1.1 User Evaluation

Due to the rigour of user testing, we use a combination of qualitative and quantitative research methods to evaluate the quality of SafeTweet from both objective and subjective perspectives and put forward suggestions for improvement through analysis [1].

1.1.1 Objective Perspective

The objective view is to collect and analyse user behaviour data. User behaviour data refers to the data collected through users' actual behaviour. We explain to the users how to use the website. The user is then assigned a task list and asked to complete the tasks one by one in the order of the task list. At the same time, we use screen recording to record data. The data we need is how long it takes the user to complete each task, the wrong path of action, the failed mission, etc. After the user test, we will fill in the summary table according to the user's screen record and record the time-consuming and failed tasks. Analyse any chores that take too long or fail. Analyses the faults caused by viewing user operation paths and provides corresponding improvement suggestions.

1.1.2 Subjective Perspective

Subjective data collect user preference data. User preference data means acquiring users' emotional feelings through questionnaires and interviews. In user-centred website design, Tandon once proposed that the critical factors for measuring website usability are design, performance, usability, aesthetics, and satisfaction [2, 3]. Cry adds to his opinion that there are three main elements of user-centred website construction: information content, navigation design and visual presentation [4]. In addition, Shao uses interactivity, content and navigation to explore the impact of social networks on user satisfaction and viscosity [5]. These factors are used to evaluate and test their products. After considering the product characteristics of social networking sites, the quality factors of SafeTweet are determined as follows: content, usability, interactivity, visual presentation and subjective feelings brought to users.

To further evaluate the influence of influencing factors on the quality of social networking sites, we need to explore further the influence degree of each of the five influencing factors currently locked. We use AHP (Analytic Hierarchy Process) to do the weight analysis on the five influencing factors of SafeTweet, to get the importance of these factors [6]. As seen in appendix A, we conducted the AHP analysis by issuing 20 questionnaires. The scoring criteria are as follows:

Level	Questionnaire Score	Scale
Extremely important	5	9
Bounded by very important and extremely important.	4.75	8
Very important	4.5	7
Bounded by more important and the very important	4.25	6
More important	4	5
Bounded by slightly important and more important	3.75	4
Slightly important	3.5	3
Bounded by equally important and slightly important	3.25	2
Equally important	3	1
Bounded by equally important and slightly secondary	2.75	1/2
Slightly secondary	2.5	1/3
Bounded by slightly secondary and relatively secondary	2.25	1/4
Relatively secondary	2	1/5
Bounded by relatively secondary and the very secondary	1.75	1/6
Very secondary	1.5	1/7
Bounded by very secondary and extremely secondary	1.25	1/8
Extremely secondary	1	1/9

Figure 1 - Proportional scaling regarding the level of importance.

The questionnaire we designed includes ten Likert scale questions, and the influencing factors are compared in pairs. The Likert scale measures people's attitudes scientifically approved way [7]. We then selected the questionnaire for experienced social network users and invited them to answer it. Thirty-one questionnaires were issued, and 30 were recovered, with a recovery rate of 97%. After the questionnaire was collected, the collected results were averaged to construct a 5x5 judgment matrix. SPSS was used to calculate the weight of the matrix using AHP, and the weight values of the five influencing factors were obtained, as shown in the figure below. A subsequent consistency test of the questionnaire results is required to ensure no logical problems with the matrix. After calculation, the CR value is 0.0836<0.1, so the consistency test is passed, and the weight value is reasonable.

	Content	Visual Presentation	Usability	Subjective Feelings (e.g., safety, respect, pleasurable)	Interactivity	Eigenvalue	Weight
Content	1	5	3	1	2	1.9744	33.10%
Visual Presentation	0.2	1	0.3333	0.3333	2	0.5365	8.99%
Usability	0.33	3	1	1	4	1.3195	22.12%
Subjective Feelings	1	3	1	1	5	1.7188	28.81%
Interactivity	0.5	0.5	0.25	0.2	1	0.4163	6.98%

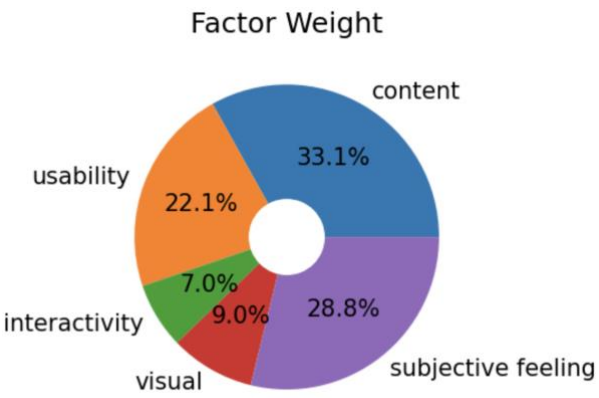

Figure 2 – Weights of influencing factors.

