Impact of UI Design on Cybersecurity Incident Reporting: A Form vs. Chatbot Comparison

Joyel Porunnedath Biju, Joshua Porunnedath Biju
Schmalkalden University of Applied Sciences
Schmalkalden, Germany

1 INTRODUCTION

Cybersecurity incidents pose a growing challenge for organizations, making efficient incident reporting systems essential for documenting threats and ensuring timely responses. A well-structured reporting process helps security teams assess risks, track vulnerabilities, and take corrective action, ultimately strengthening an organization's defense against cyber threats.

Employees play a crucial role in this process by reporting security incidents through dedicated user interfaces (UIs). However, the usability of these interfaces significantly impacts reporting efficiency. A poorly designed UI can lead to delays, incomplete reports, and user frustration, whereas an intuitive interface can streamline the process, reducing errors and improving response times. Selecting the right interface design is key to ensuring that users can report incidents quickly and accurately.

This study explores two distinct UI approaches for incident reporting, implemented in a web application designed for laptops and desktops. Variant A, a traditional multi-step form, requires users to manually enter all details without system assistance, while Variant B, a chatbot-based UI, presents a guided conversational approach with predefined options, reducing manual input effort and improving navigation. The chatbot structure provides a single scrolling view, making it easier for users to review and complete the report seamlessly.

The study compares both variants to assess their usability and effectiveness, examining how interface design influences user interaction. Findings reveal that Variant B provided a more structured and intuitive experience, allowing users to complete reports with less effort and fewer interruptions. The guided responses in Variant B reduced complexity, making the process smoother and potentially more accessible to users unfamiliar with security reporting.

This research is conducted in a controlled setting, where the web application is not deployed, focusing solely on interface evaluation. While real-world factors such as long-term user adaptation and organizational security policies are beyond the study's scope, the findings offer valuable insights into designing more user-friendly incident reporting systems.

The report is structured as follows: Section 2 reviews existing UI guidelines and related research. Section 3 presents the prototypes, collected data, and analysis. Section 4 discusses key findings, limitations, and future improvements.

2 RELATED WORKS

Incident reporting is an important part of cybersecurity, requiring interfaces that balance usability, completeness, and engagement. Research has explored two main approaches: step-based progress indicators in traditional form-based reporting and chatbot-guided interactions.

Busetti and Scanni [1] state that incident reporting systems help identify threats, respond to attacks, and improve cybersecurity policies. They found that when reporting is simple and easy to understand, more users are likely to participate. However, complex forms, poor communication, and fear of reporting reduce engagement. To address these challenges, they suggest using interactive feedback and adaptive interfaces to make the reporting process more efficient.

Building on this, Conrad et al. [2] emphasize the role of progress indicators in form-based interfaces. Their research shows that users are more likely to complete tasks when they can track their progress. Material Design guidelines reinforce this by integrating UI elements such as steppers, tabs, and progress bars [5], which simplify multi-step processes. Variant A follows this principle by providing a structured, step-by-step form-filling experience.

An alternative approach to structured forms is the use of conversational interfaces. Siddique [9] discusses the IBM Carbon Design System, which provides flexible UI components that adjust dynamically based on user needs. While not originally designed for chatbot interactions, its modular design allows chatbot elements to be incorporated into reporting systems. The IBM Carbon framework [7] includes UI elements such as chat bubbles, message threads, and interactive text fields, making chatbot-based reporting (as in Variant B) more intuitive and user-friendly.

Additionally, Song et al. [10] introduce BURT, a chatbot that guides users through the bug reporting process by asking structured questions and providing real-time feedback. This approach reduces cognitive load and improves accuracy. Similarly, Yu et al. [11] found that real-time progress indicators improve engagement, reinforcing the effectiveness of structured feedback in step-based reporting. While chatbots improve efficiency, step-based interfaces help maintain clarity and completeness in reporting.

Elson et al. [3] compare the effectiveness of web forms and chatbots in reporting suspicious activity. Their study found that chatbots improve accuracy due to their interactive nature. However, they did not find significant differences in user trust, anonymity, or perceived usability between chatbots and web forms. This confirms that while chatbots like Variant B offer structured reporting, they may not necessarily increase user confidence in cybersecurity reporting.

Similarly, Hristova [6] examines chatbot-based and form-based reporting in medical device incident tracking. The study found no major difference in information retention between the two approaches. However, chatbots were associated with higher reporting rates and improved user engagement. This suggests that Variant B may help users report incidents more quickly and with fewer errors while maintaining completeness.

While these studies provide insights into form-based and chatbot-based reporting, little research has directly compared their impact on user confidence, error rates, and efficiency in cybersecurity incident reporting. This study aims to address this gap by empirically evaluating the strengths and limitations of both approaches.

