reason could be lack of **guidelines or taxonomy to understand HCDs**, similar to issues we found in reporting usability defects in prior work (Yusop et al., 2016a, 2020b). If there were such a comprehensive list to help report, evaluate, and fix HCDs, then the management of HCDs in the SE domain could be better. For instance, IP-5 mentioned, *“A set of standards for human-centric defects could benefit all developers. A list of things related to defects which developers can refer to and see what is applicable to them rather than waiting for the customer to provide them feedback and repeat the cycle”.* The taxonomy could categorise HCDs and test conditions based on culture, language, gender, age groups, educational background, and physical or mental disabilities. For instance, IP-7 stated, *“The standards should contain human factors such as from colourblind perspective, make sure that the UI is interpretable”.* A broader taxonomy or a standard such as Design for All practices, Development practices, and W3C Accessibility Standards¹⁷, could be useful.

RQ5 Key findings.

To improve HCDs management, some suggestions are:

- increased HCDs detection/classification automation;
- a standard taxonomy of HCDs for practitioners;
- improved HCD information capturing support in defect tracking tools.

6. Discussion and Recommendations

6.1. Key findings

Nature of human-centric defects. Khalajzadeh et al. (2022) describe human-centric issues as the issues end-users may encounter when utilising mobile applications due to their

unique human characteristics and limitations. HCDs occur across SE domain in mobile and web applications as subjective and nuanced defects that may arise due to different human perspectives or are formed inadvertently during the development and testing of the software. Technical defects are more definitive and occur for all users. It is important to actively include HCDs in current defect management strategies. This will help to prioritise and draw attention to the issues experienced by a diverse range of individuals who report defects. There should be more awareness regarding HCDs to create a comprehensive working model to report, manage and solve such issues.

Ramifications of software practitioners’ human aspects in handling of HCDs. HCDs are influenced by various factors related to end users, such as their personality traits, cultural or ethnic background, experience, emotions, education, gender, age, and language proficiency. Studies such as (Grundy et al., 2018; Grundy, 2020; Grundy et al., 2020b) emphasise the difficulties arising from the inability to identify diverse end-users while addressing human-centric issues. Developers themselves possess a wide range of human characteristics, which often differ greatly from those of their end-users. Therefore, the identification and management of HCDs is frequently based on personal judgment and interpretation. With the help of experience, there is a knowledge and learning curve that helps in creating the best possible mitigation. The utilisation of experience in managing HCDs can lead to an improved process for identifying, reporting, and resolving human factors in the Software Engineering (SE) field. Incorrect or unclear communication might result in the mishandling of HCDs. Hence, effective management of HCDs relies heavily on communication, experience, and educational background. The incorporation of diversity in organizations and teams can result in a variety of viewpoints and experiences, ultimately enhancing the design, development, and testing of solutions. For instance, if we allocate the problem to a developer who possesses distinct human characteristics compared to another, it can be analysed from an entirely new standpoint and contribute to a more effective resolution of the issue. Having a comprehensive understanding of the human aspects and their influence on the SDLC will greatly assist in effectively managing HCDs and developing improved software solutions.

Impact of team and organisational context. An organisation’s size, team climate, and geographic location influence the identification and resolution of HCDs. Organisation size impact depends on how large and smaller organisations handle defects. There are positives and negatives of how the defects are managed in large and small organisations due to resources, budgets, management and bureaucracy. These aspects impact the management of HCDs significantly. The reporting and resolution of HCDs can be improved if teams and organisations actively embrace inclusivity and diversity. **Team philosophy, diversity, and culture have an impact on the handling of HCDs in a product.** If the team culture fosters a process that prioritises the needs and experiences of individuals, then any human-centred design issues can be

¹⁷<https://www.w3.org/WAI/standards-guidelines/>

discovered and resolved effectively. (Hidellaarachchi et al., 2021) stated cultural diversity can lead to communication issues and eventually lead to a lack of awareness problems. Translation challenges may arise when dealing with defects due to variations in culture and environment across different geographic regions. The influence of indigenous colloquialisms or dialects, work methodologies, cultural conventions, and notions of hierarchy can significantly impact several facets of SE tasks and the anticipated user requirements for software. These factors may hinder the ability to fully understand and resolve an HCD, perhaps prolonging the process. Thanks to globalization and the widespread use of video conferencing, agile methods, and defect-tracking tools in global software development work, the influence of geographical differences on HCD resolution has been reduced.

