User-centered Design in AI for Social Impact

Herman Saksono[1,2][0000−0002−1148−7534]

[1] Northeastern University, Boston MA 02115, USA h.saksono@northeastern.edu
[2] Harvard University, Cambridge MA 02134, USA

Abstract. By incorporating user-centered design into the AI for Social Good pipeline, we ensure that the AI solutions will ultimately be impactful and usable. This chapter touches on the basic approaches of user-centered design, namely interviews, paper prototyping, and quantitative experiments. These tutorials are the starting point to conduct more studies about the impact of technology on people and society.

Keywords: User-centered design · Human-Computer Interaction.

1 User-Centered Design for Social Impact

During the I\O 2018 developer conference, Google announced their latest Artificial Intelligence (AI) innovation. The deep learning AI could diagnose diabetic retinopathy using the patient's retinal scans with more than 90% sensitivity and specificity [13]. Google described that the AI could detect diabetic retinopathy early on and at scale, thus preventative measures could be taken, especially in places where access to healthcare is constrained [23].

A few months later, Google ran a field trial in health clinics across Thailand. About 7,800 miles away from Google's headquarter in California, Thailand is a middle-income country in mainland Southeast Asia. Beede et al. led qualitative interviews and observations during the field trial [2]. They found that the health clinics had less reliable retinal photography processes and less reliable internet connections. As a result, 21% of the patients' retinal images cannot be processed by the AI. Furthermore, the less reliable internet also reduced the diagnosis capacity. When a clinic lost its internet connection for two hours, the AI-supported diagnosis capacity dropped from 200 patients per day to 100.

This is not a story to undermine AI. Rather, this is a story about the many sociotechnical work we must do to create AI for social impact. Of course, such work could be delegated to people who do field studies. However, not engaging at the sociotechnical level would be a missed opportunity for AI developers. Developers are experts in the opportunities and limitations that AI could provide. By engaging with the social complexities of technologies, developers could access their expertise to address the problem. User-centered design is one more toolkit to ensure that technological solutions are both useful and usable in our social world.

Useful, as in, findings ways demonstrate how an AI tool will be concretely beneficial to humans. Then, *usable*, by making sure that people can use the AI tool to produce useful outcomes.

From a user-centered perspective, we use both the model's performance and positive human outcomes to evaluate the effectiveness of our AI tools. More concretely, if we are developing an AI to help a medical diagnosis, we need to demonstrate that humans will make accurate diagnoses while using the tool. Although the machine learning model performs well, the model will not be used in isolation. Instead, humans' personal and social factors play a large role in determining the effectiveness of AI.

An example about how personal and social factors play a role in AI effectiveness is the studies by Jacobs et al. in understanding how clinicians would use an AI to select the most appropriate antidepressant treatment. Through a quantitative study, they show how incorrect AI recommendations led to lowered accuracy even when explanations are provided [17]. In other words, when the AI gives inaccurate recommendation, clincians could end up prescribing wrong treatments.

Jacobs et al. also conducted a qualitative interview study, and they found clinicians often had a limited time with the patients [16]. As a result, the clinicians did not have enough time to determine whether they can trust the AI predictions. This creates a conundrum where incorrect AI predictions could lead to incorrect clinical treatments, while at the same time AI explanations could not address this issue because of a larger issue with the healthcare system.

This example and Google's example show that models that perform well could be less effective when facing sociotechnical barriers.

With user-centered design, the approach is reversed. We first study the personal to socio-technical complexity of the problem, build prototypes, develop the most appropriate technological solution, and finally evaluate the tool's effectiveness in addressing the human problem. This approach saves resources (e.g., time, funding, person-hour). If we build an AI tool without fully understanding how the tool will be used by people, then there is a chance that the tool will not be effective enough. An ineffective tool would be wasteful of resources and the stakeholders' time.

User-centered design could tell us that the AI tool for detecting diabetic retinopathy should be made for retinal scans with more imperfections. Or perhaps, the solution is not just optimizing AI diagnosis, but using AI to improve access to healthcare. Thus, it is important to keep in mind that by getting a better understanding of people's thoughts and actions, we will end up with solutions that look completely different from the first solution we had in mind.

