



Calibrating trust and crafting explanations

About this activity

In this activity, you'll imagine situations where users could under-trust or over-trust your AI feature. Then, you'll create a strategy for creating explanations and test these with potential users.

1. Brainstorm situations where there is no trust.
2. Brainstorm situations where trust exists, and is maintained.
3. Brainstorm situations where there is too much trust.
4. Using a 2x2 matrix, determine which user interactions require an explanation, and what kind.
5. Make sure that the explanation visuals and messaging make sense and are helpful for your users.



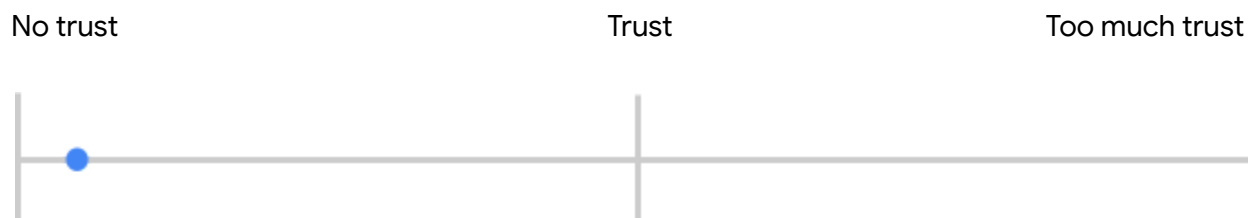
You may need to conduct additional UX research to answer some questions in this activity. Block out time to get as many cross-functional leads as possible together in a room to work through this activity, working through each question as a group. Then capture answers in the boxes provided. Review your answers as a team to determine what approaches your product will need to take to help users calibrate their trust.



1. No-Trust experiences and interactions

As a team, brainstorm what kinds of experiences and interactions would decrease trust in your feature's AI. Identify the underlying data sources, system data and user knowledge that could impact the calibration.

Example product: AI that classifies a skin condition.



User Group A

Example user group: Doctors of patients using the AI system

Example scenario: Patients see doctor with a condition that was mis-classified by the AI.

System data impacting calibration	<i>Example: image of current condition submitted by user, label of skin conditions in training data</i>
User knowledge impacting calibration	<i>Example: user's prior medical history</i>



User Group B

Example user group: Patients using the AI system

Example scenario: Patients see a doctor with a condition that was misclassified by the AI.

System data
impacting calibration

User knowledge
impacting calibration



2. Sustained and calibrated trust experiences and interactions

As a team, brainstorm what kinds of experiences and interactions would maintain or sustain trust in your feature's AI. Identify the underlying data sources, system data and user knowledge that could impact the calibration.

No trust

Trust

Too much trust



User Group A

Example user group: Patients using the AI system

Example scenario: Patient uses AI system to identify a common and temporary condition, like poison ivy, and receives a recommended treatment that works.

System data impacting calibration	
User knowledge impacting calibration	



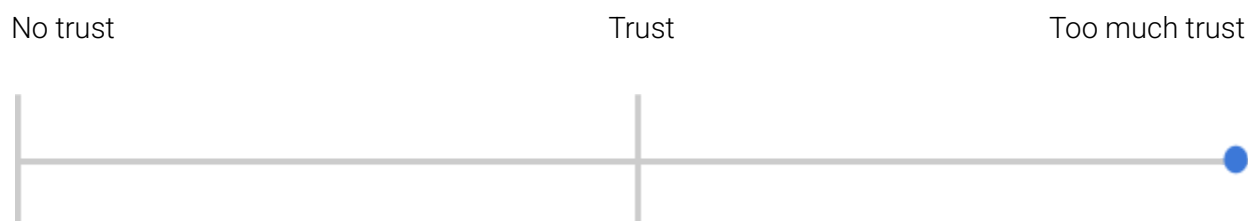
User Group B	
System data impacting calibration	
User knowledge impacting calibration	



3. Too much experiences and interactions

As a team, brainstorm what kinds of experiences and interactions would decrease, maintain, or inflate trust in your feature's AI. Identify the underlying data sources, system data and user knowledge that could impact the calibration.

Example product: AI that classifies a skin condition.