As can be seen, from the above figure, people do not have a high demand for social networking sites' visual presentation and interactivity. The proportion is about 9% and 7%, respectively. Previous research has shown that the usability of a website depends on its visual complexity to some extent [8, 9]. Compared with sites with high visual complexity, users will get more pleasure from sites with low visual complexity. Popular social networking sites also have an excellent unified design and simple UI. Too many elements will make users unable to find the key point, and the user experience will worsen. The interaction on social networking sites is relatively fixed and essential, like comments, likes, etc. Therefore, users do not expect too much from the interaction of social networking sites. At present, the primary interaction functions of SafeTweet can already meet users' interaction needs on social networking sites.

Social networking sites' content, subjective feelings, and usability have received great attention. It can be seen that users are eager for and accustomed to obtaining high-quality information from the network open platform. Another point worth paying attention to is the subjective feelings of users. Social networking sites are often flooded with information, including users' private information. SafeTweet, a private social networking platform dedicated to giving people a voice in the workplace, has more privacy problems. Therefore, for the main body of this website, that is, users, their subjective feelings are significant. SafeTweet wants users to feel respected, safe and pleased when using it. When the above sentiments are satisfied, users will have more positive emotions towards SafeTweet, thus improving their satisfaction with SafeTweet [10]. In addition, usability is often reflected in the user's interaction with the website. Such as site fault tolerance, convenient and intuitive site navigation, and pages are working normally. If the user experience is not good in the process of website interaction, it will directly affect the user's evaluation of the website [3, 11]. Below is a flow chart for the entire evaluation section.

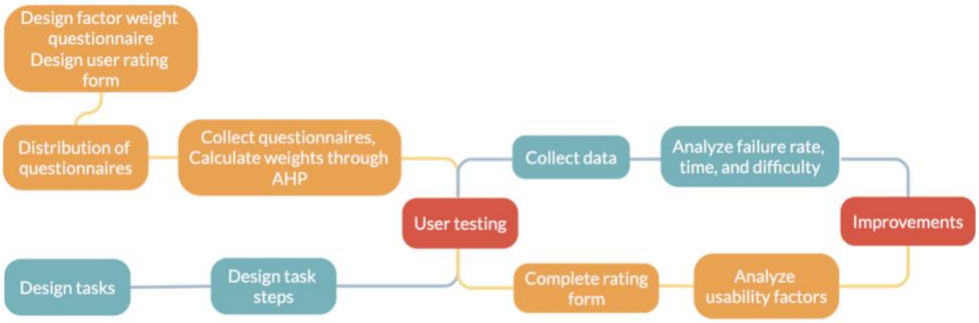

Figure 3 – Overall assessment process.

1.1.3 Evaluator

With the determination of SafeTweet influencing factors and their weights, the next step is the user test. We selected 30 volunteers to participate in our user test. They included 28 people aged 19 to 27 who used social networking apps regularly. Two participants, aged 37 and 45, are unfamiliar with popular social networking sites. There are 20 participants in the Internet industry, including students, developers and software product managers. In addition, there are ten non-Internet industry participants, including Japanese, medical, business, public security, etc. We first selected two pilots with experience in testing related products to conduct overall control of the test, including the introduction before the test, test sequence, test topic and test quantity.

1.1.4 Evaluation Processes

We divided the entire evaluation process into three parts. In the first part, we introduced our testers to the basics of SafeTweet. It was designed to protect free speech in the workplace by hiding users' information and content and introducing testers to the basics of SafeTweet. In addition, participants should be informed that SafeTweet is currently a Prototype. The definition of the prototype should be presented to the testers so that the inexperienced testers can have a basic understanding of prototype, to improve the accuracy of test results. After completing the basic introduction, ask the tester if they have any other questions and answer them.