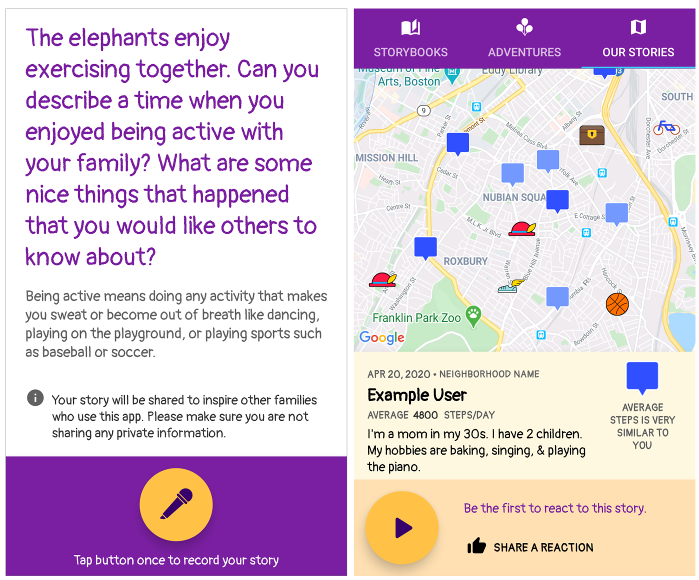
For instance, Saksono et al. conducted interviews on how fitness sensors would be used by families of low-income neighborhoods in the United States who typically face more crime concerns [31]. Indeed, these crime concerns are real. Some families could not be active outside to meet their step goal because they could be fatally shot. For example, one mother said:

> I ignored [the reminder from Fitbit]. I'm like okay. When [the app said] like, "*You have, like, 244 more steps. Okay, let's see if I can walk in the house*" but as far as getting out outside, that would be my last walk.

As a result, the fitness sensors could end up being useless. However, the study also found that the notion of neighborhood safety is not uniform. Some places were seen as safe and others were seen as less safe. These feelings of “safeness” are the byproduct of frequent and consistent social interactions among the people in the neighborhood.

Informed by this finding, instead of modifying the fitness sensors, Saksono et al. built a mobile app called Storywell that shows other families who use the fitness sensors on the neighborhood map (Fig. 1) [32]. The app also allows families to interact with other families virtually using audio stories.

Fig. 1. Screenshot of the story map in Storywell [32]. Users can record their healthy living stories (left) then share them on the map so that people can learn from each other (right).



The example above shows being open to different kinds of solutions could help to create innovative solutions for addressing human problems.

This chapter on user-centered design is specifically designed to help AI developers to create solutions that address human problems. Specifically, by equipping AI developers with practical methods for understanding people’s thoughts and actions, and also how to understand the socio-technical complexities of deploying technological solutions. The goal is to make it easier for those who just started conducting user-centered design.

The methods in this chapter include tutorials for in-depth qualitative interviews, paper-prototyping, and quantitative experiments. There is a subsection on ethics and theories about human action. These are not designed to bore, but rather a set of intellectual tools to articulate and think about sociotechnical solutions.

As you progress through this chapter you will notice that the methods are quite familiar. Indeed, user-centered design methods are well-established research methods. Being familiar with methods will make it easier for you to adopt the user-centered design. Let's start with the most accessible method that everyone has done at least once in their life: interviews.

2 Tutorial: In-depth Interviews for Needfindings

Asking the right questions is a window to understand a world that we did not understand before. By doing interviews, we can recognize AI users' thoughts and actions. We can also learn the social and organizational processes where the AI will be deployed. By understanding these complexities, AI developers could design tools that are aligned with the target users' thoughts and practices, as well as the social world in which they operate.

The end goal of interviews is to get the rich data for justifying our AI designs. In interviews, the data is the interview recordings or transcripts. They should tell us about people's experiences and thoughts. For example, from interviews we can learn that fieldworkers who will use the AI are often rushed to make the decisions. We can also learn about the workplace situations that led to rushed decisions (e.g., time-sensitive projects, the large number of tasks that needs to be attended to). By learning the situations where the AI will be used, we can develop AI solutions that works optimally in the real-world.

Interview data is rich because we can ask about *what* people prefer as well as *why* they prefer that. For instance, we can ask people what kind of predictions are useful for them and why they prefer those predictions. The "what" questions tell us what kind of prediction tools are useful. Meanwhile, the "why" questions will tell us the real-world complexities that could make or break the prediction tools.

Another benefit of interviews is that we would be able to get a multivocality of perspectives, which will help make the AI tool more resilient. For instance, the director of a non-profit would want an AI to identify the poverty level at the neighborhood level using their internal data. The field workers of the non-profit would be happy to help support the AI development, but they also know that identifying the poverty levels might not be useful in helping their work. They might also tell you the problems with incomplete data. Finally, the beneficiaries of the non-profit might be excited about AI. Still, they might tell you that they have so many strengths that were overlooked by the non-profit (e.g., youth organizations, community leaders, local businesses).