To further explore these approaches, the next section presents our study's methodology, including prototype development and hypothesis testing.

3 PROTOTYPE

Following are the designed prototypes and evaluation. Section 3.1 presents the design of Variant A (form-based UI) and Variant B (chatbot-based UI). Section 3.2 outlines the research approach, followed by users' recorded performance data in Section 3.3. The analysis of the collected data is then presented in Section 3.4.

3.1 Design

For this study, we developed two Variants for incident reporting: a form-based UI (Variant A) and a chatbot-based UI (Variant B). Both Variants ensure that users complete all required fields before submitting the report but differ in how they interact with the system.

Development of Variant A: Form-Based UI Variant A follows a traditional form layout, where users manually enter details into fields step by step. A progress indicator helps users track completion, but typing everything manually may increase effort and errors.

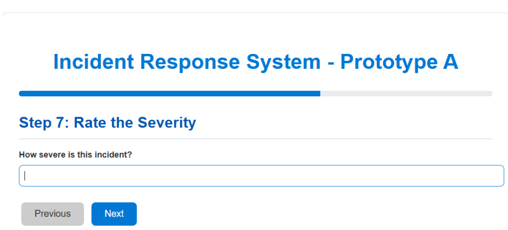

Figure 1: Variant A: Form-Based UI for Incident Reporting

Development of Variant B: Chatbot-Based UI Variant B was designed as an interactive chatbot that guides users through the reporting process. Instead of typing all details manually, users select predefined options or enter short responses, making it easier to complete the report.

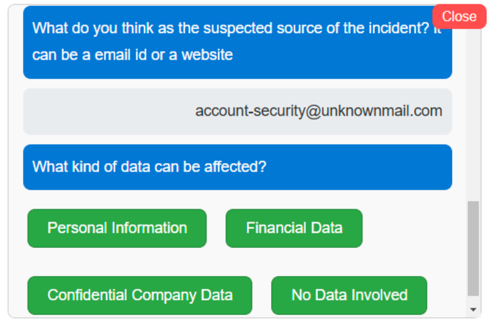

Figure 2: Variant B: Chatbot-Based UI for Incident Reporting

We believe Variant B will be preferred because it offers a single scrolling view, smooth continuity, and predefined answers, reducing user effort while keeping the process structured and intuitive.

3.2 Methodology

To evaluate the impact of different interaction styles on usability, three hypotheses were formulated:

- **H1:** Users will make fewer errors in Variant B than in Variant A.
- **H2:** Users will complete tasks faster in Variant B than in Variant A.
- **H3:** Users will prefer Variant B over Variant A in terms of usability and engagement.

These hypotheses are designed to be evaluated on laptops or desktop devices, as all users interacted with the system in a controlled environment where the application was not deployed.

To test these hypotheses, two key factors were measured:

- Task performance: Measured by time taken to complete the report and the number of errors made.
- User preference: Evaluated through a Standard Usability Scale (SUS)[4] questionnaire after interacting with both Variants.

A website containing both instrumented variants was developed for this study. The website is not hosted, requiring users to access it directly.The source code is available on GitHub. The website flow follows this structured sequence:

- Users enter their personal details (name, age, gender).
- A task description presents them with an incident scenario to report.
- Users complete the incident report using Variant A (form-based UI).
- Users complete the incident report using Variant B (chatbot-based UI).
- A SUS questionnaire evaluates their experience.

Errors are determined by comparing user responses with predefined correct answers. A limitation of this approach is that it assumes users will provide exact responses, which may not always be the case. Users are provided with a printed copy of the task description, allowing them to refer to it while filling in information based on the scenario.

The system was implemented using:

- Frontend: HTML, CSS, JavaScript
- Backend: Python (Flask)
- Database: SQLite (stores task completion time and error counts)

A within-subjects design was chosen to minimize variability between participants, ensuring that differences in task completion time and error rates result from UI design rather than individual skill levels.

A total of 30 participants, aged between 20-30 years, took part in the study, with 43.3% female participation. To test H1 and H2, a paired t-test[8] is conducted to compare completion times and errors between Variant A and B. In addition to overall performance evaluation, we also analyze completion time and error rates across gender to identify potential differences in user interaction patterns. For H3 (user preference), the SUS questionnaire is used, where users rate their experience for each Variant separately. The SUS questionnaire was chosen as it is a widely accepted and validated tool for measuring user experience across different interfaces. We will take the mean SUS score for both variants to quantify and compare overall user satisfaction.

3.3 Report on Data

A total of 30 participants (17 men and 13 women) successfully completed both Variant A (Form-Based UI) and Variant B (Chatbot-Based UI) before filling out the System Usability Scale (SUS) survey. During data collection, all participants followed the structured task procedure, and no incomplete responses were recorded.

