Core assignment: Evaluation [+10] Group

Phase: Evaluation
Due date: Friday, March 14 at 11:59pm PT
The goal of this stage of the project is to evaluate your high-fidelity prototype with potential users.
Course learning objectives this assignment facilitates: (1) Create an interactive system grounded in user research, iterative prototyping, and evaluation; (2) Explore and express complex problems, design choices; and (3) Reflect on what you know, don't know, and how to learn what you don't know.
Grading: This is a group assignment. All members of the group are expected to participate fully. Each group will receive one grade. / Every member will receive the same number of points.
What to submit: Add a .MD file in your Github repo with the responses to the following questions. **Submit the github repo link to BruinLearn. Only one person needs to submit for the group.**

Conducting user evaluation [+5]

Your goal is to assess the usability of your system.
1. **[+5] Evaluate your system with at least 10 participants. Write and submit 0.5-1p of notes for each participant.**
2. **[+1] DEPTH: Evaluate with 5 more participants. Feel free to make changes to your system between the first and second round of evaluation. If you do make changes, summarize the changes you made and why.**

Evaluation notes can be found in the file:

After your evaluation [+4]

1. **[+4] Analyze your data and write up your key findings. The findings should be about 0.5-1p for each motivating question and any other interesting findings.**
   - **For any qualitative data where you cannot easily remember the details of the results, a thematic analysis is required. When you conduct a thematic analysis, include your codebook.**
   - **For any quantitative data, submit a script for analysis + your data. Recommendation: Create a notebook for your analysis. Someone should be able to run your notebook to reproduce your results.**

Evaluation analysis of the interviews and quantitative data can be found in the file:
Evaluation_Result_Analysis

The python script analyzing the quantitative data can be found in the file:
evaluation_analysis.py

- The script looks at typical statistical metrics like mean, median, and standard deviation, correlation between variables, outliers, and tries to construct a regression for total time on the app.

Group Reflection [+1]

1. **What is one thing that went well in your evaluation?**
   a. One thing that went well in our evaluation was the standardized set of questions that we had prepared. Having that prepared beforehand, ensured that we were able to get all the vital information that we wanted, while also allowing our users to naturally give their own feedback while using the application.
2. **What is one thing that you wish you could have done differently?**
   a. We wish that we were able to add some more of the features that we wanted to add, based on our initial pilot, before doing these pilot evals. We wanted to add a way for the LLM to cite where exactly they got this information from in the "Learn More" section, but due to time constraints and focusing on other improvements, such as a transition to React, we were not able to complete this. We noticed that some of our users wanted us to add quotes from the Privacy Policy directly, and it would have been cool if we were able to implement that.
3. **How, if at all, did your participants represent the personas you intended to design for?**
   a. I believe our participants represented the personas that we intended because a lot of our participants are people who have not read the privacy policy before. Through our system, they learned a good amount of information that they were not aware of previously. For example, we had users interact with the information about Medical and Financial data for an average of 13.33 seconds, and some users even had questions or expressed concern about these topics, which suggests that they had not even realized the potential privacy risks before using our app.
4. **How do you think this impacted your results?**
   a. This impacted our results by reinforcing the idea that users are interested in a concise, digestible version of privacy policies. Many of our pilot users engaged with key pieces of information and expressed a desire to learn more, indicating that our system probably filled an existing knowledge gap.
5. **Based on the above, what does this say about the potential applicability of your system?**
   a. Based on what we observed through our pilots, our popup has good potential to improve user awareness of privacy policies, especially for users who may overlook this information. By presenting key points in an easy to digest and accessible way, the system can help users make more informed decisions about their data privacy.
6. **What new questions do you have based on your evaluation?**
   a. Some questions we have include:
      i. What is the optimal level of detail to include without overwhelming users?

ii.   How can we further personalize the experience based on individual privacy concerns?

iii.   What is the best way to enhance the credibility of our system?

Did you use a generative AI tool for this assignment? If so, which tool(s) and how?
Yes, we used ChatGPT to help determine good statistical tests/metrics to analyze our quantitative data.

How much time did you spend on this assignment as a group? Individually?
We each spent around 1.75 hours doing the different evaluations, as we did 5 each.

Brooke + Nikhil spent at least 2-3 hours each trying to shift our frontend to react, only to scrap it due to it being unfeasible.

Brooke spent 2 hours writing up the eval notes.

Nikhil spent 1 hour doing our qualitative analysis.

Sruthi spent 1 hour writing up the evaluation reflection.