By interviewing people across all roles, we will better understand the ways the AI will be beneficial and in what ways it will break. In turn, we will be able to develop a more resilient AI because we are leveraging the strengths of organizations and the community while at the same time mitigating the complex barriers.

From these multivocality of perspectives, you can also identify ethical issues in the AI tools. The domain experts (e.g., stakeholders, topic experts, field-workers, community members in which the AI will be deployed,) will be able to pinpoint how the tool could be harmful. Therefore, by conducting interviews with the people who will be impacted by the AI, we will have the opportunity to address ethical issues early on during the development process. This approach put less stress on the developers compared to findings issues emerge during deployment after much time and efforts was spent during development.

Next, we will discuss the steps to conduct in-depth interviews for need-finding.

2.1 Develop the research questions

Begin by developing a broad research question, then refine it. It is important to develop strong research questions because they will guide you throughout your interviews. The research questions should be broad enough to allow unexpected findings. On the other hand, the research question should not be too broad and causes the researchers to miss the ultimate goal of the interviews. Examples of a research question for an in-depth interview, include:

How do fieldworkers decide ...?
How do the organization and the community support fieldworkers' decision making for ...?

Interviewing, and qualitative research generally, is an intuitive process. Often, the goal is not to confirm certain technology design requirements, but also to learn requirements that we have not known before. Therefore, it is valuable to approach interviews with open and curious attitudes. Hyper focusing interviews with a specific set of hypotheses will bias us away from learning phenomena that could make or break our AI tools.

2.2 Identify the stakeholders to be interviewed

Begin by listing the people who will be affected by the AI tools. This may include people within the structure of the organization (e.g., directors, middle-managers, field workers) as well as people outside the organization (e.g., clients, beneficiaries, local leaders, community members, domain experts).

The next step is identifying which of the roles will be interviewed. The goal here is to balance the depth and multivocality of insights, versus the project's constraints such as time, human resources, and funding. We want to optimize the interviews to get as many insights as possible while ensuring that we can do the interviews within the constraints.

Once you identify the roles to be interviewed, the last step is to concretely determine the people of those roles or where you could recruit participants for those roles. This should be followed by recruiting and scheduling the interviews.

Important: Use intersectionality for identifying the roles to be interviewed. In intersectionality, people can have multiple marginalized identities and these identities compound to create distinct forms of oppression [12, 25]. Consequently, when an AI tool is harmful, people with different intersecting marginalized identities could experience harm in different ways. These marginalized identities include gender, race, disability, and socioeconomic status.

For example, white women who are harmed by technology would have a different harmed experience than Black women, because Black Women could be oppressed for being Black and women. Thus, interviewing Black women will enable them to share different ways in which a technological solution can be harmful.

Tip: Consider doing multiple interviews. There are two reasons for having more than one interview. First, some people might not be comfortable sharing their views with a complete stranger. Thus, multiple interviews are an opportunity to earn trust and rapport. Second, we often learn the salience of unexpected insights after interviewing several participants. Thus, it is beneficial to interview prior participants and check whether they agree or disagree with the insights. Conducting two interviews with the same person in a one-month interval, for example, will allow some space to confirm insights that you have identified.

2.3 Develop the interview guide

Appendix 1 shows the structure of a typical interview. An interview begins with the introduction of the interviewers and an overview of the interview's purpose. This is followed by the interviewer emphasizing that the interview data will be anonymized and interview participants can stop their participation at any time.

Next, the interviewers would continue by asking questions about the participants' background and their knowledge of the organization and the community. As the interviewer asks the questions, use this opportunity to build rapport and develop participants' trust.

Then, the interviewers could ask open-ended questions and follow-up questions about the participants' thoughts and experiences with the existing or current solutions (technological and non-technological). The goal is to identify in what ways an AI solution could solve a current problem and also how an AI solution would not work.

Since the goal of interviews is to understand the "why", it is a good practice to ask open-ended questions to get answers that you would not expect. In many cases, we also get important insights through the follow-up questions.

2.4 Pilot the interview guide

It is desirable to practice your interview questions with someone you are comfortable with. The goal is to identify issues with your interview guide and also to make sure you can comfortably carry out the interview.

2.5 Conduct interviews and do observations

Now that you have the interview guide ready and the interview scheduled, you can start conducting the interviews. Your goal should be to collect as much relevant data as possible, in the form of interview recordings and fieldnotes. Eventually, you will use the data to back your AI design decisions.

