# Investigative Proposal for AI Trust Issues in Users


PROPOSED BY ROHAN SOMJI

# THE PROJECT

ROHAN SOMJI

Consider the paradigm: **More power to the user**, Empowering users by integrating AIs should be about making the AI something the user can rely upon in a plethora of practical user contexts.

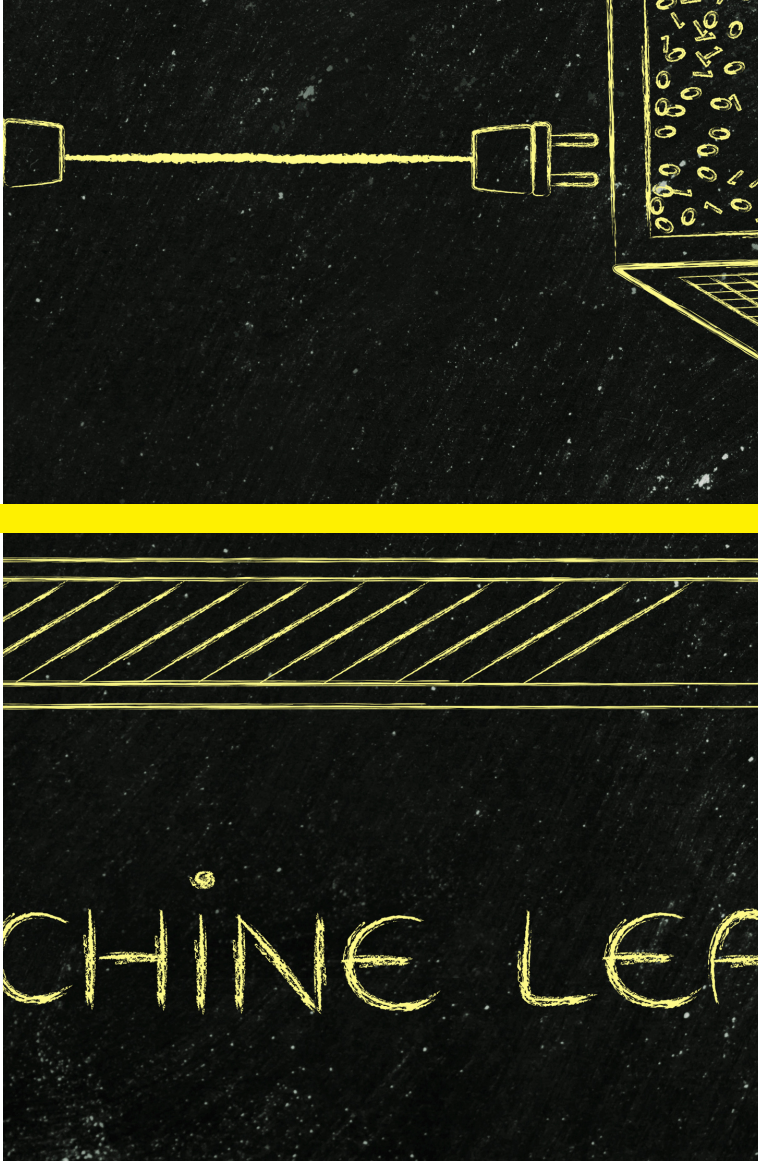
# THE FACTORS THAT SHAPE THEIR EXPERIENCE

**We first need to understand what factors shape a user's experience of AI?** Specifically, we want to understand what are some of the factors which affect people's experience of trust in AI?

Prior research has explored: **"What is the role of trust and reliability in automation?"** (Desai et al., 2009; Lee & See, 2004) But these questions have been asked in robotic automation and not digital ML-based AI automation.

With regards to ML-based AI, researchers have focused on XAI (Explainable AI) and how it can improve trust in AI by displaying to the user how the AI got its results. **XAI** hastily offered as the solution, **presumes a lack of transparency as the root cause of the user's trust issues with A**I. (Du et al., 2019, p. 21:2; Wang, et al., 2019, p. 2; Yang, 2020; Yang, et al., 2020;)

The second question to address must be: **What factors other than lack of total transparency affect user's trust in AI systems?** AI-specific trust issues may also arise from cultural depictions in fiction and nonfiction.




We may need to account for such trust issues in our AI user interfaces just as much as transparency issues. **In what ways can the user interface of AI systems help the user trust the AI system?**

# PRIOR WORK: ROBOTICS & XAI

This is equivalent to what Steinfeld has done in robotics; by developing **metrics for trust and reliability in automation** he was able to gauge how users operate physical robots (Steinfeld, 2009, pp. 4). Such studies are available in robotics automation but not many in digital automation.