To evaluate task performance, we measured completion time and error rates for both variants. The paired t-test results for errors are presented in Table 1, and the paired t-test results for completion time are provided in Table 2.

Table 1: Paired T-Test Results for Errors

Metric	Value
Variant A (Mean ± SD)	5.47 ± 2.01
Variant B (Mean ± SD)	2.13 ± 1.07
t-value	10.95
p-value	<0.00001

Table 2: Paired T-Test Results for Completion Time

Metric	Value
Variant A (Mean ± SD)	455.32 ± 164.49
Variant B (Mean ± SD)	193.93 ± 68.11
t-value	8.22
p-value	<0.00001

In addition to the primary analysis, we compared performance across gender to assess differences in completion time and error rates. The following figures, Figure 3 and Figure 4, present the gender-based comparison.

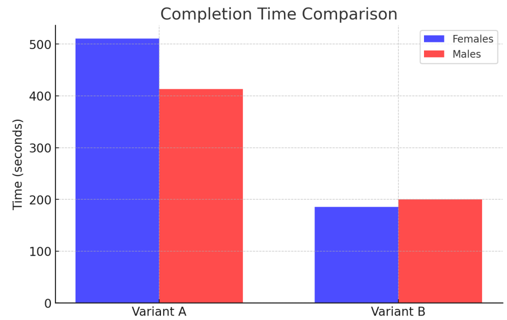

Figure 3: Comparison of Completion Time Between Genders

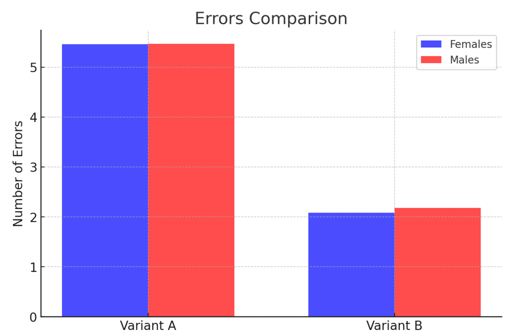

Figure 4: Comparison of Errors Between Genders

Additionally, System Usability Scale (SUS) scores were calculated for both variants. The SUS interpretation scale is shown in Table 3, while the SUS scores for Variant A and Variant B are presented in Tables 4 and 5, respectively.

Table 3: SUS Score Interpretation Scale

SUS Score	Grade	Adjective Rating
> 80.3	A	Excellent
68 – 80.3	B	Good
68	C	Okay
51 – 68	D	Poor
< 51	F	Awful

Table 4: SUS Score for Variant A (Form-Based UI)

SUS Score	Grade	Adjective Rating
42.16	F	Awful

Table 5: SUS Score for Variant B (Chatbot-Based UI)

SUS Score	Grade	Adjective Rating
89.00	A	Excellent

In the next section, the found information is analyzed.

3.4 Analysis

If we compare the reported results with the initial hypotheses presented in Section 3.2, the following can be concluded:

- **H1 is accepted.** Users made fewer errors in Variant B (2.13 errors) than in Variant A (5.47 errors), a statistically significant difference ($p < 0.00001$).
- **H2 is accepted.** Users completed tasks faster in Variant B (193.93s) than in Variant A (455.32s), with a significant difference ($p < 0.00001$).
- **H3 is accepted.** The SUS score for Variant B (89.00 – Excellent) was higher than for Variant A (42.16 – Awful).

- Error rates were higher in Variant A (5.47 errors) than in Variant B (2.13 errors), with a statistically significant difference (t = 10.95, p < 0.00001). In Variant B, female participants made 5% fewer errors than males, whereas in Variant A, error rates were similar across genders. Since users could modify responses before submission, corrections before the final input were not recorded.
- Completion time was longer in Variant A, with an average of 455.32s, compared to 193.93s in Variant B ($t = 8.22$, $p < 0.00001$). Female participants took 23.69% longer than males in Variant A but completed tasks 7.38% faster in Variant B.
- SUS scores were measured for usability. The System Usability Scale (SUS) scores were 42.16 for Variant A and 89.00 for Variant B.

In summary, users made fewer errors, completed tasks faster, and rated Variant B significantly higher in usability than Variant A.
In the next section, these results are discussed in detail.

4 DISCUSSION

During the user study, data was collected and analyzed in the previous section. The findings from this analysis are now discussed.

Users clearly preferred Variant B, which allowed smoother task completion, whereas Variant A required more effort and manual input. While we expected Variant B to perform well, its strong user acceptance was notable, suggesting it was not only efficient but also intuitive and comfortable to use. The guided responses in Variant B structured the process, making reporting effortless, though at the cost of flexibility.