Current defect tracking tools need enhancement for better HCDs management. Defect reporting and tracking tools significantly impact the process of reporting and resolving HCDs. For instance, if the quality of the HCD report is poor, the resolution of the reported issue will be delayed or perhaps disregarded. Proficiency in utilising defect management tools is primarily acquired through hands-on experience and practical use. However, awareness regarding the characteristics of the tools for reporting and managing HCDs is currently limited in many organisations and teams leading to substandard HCD reports.

Due to a lack of understanding about the key aspects of different HCDs that are most critical to capture, and limited templates containing key HCDs characteristics to capture in current defect reporting tools, developers are not able to adequately capture and track HCDs. This is supported by related prior work e.g. usability defect capture, where a wizard-based approach for reporting usability defects, especially for novice users, was proposed Yusop et al. (2016a). Specific HCD-oriented templates or checklists while reporting issues should be investigated. However, with HCDs a checklist would need more comprehensive scoping from different human aspects to common end-user problems. Overall, a checklist will comprise what characteristics will be beneficial for a developer to understand, appreciate and fix the HCD. It may help a reporter in effectively capturing information as well as speeding up the process and reducing delays. Hence, a potential research avenue to capture and manage HCDs effectively (Section 5.6).

Other reporting tools features, such as complexity, poor searching capability, lack of integrated tools, and lack of attachment of diverse file types, would better support HCD management. Many defect reporting tools additionally have been shown to have accessibility difficulties for users, a too-general structure of defect reports, primarily focus on handling quantitative issues, and provide a lack of assistance for managing a variety of HCDs Huynh et al. (2021).

Defect reporting and tracking tools alone are not sufficient for effective management of HCDs. Huynh et al. (2021) provided some insights on preliminary ideas for improved human-centric defect reporting. They captured a subset of

humans' underlying conditions via personas such as vision impairment, colour blindness, hearing impairment, dyslexia, dexterity and aphasia. They developed initial prototypes to support human-centric defect reporting based on the personas. While this work is commendable, this work did not discuss the practitioners' perspective on reporting. According to SE practitioners, the proficiency of reporters is crucial in efficiently reporting and fixing an HCD. Proficiency in reporting is acquired through practical experience. Based on the feedback we have received, it is evident that gaining more experience is crucial for detecting and resolving HCDs. Onboarding a new employee requires a significant investment of time, energy, and financial resources, which many organisations are unwilling or unable to allocate. Therefore, a deeper understanding of the process and technologies used for tracking HCDs is necessary in order to produce high-quality reports. In addition, defect reporting and fixing must prioritise various human factors related to their end-users and comprehend how these factors contribute to HCDs in their software.

Determining more suitable training of reporters and software developers in identifying, documenting, triaging, fixing and updating end users on HCD-based updates needs to be undertaken Yusop et al. (2020b). This could include learning diverse end user needs, HCD taxonomy features, HCD-supporting tools, and diagnosing diverse HCD causes. Training in mechanisms for better feedback from end users, inclusion of end users in diagnosing and fixing processes, and consultation with target end users on suitable fixes need to be investigated.

There is a lack of HCD taxonomy. One possible explanation for the limited awareness about HCDs among many reporters and developers is the absence of sufficient tools and resources to support their understanding. This aligns with previous studies into usability defect reporting Yusop et al. (2016a, 2020b), that found that many usability defects suffer from an inconsistent taxonomy. They recommended an improved usability defect taxonomy that can better capture meanings and names of usability defect attributes to improve reporting, triaging, understanding, and fixing of these defects. Developers are unable to effectively capture and track HCDs because there is no consensus on which components of HCDs are most important to include, resulting in a lack of comprehensive templates in current defect-reporting solutions. Since our definition of HCDs are more comprehensive and encompass a wider range of human-centric defect categories, an area of focus for additional research and practical utilisation should be the development of a taxonomy for HCDs. This could include various categories, edge situations, severity levels, and the necessary content for defect reports. This would aid developers with better identification and prioritisation of HCDs, improve defect report structure to accommodate key HCD information, and provide triage and test conditions based on culture, language, gender, age groups, educational background, and physical or mental disabilities (Section 5.6).