When it comes to building trust, XAI (explainable AI) has become the major approach. **The shortcomings of XAI** are that while it is possible to approximate a set of human-understandable rules, no such approximation will ever map onto the actual functioning of deep neural networks that are the core of AI (Elton, 2020). Elton calls for a self-explaining AI approach. Whether self-explaining AI is better than explainable AI or interpretable AI is not important. What is important is the fact that the possibility of alternative strategies suggests that we still don't understand user trust issues in AI and ...

... **having chalked out transparency as an issue on the basis that AI systems are complex in function does not mean that no other factors exist.**

Ultimately the reason we construct any of these approaches is to improve trust and reliability on AI systems and we must investigate if there are any other factors as well.

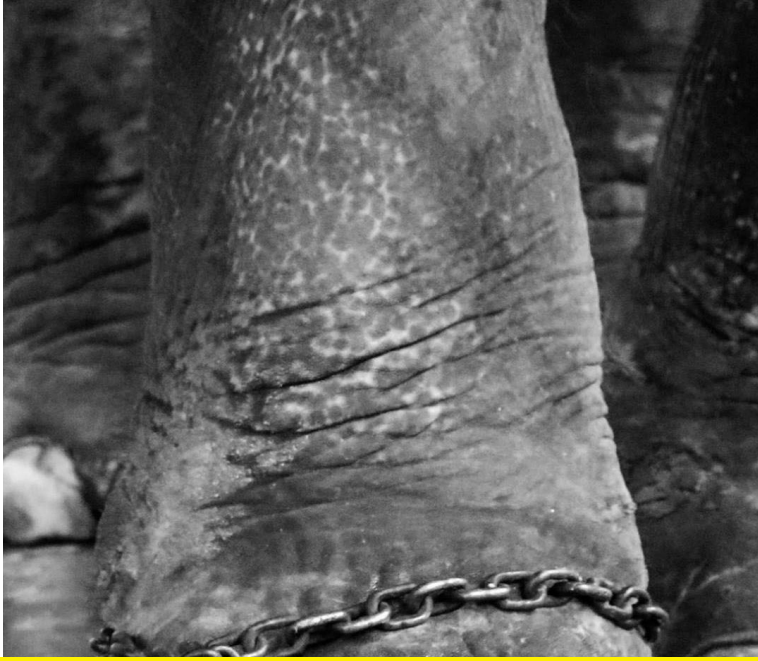
# THREE METHODS

Since this entire venture is user-centric, it is vital that we conduct **a survey with preliminary questionnaires to capture the mass sentiment.** These survey questions will later act as seeds to generate questions for in-depth interviews. Ultimately it is in the **one-on-one interviews**, where we will learn the details of the experience of trust in AI. These surveys should have a **Likert scale** with germinal questions that can later be probed to develop deeper questions (in the interview stage) based on how responders respond to the surveys. For example, a survey question might say "I think google answers my questions well?" the options being: strongly disagree, disagree, neither agree nor disagree, agree and strongly agree.

Based on what the answer is, we can later develop interview questions that can yield more qualitative data: **"Help me understand the process you go through when searching a query on Google?".**

Alternatively, we could also organize a f**ocus group** as a way to brainstorm ideas in the early stages; ideas about other factors apart from transparency that affect trust in AI (Breen, 2006, p. 465). We could use the themes that emerge from these focus groups to then generate **extensive questionnaires** to understand if certain sentiments are universal.

For example, an idea that
“AI does not actually understand what we want!” could be a focus group
a statement that could later translate into a question “Google search does not understand what I am searching for?” on a Likert scale from strongly agree to strongly disagree.



# LIMITATIONS

A potential issue with using these methods is that **if we miss good questions in the survey round or good prompts in the focus group we might miss a larger wave of sentiment** about AI in general and rather end up fixating on specific issues. The same applies to a focus group that begins fixating on one single issue in AI. It could stall idea generation in brainstorming sessions and leave us with only one theme to explore (like everyone talking about not wanting to be dictated by algorithms!). Moreover, the Likert scale itself is open to 2 well-known biases -

1. **Acquiescence bias:** a tendency to agree or select a positive response option (Baron-Epel et al., 2010)
2. **Social-desirability bias:** a tendency to go with the socially accepted position (Krosnick, 1999)

Both these biases mean that questions have to be carefully worded and provided balanced scales i.e. equal ratio of positive and negatively worded questions (Experience, W. L., 2020) But despite that, the Likert scale is still left at the mercy of the questionnaire maker and will impact the analysis of emergent themes.