One of Variant B's key advantages was its single scrolling view, which provided continuous interaction, unlike Variant A's step-based format, which may have disrupted workflow. Additionally, Variant B minimized errors by offering predefined responses, reducing manual input effort and the likelihood of mistakes, especially for users unfamiliar with incident reporting systems.

The results confirmed that users completed tasks faster and made fewer errors in Variant B, demonstrating clear efficiency and accuracy benefits. Gender-based trends showed that female participants benefited more from structured guidance, making fewer errors and completing tasks faster in Variant B than in Variant A, suggesting that chatbot-style guidance may be particularly effective for users less experienced with structured reporting environments.

Task familiarity may have influenced user perception. Since participants used Variant A first, they had already internalized the reporting structure, which may have contributed to the smoother experience in Variant B. If the order had been reversed, Variant B's structured responses might have influenced how users performed in Variant A. While Variant B's ease of use was evident, part of its advantage may stem from this learning effect rather than just UI design.

This study had key strengths. It was a controlled experiment with structured tasks, eliminating external confounding factors. The task scenario simulated real-world cybersecurity reporting, making findings relevant for practical applications. The results were statistically significant, confirming notable performance differences between the UIs.

However, some limitations should be noted. The small sample size limits generalizability, and including cybersecurity professionals could provide deeper insights. Variant B's chatbot lacks adaptability, and long-term studies are needed to assess whether users adapt to Variant A over time or if Variant B's structured approach becomes restrictive. Future research should explore how user expertise influences UI preference and performance.

If given unlimited time and resources, the ideal incident reporting system would integrate structured forms with AI-driven assistance, allowing users to switch between manual input and guided interactions. Adaptive AI models could dynamically adjust questions based on user responses and incident severity, ensuring both efficiency and flexibility while maintaining reporting accuracy.

Looking ahead, AI-driven systems could automate threat detection using NLP and predictive analytics, identifying incidents from conversations, security alerts, or anomalies—reducing manual reporting while improving response time and accuracy. However, real-world testing by cybersecurity professionals is essential to refine AI-driven solutions and maintain human oversight.

Overall, this study highlights how structured guidance in incident reporting improves efficiency while real-world applications demand flexibility. While Variant B was preferred for ease of use, Variant A aligns more closely with traditional security reporting. Future research should explore AI-driven solutions that optimize both usability and adaptability, balancing automation with human decision-making to enhance cybersecurity reporting systems.

REFERENCES

[1] Simone Busetti and Francesco Maria Scanni. "Evaluating Incident Reporting in Cybersecurity: From Threat Detection to Policy Learning". In: *Government Information Quarterly* 42.1 (2024). DOI: 10.1016/j.giq.2024.102000.

[2] F. G. Conrad et al. "The impact of progress indicators on task completion". In: *Interacting with Computers* 22.5 (2010), pp. 417–427. DOI: 10.1016/j.intcom.2010.03.001.

[3] J. S. Elson et al. "Analyzing the Effectiveness of a Web Form vs. Chatbot for Suspicious Activity Reporting". In: *Proceedings of the 57th Hawaii International Conference on System Sciences* (2024). URL: https://hdl.handle.net/10125/107243.

[4] Usability Geek. "How to Use the System Usability Scale (SUS) to Evaluate the Usability of Your Website". In: *Usability Geek* (2024). URL: https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/.

[5] Google. *Material Design Guidelines*. 2024. URL: https://m3.material.io/.

[6] M. Hristova. "Chatbot vs. Form Systems for Medical Device Adverse Event Reporting". In: *Master's Thesis, University of Bath* (2024). URL: https://essay.utwente.nl/97739/1/Hristova_MA_BMS.pdf.

[7] IBM. *IBM Carbon Design System*. 2024. URL: https://carbondesignsystem.com.

[8] JMP Statistical Knowledge Portal. "Paired t-Test: Definition, Formula, and Examples". In: *JMP Statistics Knowledge Portal* (2024). URL: https://www.jmp.com/en_nl/statistics-knowledge-portal/t-test/paired-t-test.html.

[9] A. Siddique. *IBM: Carbon UI: Concepts and Design*. Tech. rep. 2023. DOI: 10.13140/RG.2.2.36748.08325.

[10] Y. Song et al. "BURT: A Chatbot for Interactive Bug Reporting". In: *arXiv preprint* (2023). eprint: 2302.06050. URL: https://arxiv.org/abs/2302.06050.

[11] H. Yu et al. "How Much Progress Did I Make? An Unexplored Human Feedback Signal for Teaching Robots". In: *arXiv preprint* (2024). eprint: 2407.06459. URL: https://arxiv.org/abs/2407.06459.