There is a lack of automation for reporting, managing and fixing HCDs. The process of identifying and reporting HCDs is still highly manual. Introducing intuitive mechanisms to improve the process, detection, reporting, and fixing could be beneficial. For instance, in the case of a reported HCD, automation can be employed to identify the precise category of HCDs, prioritise multiple HCD reports related to the same issue, and establish a defect resolution procedure tailored to that particular HCD. (Plein and Bissyandé, 2023) provides the usage of ChatGPT¹⁸ in understanding and reproducing the bug reports. Their research showcases that ChatGPT automatically addressed half of the reported bug reports. This inclusion can extremely help in reporting and fixing of HCDs. Hence, more research needs to be conducted in this area.

6.2. Recommendations for Practice

We highlight the limitations of existing industry practices and propose key recommendations to capture and manage HCDs effectively, based on our study findings.

There needs to be more awareness regarding HCDs to create a comprehensive working model to report, manage and solve such issues. Due to the subjectivity of HCDs, different perceptions make it difficult to capture and understand HCDs (Huynh et al., 2021). Every individual will perceive applications differently, and the usage criteria vary resulting in defects. Hence, awareness and understanding of HCDs are important in the current SE domain.

Create diverse work environments: Organisations and teams can create diverse, inclusive, and experienced environments to manage and promote understanding regarding HCDs efficiently. HCDs should be actively considered as a part of the current defect management process to emphasise and highlight the problems diverse reporters face (Sections 5.3, 5.4).

Provide human-centric issue training: Culture and sensitivity training should be provided to all practitioners to understand different human aspects and their impact on defect reporting and fixing. Yusop et al. (2016c) stated that usability-related training to respondents leads to a better understanding of usability defects. This study emphasises standardised training methods for all HCDs, rather than focusing solely on usability problems. Practitioners should be provided with templates and targeted practices for more accessible learning and training to ensure human-centric requirements are not missed (Sections 5.3, 5.4).

Utilise more customer-centric processes: Such processes should be created by teams to help impact end-users with a feedback loop to improve the existing processes. Co-creational spaces to enable software teams to capture critical human-centric requirements related to their software have been suggested by Grundy et al. (2020b). In addition to co-creational spaces, organisations and teams can create a change management process based on end-users feedback, which can help highlight HCDs efficiently (Section 5.4).

Consider more diverse users: Diverse user needs should be considered with solutions formulated around their needs. Identification practices such as trial and error, categorisation of issues, user's expected behaviour, accessibility issues, a recreation of user's approach and testing for edge conditions should be undertaken by teams. Yusop et al. (2016a) provided key recommendations around usability defect attributes to capture attributes at a fine-grained level to support defect correction, which could be established for other HCDs as well. Teams need to cultivate an approach for focusing on HCDs in their development and testing practices (Section 5.2).

Create HCD-specific report templates: HCD report structure needs to be improved by accommodating different categories, tags and information for different types of HCDs. The defect priority and severity should be adjusted based on the demographics impacted to reduce bias. Yusop et al. (2016a) recommends a better severity, prioritisation, and custom reporting form to accommodate different types of usability defects. However, with regard to HCDs, different categories of HCDs need to be captured, reported and diagnosed to understand HCDs appropriately (Section 5.5.1).