Some interview participants will be eager to share their thoughts and experiences, and some will be more reserved. There are several approaches that you can use to make sure you learn as much as possible in a short 1-hour interview. There are some strategies to conduct interviews effectively [22, 24, 7]. We will distill these strategies below:

1. **Explain the goals of the interviews clearly.** Communicating the interview goals is a way to earn trust and build comfortable interactions, which in turn help the interview participants to share their thoughts and experiences.
2. **Indicate that participation is voluntary**. There is often a power imbalance between the interviewer and the participant. Therefore, equalizing the power imbalances would help facilitate comfortable interactions. One approach to equalizing power is emphasizing that the people's participation is voluntary. That means participants can choose to stop their participation any time they want. By stating so explicitly, the interviewer gives more power to the participants.
3. **Paraphrase participants' statements**. Paraphrasing shows that you have been actively listening to their stories genuinely. It's also an indirect way to invite participants to share more stories. But, more importantly, paraphrasing helps the participant's to verify the interviewer's understanding. If the paraphrasing is inaccurate, the participant can clarify it immediately.
   Paraphrasing can be in the form of saying "It sounds like ..." or "I think you're saying that ..." [24]. For example, when a participant explained how it was hard for them to do something because of organizational constraints, the interviewer could ask, "It sounds like you already know what you need to do, but your organization's rules made it difficult for you to act?"
   By asking such questions above, we are inviting participants to confirm and elaborate their stories. Sometimes, the interviewer misunderstood what the participants said, and paraphrasing the participants' statement gives them the opportunity to correct the interviewers' understanding.
4. **Clarifying participants' statements**. Sometimes our understanding of the participants' stories might be different from the participants. If you are uncertain, even slightly, it is a good practice to ask for clarifications.
   Clarification questions can function in two ways. First, clarification questions invite participants to give concrete examples. Examples are useful because they gave us concrete descriptions of what happened. In turn, it helps the interviewer to understand participants' thoughts and actions as well as the participants' perspectives. One example is, "You said that making decisions is hard because of the rules of your organization. Can you give me an example of that?"

Second, clarification questions invite participants to share their interpretations of the examples that they have shared. By knowing how people evaluate events, we will be able to know the meaning of the event. One example is, "You mentioned the challenges you had when making decisions, I was wondering how you feel about having those challenges?"

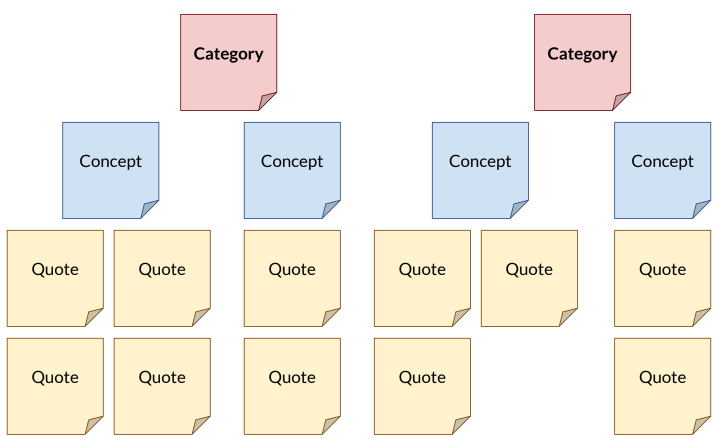
5. **Summarizing participant statements.** Summarizing is the final tool to understand people's thoughts and actions. It should be regarded as a final tool because we do not want to box the participants' statements prematurely. Typically, once you have conducted multiple interviews and identified important insights, then summarizing a participant's statement is helpful to define the key insights. Examples of summarizing statements include, "Based on what I learned from you, making decisions is hard mainly because of the rules in your organization. Is that accurate — or not accurate — based on your experience?"

Notice how we used tentative languages in the examples for paraphrasing, clarifying, and summarizing above. Tentative languages are often preferable because they are not leading the participants to give a certain answer. Asking open-ended questions, providing a range of answers (e.g., accurate or not accurate), and using tentative words (e.g., it seems, it sounds) are strategies to avoid the risk of leading the interview participants.

2.6 Analyze the interview transcripts as soon as possible

There are many qualitative analysis methods, such as Thematic Analysis [5] and Grounded Theory [11]. For this chapter, we will use a simpler form of qualitative analysis called Affinity Diagramming (Fig. 2).