The second part is to complete the tasks according to the task list, which is provided in appendix B. The task sheet contains all the functions that users can interact with SafeTweet. The task sheet is designed according to the classification of parts of the MoSCoW table. The categories include Basic Functions, Send Posts, Information Detection, Read Posts, Decrypt Posts, Comment Posts and Search Posts. Testers need to complete all tasks under each functional category. The entire test process needs to be recorded to record the completion time of each task and the problems encountered during the task. The moderator should not answer

any questions while the participant performs the test. Participants are required to complete all tasks in the task list by themselves.

The third part is the questionnaire survey after the test. The questionnaire also includes three parts. The first part is the system availability scale. Brooke first compiled SUS in 1996 [12]. It can scientifically quantify user experience and measure the overall perceived availability of products or systems after completing a series of task scenarios, avoiding premature attention to details early. It is a robust system usability measure [13, 14]. SUS consists of 10 questions, with odd numbers as positive statements and even numbers as negative statements. Each question is rated on a scale of five, from strongly disagree to strongly agree [15]. SUS has two sub-scales. The fourth and tenth questions constitute the "Learnability" sub-scale. The remaining eight items constitute the "Usability" subscale [14].

The second questionnaire is the user experience questionnaire, which is provided in appendix C. This part of the questionnaire was scored according to SafeTweet's influencing factors, and the Likert scale was still used. The influencing factors are five factors through AHP weight analysis: Content, Usability, Interactivity, Visual Presentation and Subjective feelings. In each influencing factor, specific sub-influencing factors are divided in more detail. The third part of the questionnaire is the short answer part. The questions include the overall view of SafeTweet, the likes and dislikes of SafeTweet, and the impact of latency on usage when monitoring sensitive information.

1.1.5 Assessment Steps

Because of the pandemic, all tests are being conducted online. Each tester was asked to download sunflower Remote Control, a system that could connect two devices remotely. After the tester is paired with the host's computer, the tester can operate on the host's device to complete the test.

1. The test time for all evaluators is 9:00-12:00 am and 2:00-5:00 PM. A tester performs each test.
2. The host introduces SafeTweet to the testers.
3. The moderator introduces the task list to the tester. If participants have any questions, the facilitator will answer them.
4. Before the test starts, enable screen recording and record the whole test process.
5. Evaluators need to complete each task in sequence, and the successful completion will be recorded as the success of the task. If the evaluator could not complete or abandon a task, it was a task failure. The moderator should use a stopwatch to record the completion time of each task synchronously on the schedule.
6. The moderator shall not instruct the tester to complete the task during the test.

7. After the task, participants should complete three questionnaires: SUS, UEQ and short answer sections.
8. After the test participants complete the questionnaire, the host should retrieve the task sheet and questionnaire and adequately store it together with the screen recording of the renamed participant. A complete task list ensures that 80% of the tasks have been completed and that the analysis data is valid.

1.2 Result Summary

1.2.1 Objective Analyse

This section summarises how long it took each participant to complete each task, the total time it took to complete all tasks, and the average time for each task.

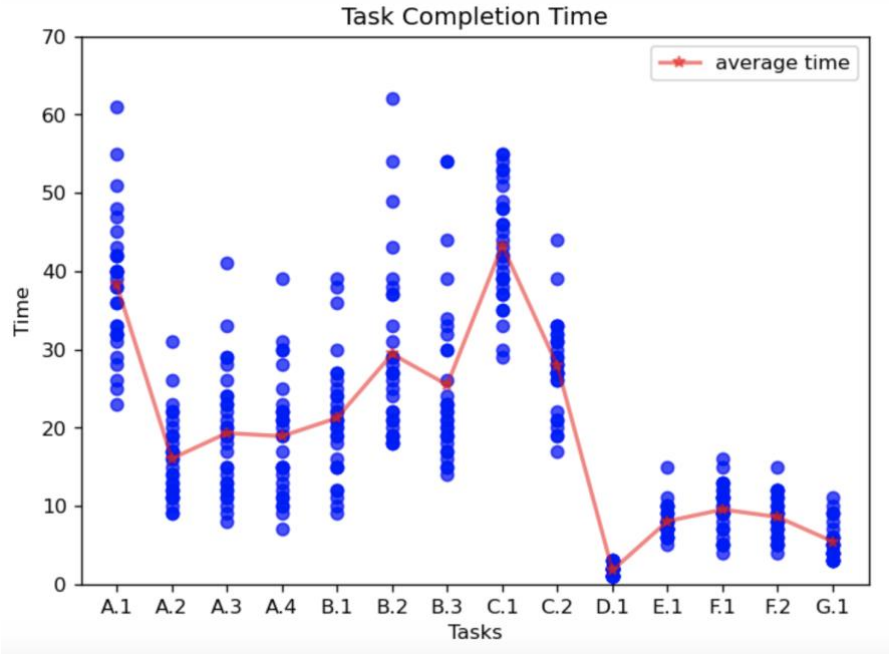

Figure 4 – Task completion time.