User Group A

Example user group: Patients using the AI system

Example scenario: Patients with pre-cancerous cells doesn't double check the app's diagnosis

System data impacting calibration	
User knowledge impacting calibration	



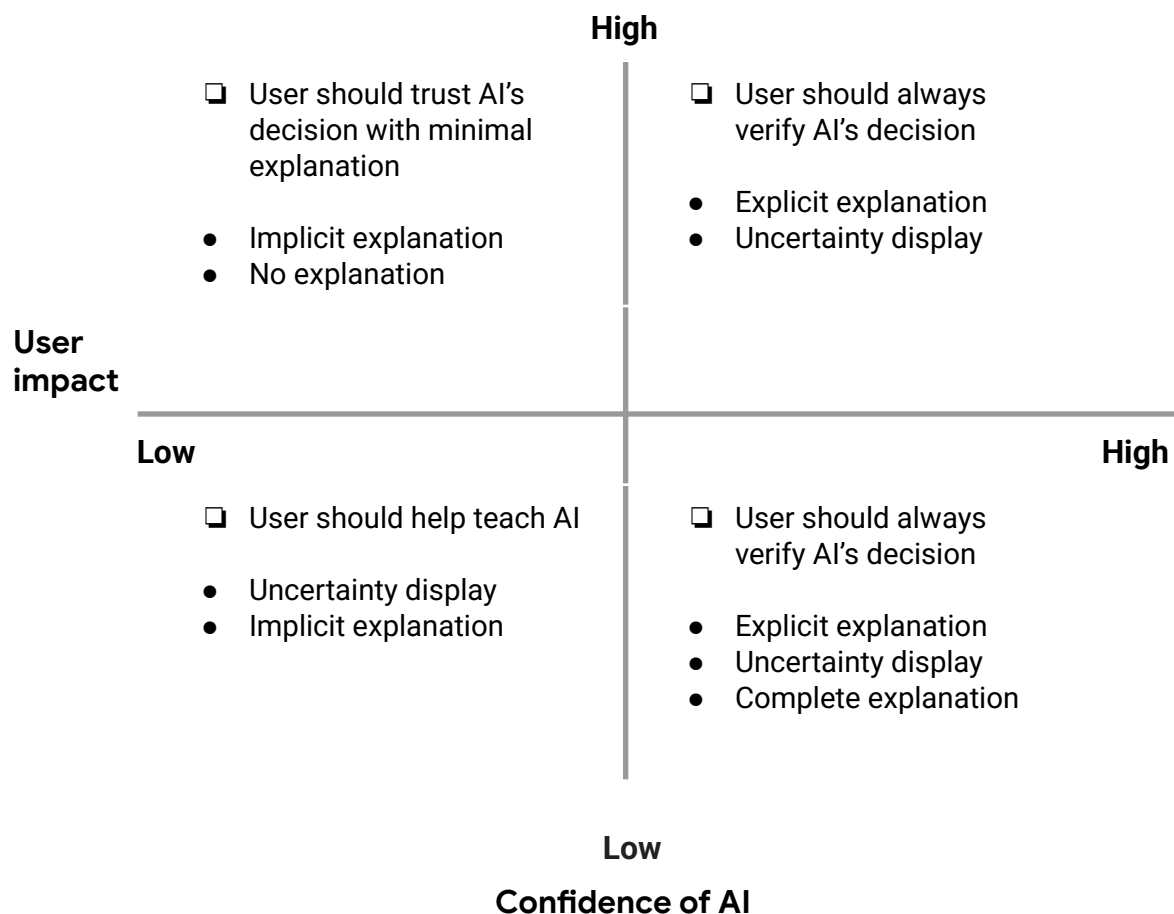
User Group B	
System data impacting calibration	
User knowledge impacting calibration	



4. Explanation strategy

After mapping the range of interactions that could decrease, maintain, or inflate trust, decide which interactions require an explanation. Use the 2x2 template of “User impact” vs.

“Confidence of AI” to help narrow down which explanations you want to test with users. Once you have consensus on the interactions that require an explanation, use the templates below to draft the explanation copy for user testing.





Explanation messaging

Product interaction	
Users	
Technical understanding	☐ expert ☐ non-expert
Explanation type	
Explanation text	



Messaging templates

Explicit explanation

Example:

"You are seeing this video recommendation because you often watch cooking videos."

"This is most likely ____X____, because ____Y____."

Displays of uncertainty

Example: "Prediction: [category] XX%"

Alternatives & most likely outcomes

Example: "Most likely X, Y, or Z"



5. Test with user research

Use the explanations you have drafted in user research. Multiple research efforts will be needed to understand any trust concerns users anticipate, and arrive at the ideal explanations for your user groups.

Research protocol questions

- On this scale, show me how trusting you are of this recommendation. [show scale]
- What questions do you have about how [name of product / feature] came to this recommendation?
- What, if anything, would increase your trust in this recommendation?
- How satisfied or dissatisfied are you with the explanation written here?

Once you're through with the above, ask additional questions to gauge user understanding of your specific AI system and the actions users can take.