

Present the pilot user with a brief statement of the scenario and task. Ask the pilot user to complete the task. Note: You might feel (very) nervous that something will break. That is OK. It's ok for the pilot user to break things as they test out your system. Be prepared to restart/recover your system when things break. Note what happened step by step. Include 0.5-1p of notes on one pilot user. Additionally, summarize in a few sentences: What happened? Why? What changes do you need to make to your system before the next pilot?

Pilot Users in Our Class

Kalyan

- Felt like there was definitely too much text, and also wasn't sure what the text meant without clarifying definitions
- Didn't like the star rating system
 - He felt like he was on yelp and had to "review" the system himself
 - Wasn't clear that it was a rating of the system done bby LLM
 - He suggested switching to something else, for instance using a 'scale' like the smoky the bear fire risk scale (that's the only example I can think of)
- Thought it needed more color to draw user in
- Needed more options for depth of understanding
 - EG, Have links that say "learn more" that have more text heavy pages, and make the main page less text heavy. Also, having definitions of what the categories mean would be helpful
- He made several suggestions for improving the UI by incorporating tools like React or Svelt, along with <https://ui.shadcn.com/>
- Overall liked the popup style but didn't like that when he clicked away from the popup it vanished
 - Suggestion to make it injectable (like Honey's popup)
- Overall, good idea just needs more fine tuning to get users truly engaged
- Prompt engineering could be improved
 - More specificity and less legal jargon

WHAT HAPPENED/WHY:

Kalyan navigated to instagram.com and clicked sign up. When he got to the sign up page, a popup appeared saying "waiting for summary..." He clicked it but nothing happened, the popup isn't interactive until the summary loads in. After a few moments, the summary popped up and he read through it. He was confused about what some of the contents meant. When he went to click somewhere else on the page the popup vanished. This is because of

the nature of the popup, they automatically close when clicked away from. This is something we improved upon in the next iteration, injecting it directly into the webpage rather than having it be a popup. This way, it persists even if the user clicks elsewhere. We also need to improve on the UI so that it is more informative and easy to understand at a glance instead of having to focus on large chunks of text.

Michael

- We asked our user to open [instagram.com](https://www.instagram.com) and create an account
- As they went to the create account page, the popup showed up but they did not see it, but rather clicked to input their information (like their phone number or email) immediately and that made the popup close without them even noticing that the popup was there
- This made us realize that we should probably inject the page with our popup, instead of displaying it like we currently are – this is to make sure that whatever just happened does not happen again, and people can actually see the popup.
 - One comment was that honey, the coupon popup, injects the popup into the page and while it might be more “annoying” to see it since it persists, as long as there is a way to close it manually then it should be fine
- In addition, we got questions about what the stars represent. It was not immediately clear as to what the rating stands for, is it ranking how much of an issue the privacy policy is or is it ranking how good the data privacy is.
- We were then given suggestions as to add words under the stars saying good or bad rather than just displaying stars. Another idea was to have a scale saying bad to good and with a line indicating the number it is ranking it as. This is so that it is much more intuitive as to what is being displayed
- In addition, since we are using LLMs, another idea is to get the response from one API call, and then send the response to another LLM to get the privacy rating with some examples as to what good ratings are vs what bad ratings are
- Another suggestion that was made was to break the text block into bullet points or have it so that users have to click to learn more

WHAT HAPPENED/WHY:

During our pilot session, our popup was unintentionally dismissed, highlighting the need to inject it into the page for visibility. The star rating was also found to be unclear, suggesting labels or a bad-to-good scale. To improve readability, we should break text into bullet points or collapsible sections. Lastly, refining our LLM pipeline by using one API call for analysis and another for rating could enhance accuracy. These changes will improve clarity and user experience for the next pilot.

Daniel

- We provided context that Daniel was a user who was interested in making a social media account (Instagram, Facebook), but was also curious about data privacy and had an extension installed to inform him on his options
- Noticed the popup when navigating to the facebook signup page
- Took the time to read thru the descriptions and ratings
- Felt more informed about Meta privacy policy
- Was unclear about the meaning of the stars for rating
 - More = good or bad?
- Wanted shorter summaries for each issue, or highlights
 - Access to financial data YES/NO
 - Data Storage -> Length
- Thinks a injected popup could work better but is fine with current extension popup


WHAT HAPPENED/WHY:

We gave Daniel better context surrounding our app since he was the last user we gave this pilot to. He knew that there would be an extension, so he wasn't surprised when there was a popup when he opened Facebook's sign up page. He took the time to read through the text and said he liked the core idea of making it easier to understand privacy policies. However, he was a bit confused by what the ratings meant like our other users. Additionally, he wasn't too impressed with the LLM summaries of the categories since they were quite vague and long considering how much they had to say. There is definitely room to improve the content of the popup, but the general layout and idea seemed to be good.