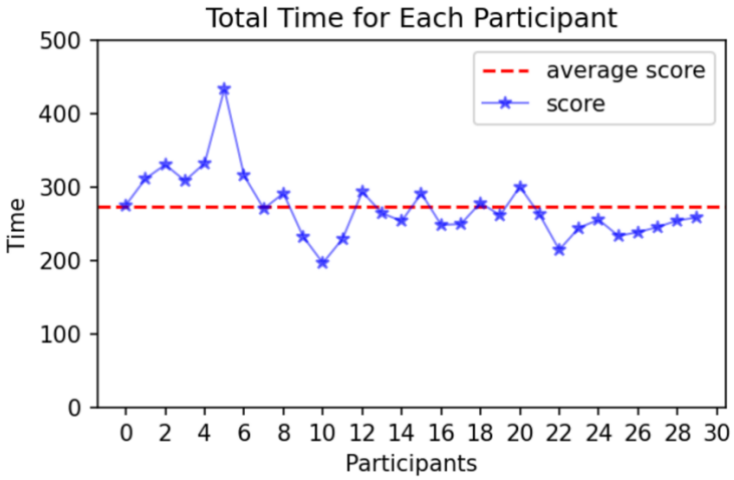

Figure 5 – Total time for each participant.

As shown in figure 4, there are 14 sub-tasks in total, and the average time to complete all tasks is 274 seconds. Evaluator No. 5 took the most extended total of 434 seconds. No. 5 participant is inexperienced in social networking sites, and the total time consumption was higher than other experienced participants. It takes time to understand the purpose of the task, figure out what each button does, and check each action carefully. Participant No.5 is also the evaluator who spends the longest time in tasks such as registration (A.1), posting sensitive information anonymously (B.2), and detection of sensitive information (C.1), so there are seemingly abnormal time nodes in Figure 5.

The C.1 task, which took the longest average time of all tasks, was 43 seconds. The mission of C.1 is to input sensitive content and use the sensitive information monitoring function of the system to detect it. The main problems are:

1. The tester has no concept of detecting sensitive information. It cannot accurately control the system's sensitivity, so it will spend some time thinking about sensitive content.
2. Some evaluators ignore that SafeTweet is a social networking site mainly based on the work scene. When they use sensitive information utterly unrelated to the work scene, the system may not be able to identify it.
3. The sensitive information system can detect various sensitive information and prompt users on which kind or more sensitive information is leaked. Some testers spontaneously hope to test what types of sensitive information the system can recognise by inputting different sensitive information.

In addition, the other long task is A.1, which takes 38 seconds on average. The main content of this task is to ask the user to register an account. As five items of personal information need to be filled in, and when setting the password, most of the testers have to think about it, so it generally takes a long time. On average, the shortest task was D.1, which lasted only 1.8 seconds. The main content of this task is that testers select the content they are interested in to give a thumbs-up, which is an interactive function.

Throughout the test, younger participants with experience using social networking sites could complete the task in a short time. In the case of Evaluator 10, the time to complete all tasks is 197 seconds. Both experienced and inexperienced social network users completed all tasks in the task list smoothly, which means SafeTweet is not limited to a user's age, occupation, or even social network experience. To sum up, SafeTweet, as a prototype, has the characteristics of simple operation and easy learning.

1.2.2 Subjective Analyse

1.2.2.1 Questionnaire One – SUS Scale

At the end of the test, each participant received three questionnaires, as mentioned earlier. The first of these is SUS. According to the research of Brooke, the average score of the current SUS scale is 68 [16]. A score above 68 indicates above average, while a score below 68 means below average. SUS scale is calculated as follows: Odd-numbered items subtract 1 from the user score. For even-numbered items, the user score is subtracted from 5. This scales all values from 0 to 4 (4 is the most positive response). The conversion scores for each user are added and multiplied by 2.5. This transforms the range of possible values from 0 to 100.