Include diverse user information in HCDs: Customer information and feedback should be integrated into the HCDs tracking process. Strate and Laplante (2013) suggests that defect report quality suffers from a disconnect between user reports and what the developers need in that report. It is also true for HCDs since HCDs could deviate depending on end-users and the report quality will also be dependent as not all users are proficient enough to provide quality reports. However, defect-tracking tools should provide basic features such as snapshots, videos, detailed descriptions, end-user surveys, templates, tags, and categories, to make the HCD reports more precise for the developer (Sections 5.5.2, 5.5.3, 5.5.4).

Better HCD report automation support: Identifying and reporting HCDs is still highly manual. Yusop et al. (2016c,a) suggests engaging automation in usability defect reporting tools could be more effective in capturing usability defects. Exploration of automated tools by introducing intuitive mechanisms to improve the process, detection, reporting, and fixing of HCDs could also be beneficial. For instance, if an HCD is reported, there could be automation to recognise the specific category of HCDs, triage multiple HCD reports for the same problem, and create a defect resolution process specific to that HCD. Similarly, analysing HCDs in relation to other defects can also lead to better identification and familiarity for the developers (Section 5.6).

Utilise HCD domain experts: Understanding and fixing HCDs can be better managed if a defect is assigned to domain experts with prior experience in HCDs. HCDs are subjective, and the understanding of defects can depend on the developer's empathy and understanding of human aspects leading to HCDs. This can be improved by organisations and teams employing diverse environments. Strate and Laplante (2013) suggested automatic defect assignment

¹⁸<https://chat.openai.com>

based on the developer's background and commit history to reduce the time to fix the issue (Section 5.2).

7. Limitations

7.1. Threats to Validity

Our strategies for addressing validity threats and ensuring the robustness of our findings are informed by the detailed discussion of validity threats in empirical software research provided by (Wohlin and Aurum, 2015; Cruzes and ben Othmane, 2017; Stol and Fitzgerald, 2020; Felderer and Travassos, 2020).

Whether participants clearly comprehended the idea of HCDs in the same way as we had intended poses the primary threat to the internal validity of our study (Cruzes and ben Othmane, 2017; Wohlin and Aurum, 2015). Our results could have been deemed less trustworthy if practitioners had not comprehended the topics or used other meanings, even though the practitioners have a working knowledge of defect management in the SE domain. To mitigate this threat and reduce participants' misunderstandings, we gave clear descriptions and examples of each topic in text descriptions in the survey and verbally in interviews. We consider this threat to be minimal because we did not spot any misunderstandings or contradictions in the open-ended survey questions provided or in answers during interview analysis. Further, providing payment or gift cards to the participants in the second round of the survey and for interviews could be another threat. We used two distinct crowdsourcing platforms, AMT and Prolific, after considerations by similar studies (Jiarapakdee et al., 2021). However, after reviewing their responses, we authorised the payments for the respondents to determine if they answered each question and matched our target participant group.

Potential misinterpretation of study questions is the main threat to construct validity. We developed our survey and interview questions with careful consideration. We conducted pilot studies with software industry professionals and research professionals with previous software development and testing experience to provide clarity to the questions. We asked our pilot study participants to explain their understanding of each concept and modified the instruments to improve clarity. Additionally, we provided practitioners with sufficient examples without creating bias so the responses could be based on practitioners' own judgment.

The limited survey size impacts generalisability. We received 104 responses, out of which 50 responses were selected. To mitigate the sample size of the survey Stol and Fitzgerald (2020), we conducted ten in-depth parallel interviews to understand the concepts of HCDs in SE. A community as diverse as software practitioners' cannot be quantified in terms of size. Therefore, it is probable that the sample used in our analysis does not accurately reflect the population as a whole. To mitigate this threat, we recruited participants from different countries and domains. Additionally, different organisations have different

interpretations of the job roles of an individual. To reduce this threat, we interviewed developers and reporters with varied skills and experience. We gauged the skills of our respondents by rating their involvement in SE activities such as development, testing, database, usability, administration, and security.