Pilot Users Outside Our Class

Corin

- Initially confused by the popup and what was going on, since it just said waiting for summary
- Popup structure was confusing since the stars are currently broken... once explained it made more sense
- Corin felt like the prompt engineering was poor and didn't really understand the policy even after reading the summaries
 - They were still very jargon-y and contained legal stuff
 - Wanted more actionable insights and understanding of the details of what it meant
 - Instead of 'summarizing' the policy he wanted the policy to be more explained which it wasn't

- Instead of text block he suggested bullet points
- He also thought static text describing key components of policy and its implications would be very beneficial to have
 - EG, describing what our 3 categories meant exactly
- Felt like the stars were good and helped him understand
 - More stuff like that would be helpful
- Corin also stated that “if I was about to make an Instagram account this wouldn’t stop me”
 - Because of this he suggested having more visuals demonstrating how dangerous/not dangerous those things are
 - EG, use red color to denote worse policies, have more visuals like the emergency/warning sign 

SUMMARY:

After a brief introduction, Corin went to instagram as if he was making an account. When the injection div appeared that said “waiting for summary” he stopped signing up and waited for it to pop up. Once It did he skimmed it and started signing up again. I prompted him to stop signing up and talk through his impressions of the popup. He said he skimmed through and kept signing up because it was 1. Very brief and 2. Too much text. He also said it just didn’t impact/wouldn’t impact his decision to sign up. In order to stop people from just skimming we need to make the popup more visually appealing and interactive. This includes having more color and visuals, but also layering the popup so that users have to click more to see other pages. This interactiveness and organization will organize the information more clearly, and also keep users more engaged.

Harshi

- This test was done on facebook.com with someone who is not in our course and does not have much previous context about the project.
- The user navigates to facebook and clicks create account. The user saw something flash but didn’t care as she immediately went to create an account.
 - This emphasized the need for us to inject the popup
- Then she opens the popup and sees these star rankings and asks “Does three stars mean it’s good or bad for privacy?”
- From there, she read some of the summaries that were generated and saw that it was pretty vague and asked how we are getting these summaries. Since she does not have a tech background, we had to explain what the LLM does.
 - She mentioned that it would be nice to have quotes directly from the privacy policy in these specific subsections

- She also said that it would be cool if the system could explain why it gave this rating, maybe comparing it to good and bad examples
- She overall thought it was a good idea but it needs more refining to make it more clear if it is good or not as well as suggested changes she can make in her usage of the app.

SUMMARY:

During our facebook pilot, the user dismissed the popup without noticing it, highlighting the need to inject it into the page. She found the star ratings unclear and the summaries vague, suggesting direct quotes from the privacy policy and explanations comparing good and bad examples. While she saw potential in the idea, she emphasized the need for clearer guidance. Before the next pilot, we should improve popup visibility, clarify ratings, and provide more detailed explanations if users want to click “Learn More”.

Bayan

- Bayan is currently a social media user but on the more cautious side since he doesn’t post too much (privacy reasons)
 - He fits the target audience well: people who use social media but are worried about how their data is being used
- Context Provided: you are looking to make an instagram account since your friends are on the platform, but are worried about the data privacy risks.
- Liked the popup since it wasn’t too intrusive and could be clicked off of if needed
 - However, since it didn’t regenerate if clicked off, that could lead to accidental click offs
- Liked the gold stars and ratings
 - Maybe add a word to summarize the rating (good, bad, horrible, etc.)
- Was curious about the LLM ratings
 - Noticed that the ratings changed between reloads of the page
 - We could cache the page ratings to keep them constant
 - it’s truly a prompt engineering issue though
- Noticed that the LLM summaries didn’t say too much, but liked that they were shorter than the actual privacy policy

SUMMARY:

During this pilot, the user was instructed to go to instagram and create a new account. They were also told they had an extension to provide data privacy information on new signups. Bayan navigated to the sign-up page then noticed the popup. He liked the popup’s layout and that the stars quickly gave him an idea of how good the app’s privacy policy was. However, he wasn’t sure how it compared to other sites or what exactly a % meant. Most of his questions were

regarding the LLM generated content on the popup. The descriptions felt similar to what would be on the actual privacy policy.