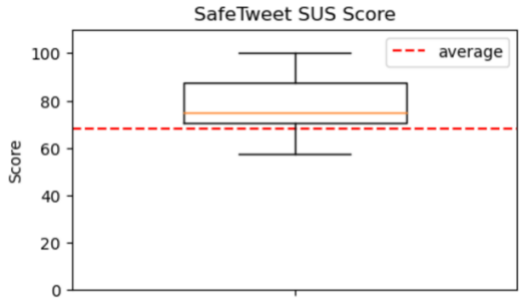

Figure 6 – SUS score box graph.

According to the box graph presented by the SUS questionnaire, the average SUS of SafeTweet is 77.25. The median and lower quartile scores are above 68, proving that SafeTweet's user experience is qualified.

1.2.2.2 Questionnaire Two – UEQ Scale

The second part of the questionnaire is based on the UEQ (User Experience Questionnaire) customised by SafeTweet. The questions are designed around the quality influencing factors conducted before the start of the test, and there are 2-9 questions under each influencing factor. SafeTweet's quality factors are mainly divided into five parts: content, usability, interactivity, visual effects and subjective feelings. Thirty participants answered the UEQ questionnaire truthfully and accurately within 10 minutes after the test, with an average score of 3.84 for all final results. According to Figure 7, SafeTweet performed above average in usability and interactivity and nearly average in content with the score bar chart of influencing factors. The visual performance of the influencing factors is poor in THE score of UEQ, followed by the subjective feelings brought to the user are also lower than the mean.

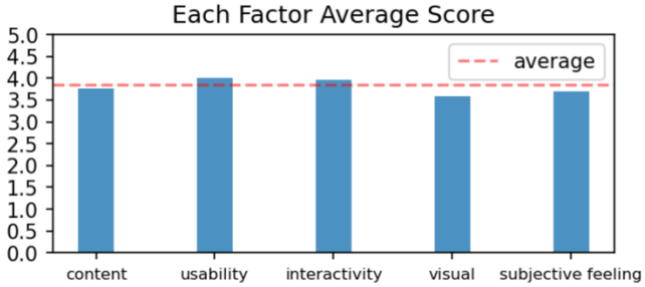

Figure 7 – Each factor's average score.

Due to SafeTweet being only a prototype for a social network, there was not enough front-end design. Since the visual effect is intuitive and can be perceived obviously by users in the test, users will give strict marks to this part. Secondly, the user experience with a lower-than-average score also reflects that the prototype product is not attractive enough to users. As for content influencing factors, SafeTweet is only a prototype at present, and its size is small, so it cannot fully display its quality potential content. So, its score is roughly in line with the average. However, in SafeTweet's usability and interactivity section, the above-average score confirms that SafeTweet's current features perform well and meet users' needs.

1.2.2.3 Questionnaire Three – Short-Answer Question

The third part of the questionnaire consists of three short answer questions, which are text-based. So, we chose the card-sorting to export the topic from the text [17]. The first question asked participants about their overall attitude to SafeTweet. We started with the "open" classification type of card classification, which exposes definitions and new taxonomy categories. In this process, we pre-classified ten responses to extract the participants' attitudes towards SafeTweet. The participants' attitudes were divided into four main categories: positive, reserved approval, neutral, and negative. We then classified the remaining 20 responses using the "closed" classification type of the card classification. The final result is as follows:

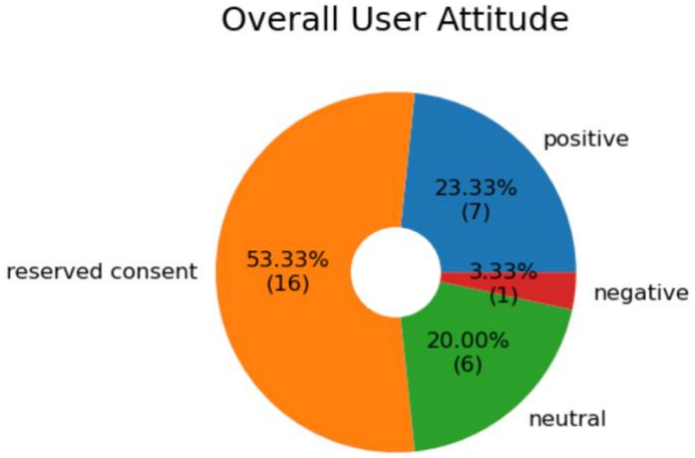

Figure 8 – Overall User Attitude.

It can be seen from Figure 8 that half of the participants have a reserved approval of SafeTweet, and nearly a quarter of the participants have a positive attitude towards SafeTweet, which means that SafeTweet has many advantages. The number of neutral testers was close to the number of positive testers who had reservations about SafeTweet. One person remained negative about using SafeTweet, for reasons mentioned in the second question.