7.2. Study Findings Limitations

Our study is naturally limited to self-reported feedback during our survey and interview participation. However, we believe our findings can be transferable to other organisations within the same country and in other countries within the same geographical zone where cultural contexts are similar (Ghaisas et al., 2013; Wieringa and Daneva, 2015). For instance, the lack of awareness regarding human-centric aspects was consistently observed across different organisational settings, suggesting that this issue may also be present in other environments. Additionally, recommendations for improved end-user feedback processes, a more descriptive taxonomy, and suitable automation are universally applicable across various industries and organisational contexts. This is supported by the similarities observed in previous studies by (Mockus, 2010; Yusop et al., 2020a). Nevertheless, we recognize that different organisational contexts, software domains and cultural differences might affect the wider applicability of our results. In future studies, we plan to include end-users as HCD-reporting participants from different cultural contexts to capture a broader range of HCDs. We focused primarily on reporting HCDs in this study and not triaging, diagnosing and fixing reported HCDs. Future studies will explore how HCDs are handled by development teams to correct them. We asked reporters for their feedback on HCDs and HCD reporting in this study. In future we want to analyse existing HCD-related defect reports in defect reporting tools for their structure, content and handling to triangulate with our survey and interview findings.

8. Conclusions and Future Work

HCDs occur in response to user perceptions or form inadvertently during the software development, therefore making it challenging to identify the problem and correct it before the products hit the market. Currently, development teams work on products based on a set of specified client requirements such as 'what a system should do', i.e. functional, and 'how it should do it', i.e. non-functional (Pandey and Pandey, 2012; Siddiqi and Shekaran, 1996). Unfortunately, the requirements often ignore the characteristics related to human-centric aspects such as personality, gender, emotions, language, engagement, ethnicity and culture, age, socioeconomic background, physical and mental challenges (Grundy, 2020; Hidellaarachchi et al., 2021). We surveyed and interviewed software practitioners to explore the concept of human-centric aspects in defect reporting and resolution in the SE domain. We received fifty responses for a survey and conducted ten interviews to understand the perception of

HCDs in SE. Based on the responses, we identified multiple areas of focus such as HCDs, identification and assignment of defects, experience role in defects, impact of human aspects on HCDs, organisation and teams role in HCDs, and management of HCDs in defect tracking tools.

Our study offers a detailed overview of how HCDs are perceived by software developers and reporters throughout the SDLC. Based on our findings, HCDs are viewed distinctly from technical defects and highlight the importance of considering human-centric aspects during the development, testing, and maintenance phases. We explored the current state of defect reporting tools, identifying significant challenges and necessary enhancements to better support HCDs. Practitioners reported that existing tools lack the capabilities to effectively manage HCDs, suggesting the need for improved automation, taxonomy, and checklists tailored to HCDs, which can lead to more effective human-centric SE practices. Additionally, highlighting the importance of diverse development teams and the role of experience in managing HCDs encourages more inclusive and comprehensive approaches within the SE domain.

The field of defect tracking is large, and handling human-centric aspects is still a major challenge. There are multiple avenues to explore, such as better discovery and triaging of HCDs, classification and categorisation of HCDs, and remediation and verification of HCDs. Based on the responses we gathered, our next direction is to explore existing detailed reporting processes and tools of HCDs. We want to study the challenges faced by diverse users while reporting their issues. Once we identify their challenges, we want to work on improving the reporting process and provide suggestions to enhance existing defect-tracking tools for better management and resolution of HCDs. This information will require close interaction with diverse users and ultimately can lead to best practices in reporting HCDs in the SE domain. We further plan to corroborate our findings from this study by analysing defect reporting and fixing in open-source repositories.

Acknowledgment. Chauhan and Grundy are supported by ARC Laureate Fellowship FL190100035.

Declarations

Conflict of Interest. The authors declare that they have no conflict of interest.

Data Availability Statement. The data is provided online.

References

Ahmad, K., Abdelrazek, M., Arora, C., Bano, M., Grundy, J., 2023. Requirements practices and gaps when engineering human-centered artificial intelligence systems. *Applied Soft Computing* .

Anderson, N.R., West, M.A., 1998. Measuring climate for work group innovation: development and validation of the team climate inventory. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 19.