# RECOMMENDATION

**We should prefer interviews and focus groups for generating ideas and themes as an inductive Emic approach is best suited for exploratory research.**

As we do not already have a theoretical framework about what factors other than transparency could be impacting the user's experience of trust in AI, we must rely on meanings that emerge from the field (as opposed to a more deductive Etic approach) (Tracy, 2019 pp. 27).

# DISCOVERY IS KEY

Findings from this research will provide us multiple factors in addition to lack of transparency, that are responsible for deficient trust in AI. XAI has been a user-centric approach but recognizing every factor before arriving at a UX solution to AI is, putting the user in the center of the design. If AI is to empower, then the user's problems, both conscious and unconscious, must be accounted for in AI-UX designs.

# APPENDIX

**Du, F., Plaisant, C., Spring, N., Crowley, K., & Shneiderman, B. (2019). EventAction: A visual analytics approach to explainable recommendation for event sequences. ACM Transactions on Interactive Intelligent Systems (TiiS), 9(4), 1–31.**

This argument does have some merit as racial bias in algorithms, poor medical decisions by recommender systems, etc have been known to occur. But Tiktok does not use XAI and yet it has a growing user base which means there are more strategies left to explore that can generate trust in user and XAI is not the only way.

**Elton, D. (2020). Self-explaining AI as an alternative to interpretable AI.**

Transparency is one approach to improve trust but consistent results is also an equally powerful method. Elton provides an alternative to XAI called Self Explaining AI where the AI system explains its process using neural networks. This points to approaches common to visual display of explanations in XAI format.

**Lee, J. D., & See, K. A. (2016). Trust in Automation: Designing for Appropriate Reliance: Human Factors. https://doi.org/10.1518/hfes.46.1.50_30392**

Lee and See argue for a deep relation between trust and reliability and go into much detail referring to communications studies. Steinfel relies on their concepts to explain automation reliability and automation capability in robotics.

We know for example that within robotics cute looking robots receive a better response (see Kate Darling's work at MIT media labs)

I want to understand how does the experience of a user change when they are aware of an AI system behind a UI and how do cultural dispositions play a role in these experiences?

**Tracy, S. J. (2019). Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/georgetown/detail.action?docID=5847435**

But an entirely different approach could be complete participant observation, where the researcher will monitor users while they are operating an AI/ML-based application like a search engine and generate themes based on their own field notes of what they observed the user do. In the end, the user could be provided a semantic differential scale-based questionnaire and their answers would provide the quantitative data that the researcher can later compare with their own field notes to draw out issues related to trust.

# REFERENCES


Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. European Journal of Public Health, 20(5), 543–548. https://doi.org/10.1093/eurpub/ckq052

Breen, R. L. (2006). A Practical Guide to Focus-Group Research. Journal of Geography in Higher Education, 30(3), 463–475. https://doi.org/10.1080/03098260600927575

Desai, M., Stubbs, K., Steinfeld, A., & Yanco, H. (2009). Creating trustworthy robots: Lessons and inspirations from automated systems.

Du, F., Plaisant, C., Spring, N., Crowley, K., & Shneiderman, B. (2019). EventAction: A visual analytics approach to explainable recommendation for event sequences. ACM Transactions on Interactive Intelligent Systems (TiiS), 9(4), 1–31.

Elton, D. (2020). Self-explaining AI as an alternative to interpretable AI.

Experience, W. L. in R.-B. U. (n.d.). Rating Scales in UX Research: Likert or Semantic Differential? Nielsen Norman Group. Retrieved December 1, 2020, from https://www.nngroup.com/articles/rating-scales/

Krosnick, J. A. (1999). Survey research. Annual Review of Psychology, 50(1), 537–567. https://doi.org/10.1146/annurev.psych.50.1.537

Lee, J. D., & See, K. A. (2016). Trust in Automation: Designing for Appropriate Reliance: Human Factors. https://doi.org/10.1518/hfes.46.1.50_30392

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, 33–40.

Tracy, S. J. (2019). Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/georgetown/detail.action?docID=5847435

Wang, D., Yang, Q., Abdul, A., & Lim, B. (2019). Designing Theory-Driven User-Centric Explainable AI. https://doi.org/10.1145/3290605.3300831

Yang, Q. (2020). Pro ling Artifcial Intelligence as a Material for User Experience Design. 126.
Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. https://doi.org/10.1145/3313831.3376301