The second question asked participants what they liked and disliked about SafeTweet. Sort the cards with the views of the participants in the first question. Among them, 55 were favourable to SafeTweet, and 22 were unfavourable. Because SafeTweet's system is so large, we use an "open" taxonomy throughout

the process so that new tabs can be added at any time. At the same time, to better analyse the results, we adopted a two-level classification method to make the results more intuitive. In addition, we have a detailed description of each label. The final card classification results are as follows:

High level taxonomy	Low level taxonomy (Count)	% of Cards	Description
Functions	Anonymity (11)	20	The function of posting anonymously.
	Encryption (15)	27.3	The function of posting encrypted information.
	Acquisition (1)	1.8	SafeTweet is a place for people to talk and exchange information.
	Detection (5)	9.1	The function of detecting sensitive information and give advice.
Feelings	Safe (4)	7.3	SafeTweet gives users a sense of security.
	Respect (2)	3.6	Users feel respected when using SafeTweet.
	Friendly (1)	1.8	Users think SafeTweet is a friendly social website.
	Free (4)	7.3	Users feel free to use SafeTweet.
Design	UI / Interface (4)	7.3	Website layout, interface, font and icon design.
Innovation	Innovate (6)	10.9	SafeTweet has new ideas as a social network website.
Potential	Value (1)	1.8	SafeTweet is commercially competitive.
	User (1)	1.8	SafeTweet will be popular with young people.

Figure 9 – Users like part.

Most testers liked the anonymity, encryption and detection features of SafeTweet. Because of these features, 10% of respondents also mentioned that they liked SafeTweet's breakthrough innovation in social networking. Users have also praised the simple interface and UI design. In addition, nearly 20% of respondents said SafeTweet made testers feel safe, friendly, and free. Some testers even think SafeTweet has many potentials, that it will be very competitive in the market and will be loved by young people.

High level taxonomy	Low level taxonomy (Count)	% of Cards	Description
Functions	Add (6)	27.3	The overall function is not enough, should develop other features.
	Detection (1)	4.5	People from different countries, races, occupations, and ages may have different definitions of sensitive information.
	Improve (4)	18.2	Some functions need to be improved, and some details need to be adjusted.
Feelings	Block (2)	9.1	The feeling that blocked by others is uncomfortable.
Design	UI / Interface (4)	18.2	The UI design is too monotonous.
Law	Legitimate (1)	4.5	Without filtering or restricting the information posted, people may make inappropriate and even illegal comments.
Potential	Competition (1)	4.5	SafeTweet is difficult to seize existing market.
Usage	User security (2)	9.1	SafeTweet may steal user information.
	Use scenario (1)	4.5	Use scenarios are limited to the workplace.

Figure 10 – Users dislike part.

Figure 10 shows what the testers think is weak. First of all, SafeTweet, as a social networking site, is not functional enough. It needs to add essential functions such as follow and @other users. Other participants mentioned that some features were poorly handled in detail, such as the inability to locate specific posts after searching for keywords. Some testers also deemed the interface to be too simple to be beautiful. In addition, some users get an unpleasant feeling of being blocked when they read anonymous posts and know they can't read encrypted posts from employees of the same company. They don't like the feeling. In the first question, another tester, which was pessimistic about SafeTweet, had three questions:

1. The detection of sensitive information is delayed, and there is no guarantee that the same set of sensitive information detection functions can be applied in different countries worldwide.
2. Not filtering anonymous comments may lead to people abusing the feature to make inappropriate comments.
3. There is no guarantee that SafeTweet will infringe on users' sensitive information.

The third question is about the delay of the detection function of sensitive information, asking whether the tester has found the delay and whether they think the delay function impacts the use of SafeTweet. We used a "closed" taxonomy of card classification, pre-defining three labels: “Not noticed”, “Noticed and accept”, and “Noticed but cannot accept”. The final results of the 30 testers collected are as follows:

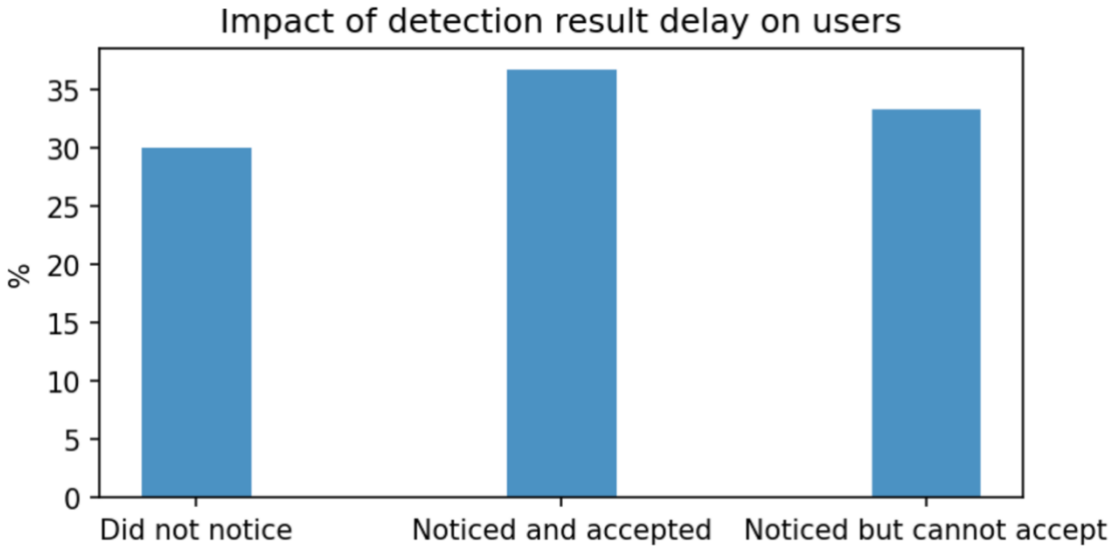

Figure 11 – Impact of time delay.

In Figure 11, nearly one-third of the participants did not realise that the detection system had a delay of about 5 seconds. Of the remaining testers aware of the delay, more than half said it was a normal system delay and had no impact on SafeTweet use. Still, a third of the participants found the delay unacceptable. They believe this is one of SafeTweet's core features and should be optimised as much as possible.

1.3 Future Work

In future work, the first direction of continuous efforts should be to improve the efficiency of sensitive information detection. The model inspection time is determined by the long-running and invocation time of the model. This needs to be addressed by optimising the algorithm and adopting better model invocation methods. The optimisation algorithm modifies the practice and flow of data preprocessing and improves the speed. The model invocation method is the faster invocation method that the author is currently trying. A more appropriate model invocation package may be needed to address this issue, or the project may be developed using the Python framework Django in the future. The second thing to tackle is more elegant encryption. SafeTweet currently uses Base64 for plain text transcoding in a sloppy fashion.

The existence and availability of encryption can be significantly improved if the key is securely transmitted using asymmetric encryption. Next, sensitive information detection standards should be considered, and corresponding adjustments should be made to the detection standards of sensitive information in different countries and regions. In addition, guiding users to use SafeTweet with their real names is also a problem. It can be considered to add recruitment and job-hunting functions in SafeTweet, which may be helpful to the real-name system of users. For anonymous comments, consider filtering your comments properly to avoid direct insults. Finally, I will adjust the UI design and continue to complete more essential functions of social networking platforms, such as following other users and uploading pictures and videos.

References

[1] Rahman, M.S., *The advantages and disadvantages of using qualitative and quantitative approaches and methods in language "testing and assessment" research: A literature review.* 2020.

[2] Tandon, U., R. Kiran, and A.N. Sah, *Customer satisfaction using website functionality, perceived usability and perceived usefulness towards online shopping in India.* Information development, 2016. **32**(5): p. 1657-1673.

[3] Dianat, I., et al., *User-centred web design, usability and user satisfaction: The case of online banking websites in Iran.* Applied Ergonomics, 2019. **81**: p. 102892.

[4] Cyr, D., *Modeling web site design across cultures: relationships to trust, satisfaction, and e-loyalty.* Journal of management information systems, 2008. **24**(4): p. 47-72.

[5] Shao, Z., et al., *Examining user satisfaction and stickiness in social networking sites from a technology affordance lens: uncovering the moderating effect of user experience.* Industrial Management & Data Systems, 2020.

[6] Podvezko, V., *Application of AHP technique.* Journal of Business Economics and Management, 2009(2): p. 181-189.

[7] Joshi, A., et al., *Likert scale: Explored and explained.* British journal of applied science & technology, 2015. **7**(4): p. 396.

[8] Chen, M., *Improving website structure through reducing information overload.* Decision Support Systems, 2018. **110**: p. 84-94.