Arora, C., Sabetzadeh, M., Briand, L.C., 2019. An empirical study on the potential usefulness of domain models for completeness checking of requirements. *Empirical Software Engineering* 24.

Baltes, S., Ralph, P., 2022. Sampling in software engineering research: A critical review and guidelines. *Empirical Software Engineering* 27.

Bass, J.M., 2015. How product owner teams scale agile methods to large distributed enterprises. *Empirical software engineering* 20.

Bettenburg, N., Just, S., Schröter, A., Weiss, C., Premraj, R., Zimmermann, T., 2008. What makes a good bug report?, in: *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering*.

Bigham, J.P., Brudvik, J.T., Zhang, B., 2010. Accessibility by demonstration: enabling end users to guide developers to web accessibility solutions, in: *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3.

Burnett, M., Stumpf, S., Macbeth, J., Makri, S., Beckwith, L., Kwan, I., Peters, A., Jernigan, W., 2016. Gendermag: A method for evaluating software's gender inclusiveness. *Interacting with Computers* 28.

Cochran, W.G., 1977. *Sampling techniques*. John Wiley & Sons.

Couper, M.P., 2000. Web surveys: A review of issues and approaches. *The Public Opinion Quarterly* 64.

Cruz, S., da Silva, F.Q., Capretz, L.F., 2015. Forty years of research on personality in software engineering: A mapping study. *Computers in Human Behavior* 46.

Cruzes, D.S., Dyba, T., 2011. Recommended steps for thematic synthesis in software engineering, in: *2011 international symposium on empirical software engineering and measurement, IEE*.

Cruzes, D.S., ben Othmane, L., 2017. Threats to validity in empirical software security research, in: *Empirical research for software security*. CRC Press.

Curumsing, M.K., Fernando, N., Abdelrazek, M., Vasa, R., Mouzakis, K., Grundy, J., 2019. Emotion-oriented requirements engineering: A case study in developing a smart home system for the elderly. *Journal of systems and software* 147.

Davies, S., Roper, M., 2014. What's in a bug report?, in: *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*.

Fazzini, M., Khalajzadeh, H., Haggag, O., Li, Z., Obie, H., Arora, C., Husain, W., Grundy, J., 2022. Characterizing human aspects in reviews of covid-19 apps, in: *9th IEEE/ACM International Conference on Mobile Software Engineering and Systems*.

Feiner, J., Andrews, K., Krainz, E., Aigner, W., Schmiedl, G., Blumenstein, K., Zeppelzauer, M., Iber, M., 2016. Convenient mobile usability reporting with useapp., in: *FMT*.

Felderer, M., Travassos, G.H., 2020. The evolution of empirical methods in software engineering, in: *Contemporary Empirical Methods in Software Engineering*. Springer.

Florea, R., Stray, V., 2019. The skills that employers look for in software testers. *Software Quality Journal* 27.

Frijda, N.H., et al., 1986. *The emotions*. Cambridge University Press.

Ghaisas, S., Rose, P., Daneva, M., Sikkil, K., Wieringa, R.J., 2013. Generalizing by similarity: Lessons learnt from industrial case studies, in: *2013 1st International Workshop on Conducting Empirical Studies in Industry (CESI)*, IEEE.

Glinz, M., 2007. On non-functional requirements, in: *15th IEEE international requirements engineering conference (RE 2007)*, IEEE.

Grundy, J., 2020. Human-centric software engineering for next generation cloud-and edge-based smart living applications, in: *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, IEEE.

Grundy, J., Abdelrazek, M., Curumsing, M.K., 2018. Vision: Improved development of mobile ehealth applications, in: *2018 IEEE/ACM 5th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, IEEE.

Grundy, J., Khalajzadeh, H., McIntosh, J., 2020a. Towards human-centric model-driven software engineering., in: *ENASE*.

Grundy, J., Khalajzadeh, H., McIntosh, J., Kanij, T., Mueller, I., 2020b. Humanise: Approaches to achieve more human-centric software engineering, in: *International Conference on Evaluation of Novel Approaches to*

- Software Engineering, Springer.
- Grundy, J., Mueller, I., Madugalla, A., Khalajzadeh, H., Obie, H.O., McIntosh, J., Kanij, T., 2021. Addressing the influence of end user human aspects on software engineering, in: International Conference on Evaluation of Novel Approaches to Software Engineering, Springer.
- Gunatilake, H., Grundy, J., Mueller, I., Hoda, R., 2023. Empathy models and software engineering—a preliminary analysis and taxonomy. *Journal of Systems and Software* 203.
- Haggag, O., Grundy, J., Abdelrazek, M., Haggag, S., 2022. A large scale analysis of mhealth app user reviews. *Empirical Software Engineering* 27.
- Hidellaarachchi, D., Grundy, J., Hoda, R., Madampe, K., 2021. The effects of human aspects on the requirements engineering process: A systematic literature review. *IEEE Transactions on Software Engineering* .
- HL, D., Dreyfus, S., 1986. *Mind over machine: the power of human intuition and expertise in the era of the computer*. New York Free Pr .
- Hornbæk, K., Frøkjær, E., 2006. What kinds of usability-problem description are useful to developers?, in: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA.
- Huynh, K., Benarivo, J., Da Xuan, C., Sharma, G.G., Kang, J., Madugalla, A., Grundy, J., 2021. Improving human-centric software defect evaluation, reporting, and fixing, in: *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, IEEE.
- James, W., 1922. *The emotions*. Williams & Wilkins Co.
- Jiarpakdee, J., Tantithamthavorn, C.K., Grundy, J., 2021. Practitioners' perceptions of the goals and visual explanations of defect prediction models, in: *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*, IEEE.
- Keenan, S.L., Hartson, H.R., Kafura, D.G., Schulman, R.S., 1999. The usability problem taxonomy: A framework for classification and analysis. *Empirical Software Engineering* 4.
- Khalajzadeh, H., Shahin, M., Obie, H.O., Agrawal, P., Grundy, J., 2022. Supporting developers in addressing human-centric issues in mobile apps. *IEEE Transactions on Software Engineering* .
- Khandkar, S.H., 2009. *Open coding*. University of Calgary 23.
- Kitchenham, B.A., Pfleeger, S.L., 2002. Principles of survey research part 2: designing a survey. *ACM SIGSOFT Software Engineering Notes* 27.
- Kitchenham, B.A., Pfleeger, S.L., Pickard, L.M., Jones, P.W., Hoaglin, D.C., El Emam, K., Rosenberg, J., 2002. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on software engineering* 28.
- Ko, A.J., Chilana, P.K., 2010. How power users help and hinder open bug reporting, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- Lee, J.H., Kim, Y.M., Rhiu, I., Yun, M.H., 2021. A persona-based approach for identifying accessibility issues in elderly and disabled users' interaction with home appliances. *Applied Sciences* 11.
- Mbipom, G., Harper, S., 2011. The interplay between web aesthetics and accessibility, in: *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*.
- Mendez, C., Letaw, L., Burnett, M., Stumpf, S., Sarma, A., Hilderbrand, C., 2019. From gendermag to inclusivemag: An inclusive design meta-method, in: *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, IEEE.
- Mockus, A., 2010. Organizational volatility and its effects on software defects, in: *Proceedings of the eighteenth ACM SIGSOFT international symposium on Foundations of software engineering*.
- Moreno, L., Alarcon, R., Martínez, P., 2020. Easier system. language resources for cognitive accessibility., in: *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*.
- Müller, S.C., Fritz, T., 2015. Stuck and frustrated or in flow and happy: Sensing developers' emotions and progress, in: *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, IEEE.
- Pandey, D., Pandey, V., 2012. Importance of requirement management: A requirement engineering concern. *International Journal of Research and Development-A Management Review (IJRDMR)* 1.