[9] King, A.J., A.J. Lazard, and S.R. White, *The influence of visual complexity on initial user impressions: testing the persuasive model of web design.* Behaviour & Information Technology, 2020. **39**(5): p. 497-510.

[10] Desmet, P., *Faces of Product Pleasure: 25 Positive Emotions in Human-Product Interactions*. 2012. 2012.

[11] Dewi, C.R. and G.S. Darma, *Website usability, satisfaction, loyalty, security perception, trust and word of mouth in e-commerce business.* Jurnal Manajemen Bisnis, 2014. **11**(2): p. 1-30.

[12] Brooke, J., *Sus: a "quick and dirty'usability.* Usability evaluation in industry, 1996. **189**(3).

[13] Bangor, A., P. Kortum, and J. Miller, *Determining what individual SUS scores mean: Adding an adjective rating scale.* Journal of usability studies, 2009. **4**(3): p. 114-123.

[14] Lewis, J.R. and J. Sauro. *The factor structure of the system usability scale*. in *International conference on human centered design*. 2009. Springer.

[15] Orfanou, K., N. Tselios, and C. Katsanos, *Perceived usability evaluation of learning management systems: Empirical evaluation of the System Usability Scale.* The International Review of Research in Open and Distributed Learning, 2015. **16**(2): p. 227-246.

[16] Brooke, J., *SUS: a retrospective.* Journal of usability studies, 2013. **8**(2): p. 29-40.

[17] Zimmermann, T., *Card-sorting: From text to themes*, in *Perspectives on data science for software engineering*. 2016, Elsevier. p. 137-141.

Appendix A AHP Questionnaire

1. Compared to the **visual presentation**, you think the **content** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

2. Compared to the **usability**, you think the **content** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

3. Compared to the **subjective feeling** (e.g., safety, respect, pleasurable), you think the **content** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

4. Compared to the **interactivity**, you think the **content** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

5. Compared to the **usability**, you think the **visual presentation** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

6. Compared to the **subjective feeling** (e.g., safety, respect, pleasurable), you think the **visual presentation** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

7. Compared to the **interactivity**, you think the **visual presentation** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

8. Compared to the **subjective feeling** (e.g., safety, respect, pleasurable), you think the **usability** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

9. Compared to the **interactivity**, you think the **usability** of a social network site is

   Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

10. Compared to the **interactivity**, you think the **subjective feeling** (e.g., safety, respect, pleasurable) of a social network site is

    Strongly disagree 1 ○ ○ ○ ○ 2 ○ ○ ○ ○ 3 ○ ○ ○ ○ 4 ○ ○ ○ ○ 5 ○ Strongly agree

Appendix B Task List

ID	Test Content	Pass	Duration
A. Basic Function			
1	Register an account.		
2	Log in with registered account.		
3	Upload an avatar.		
4	Change your email address.		
B. Send Posts			
1	Send a public post.		
2	Send a sensitive information anonymously without identity information.		
3	Send a encrypted sensitive information with identity information.		
C. Information Detection			
1	Input sensitive information in text box and check sensitive information.		
2	Input non-sensitive information in text box and check sensitive information		
D. Read Posts			
1	Give a thumb up to a post that interest you.		
E. Decrypt Posts			
1	Decrypt a post and try to copy.		
F. Comment Posts			
1	Comment a post with emojis.		
2	Comment a post anonymously.		
G. Search Posts			
1	Search posts by entering keyword.		

Appendix C UEQ Questionnaire

ID	Category	Question	Score 1 - Strongly disagree 5 - Strongly Agree
1	Content	Users can read a wealth of information.	
2		The content on the site is rich and varied.	
3		Users can get real-time information.	
4	Usability	Users can manage their personal information.	
5		Users can post information.	
6		The search function is effective.	
7		Site status and feedback are visible.	
8		Site fit people's logical thinking.	
9		Users can check for sensitive information before Posting it.	
10		Users can post information anonymously.	
11		Users can encrypt their posts.	
12		Users can decrypt messages encrypted by others.	
13	Interactive	Users can comment on other content.	
14		Users can thumb up other content.	
15	Visual Presentation	The layout is reasonable.	
16		The font design is reasonable.	
17		The color scheme is reasonable.	
18		Page elements have consistency.	
19	Subjective Feeling	Users feel secure when using the site.	
20		Users feel respected when they use the site.	
21		Users experience pleasure when using the site.	