- Plein, L., Bissyandé, T.F., 2023. Can llms demystify bug reports? *arXiv preprint arXiv:2310.06310* .
- Prikladnicki, R., Dittrich, Y., Sharp, H., De Souza, C., Cataldo, M., Hoda, R., 2013. Cooperative and human aspects of software engineering: Chase 2013. *ACM SIGSOFT Software Engineering Notes* 38.
- Puzis, Y., 2012. Accessible web automation interface: A user study, in: *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, Association for Computing Machinery. p. 291–292.
- Shamsujjoha, M., Grundy, J., Khalajzadeh, H., Lu, Q., Li, L., 2024. Developer and end-user perspectives on addressing human aspects in mobile health apps. *Information and Software Technology* 166.
- Siddiqi, J., Shekaran, M.C., 1996. Requirements engineering: The emerging wisdom. *Ieee Software* 13.
- Soomro, A.B., Salleh, N., Mendes, E., Grundy, J., Burch, G., Nordin, A., 2016. The effect of software engineers' personality traits on team climate and performance: A systematic literature review. *Information and Software Technology* 73.
- Stol, K.J., Fitzgerald, B., 2020. Guidelines for conducting software engineering research, in: *Contemporary Empirical Methods in Software Engineering*. Springer.
- Strate, J.D., Laplante, P.A., 2013. A literature review of research in software defect reporting. *IEEE Transactions on Reliability* 62.
- Vedant, Chetan, Hourieh, John, 2022. How do software developers and reporters currently perceive human-centric defects? URL: <https://doi.org/10.5281/zenodo.7978481>, doi:10.5281/zenodo.7978481.
- Votipka, D., Stevens, R., Redmiles, E., Hu, J., Mazurek, M., 2018. Hackers vs. testers: A comparison of software vulnerability discovery processes, in: *2018 IEEE Symposium on Security and Privacy (SP)*, IEEE.
- Wieringa, R., Daneva, M., 2015. Six strategies for generalizing software engineering theories. *Science of computer programming* 101.
- Winograd, T., Woods, D., 1997. *The challenge of human-centered design. Human-centered systems: information, interactivity, and intelligence* .
- Wohlin, C., Aurum, A., 2015. Towards a decision-making structure for selecting a research design in empirical software engineering. *Empirical Software Engineering* 20.
- Yusop, N.S.M., Grundy, J., Schneider, J.G., Vasa, R., 2020a. How usability defects defer from non-usability defects?: A case study on open source projects .
- Yusop, N.S.M., Grundy, J., Schneider, J.G., Vasa, R., 2020b. A revised open source usability defect classification taxonomy. *Information and software technology* 128.
- Yusop, N.S.M., Grundy, J., Vasa, R., 2015. Reporting usability defects: limitations of open source defect repositories and suggestions for improvement, in: *Proceedings of the ASWEC 2015 24th Australasian Software Engineering Conference*.
- Yusop, N.S.M., Grundy, J., Vasa, R., 2016a. Reporting usability defects: a systematic literature review. *IEEE Transactions on Software Engineering* 43.
- Yusop, N.S.M., Grundy, J., Vasa, R., 2016b. Reporting usability defects: do reporters report what software developers need?, in: *Proceedings of the 20th international conference on evaluation and assessment in software engineering*.
- Yusop, N.S.M., Schneider, J.G., Grundy, J., Vasa, R., 2016c. What influences usability defect reporting?—a survey of software development practitioners, in: *2016 23rd Asia-Pacific Software Engineering Conference (APSEC)*, IEEE.
- Yusop, N.S.M., Schneider, J.G., Grundy, J., Vasa, R., 2017. Analysis of the textual content of mined open source usability defect reports, in: *2017 24th Asia-Pacific Software Engineering Conference (APSEC)*, IEEE.
- Zhang, J., Liu, S., Wang, X., Li, T., Liu, Y., 2023. Learning to locate and describe vulnerabilities, in: *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, IEEE.
- Zimmermann, T., Premraj, R., Bettenburg, N., Just, S., Schroter, A., Weiss, C., 2010. What makes a good bug report? *IEEE Transactions on Software Engineering* 36.

A. Appendix

The survey and interview questionnaires can be found on the Zenodo platform at <https://doi.org/10.5281/zenodo.7978481>.