Fig. 2. Structure of an affinity diagram. Quotes build concepts and concepts build categories.



Affinity diagramming is a method to structure interview quotes. The goal is to identify important stories from qualitative data [28]. This is an inductive process, the structure is not preconceived but rather systematically developed from data.

Remember that qualitative analysis is an art and science [28]. It is an art because you generate insights intuitively from data, which requires you to trust your instincts. It is also a scientific process because the insights must be grounded on the data, in this case, interview data.

Qualitative analysis is also a co-production knowledge with the interview participants. Your technical expertise in computer science, math, machine learning, and artificial intelligence will complement the domain knowledge that the interview participants share during the interviews.

The steps to do affinity diagramming is as follow:

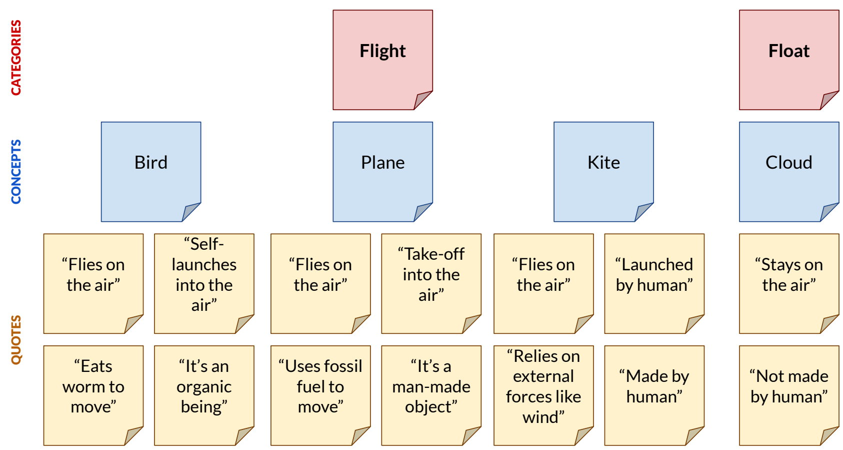
1. **Get index cards (or sticky notes)**. Then, find an empty wall or table for you and your team to lay out the cards. Affinity diagramming will happen for as long as the interview process, which means you will need to leave the cards for the duration of the study.
2. **Write down each important interview quote into a quote card** (yellow).
3. **Invite the development team to cluster similar quotes together.** The goal of having several people performing affinity diagramming analysis is not to find a single truth about the participants' experiences. Rather, the goal is to make sure the analysis captures multiple perspectives.
4. **Group similar quotes under a concept card** (blue). Write a short name on each comment card. The name should be an abstraction of the underlying quotes. By abstracting a set of similar yet slightly different quotes under one concept, you will be able to compare concepts. Eventually the concepts should be complete. When you compare two complete concepts, there would be little (if no) overlapping ideas between the two.
5. **Write memos about your concepts.** Memos are extremely valuable for three reasons. First, it is a way to materialize your thinking in writing. Second, it is an audit trail of the concept, so that months later you can always go back and retrace your decisions to structure the quotes and concepts. Third, it is an opportunity for you to ask questions based on what you observed from the interview data. From these questions, you can seek answers the in the next interviews

   Note: You may have low-level and high level concepts. Thus there can be more levels in your affinity diagram.
6. **Group similar concepts under a category card** (red). Write a short name for each category. This name should be a further abstraction of the underlying concepts. A category is useful when you can develop causal relationships between categories. Write memos to record your interpretation of each category as well as how each category relates to other categories.

   Sometimes you need to group multiple sub-concepts under one higher-level concept. You should do this when the data suggest it is necessary.

Fig 3 shows how an affinity diagram could look like, as adapted from Corbin and Strauss qualitative analysis method [11, p. 76].

Fig. 3. An example of an affinity diagram. Suppose that the analyst is an extraterrestrial being who has never seen a bird, a plane, and a kite. However, the analyst has access to quotes about what people see on the air. From the quotes, the analyst identified three concepts: bird, plane, and kite. Then, the three concepts were grouped under a category called flight. A separate category called float describes a concept called cloud that is different from bird, plane, and kite.



Important: Analyze the interview data as soon as you get your first interview transcript. This means you should do steps 2-4 as soon as you receive an interview transcript. Unlike quantitative research where you do the analysis after all of the data is collected, effective qualitative analysis is done by analyzing the data as soon as you receive it. The purpose of this approach is to build preliminary insights as soon as possible so that you can ask the next interview participants about the insights.

In a rigorous qualitative analysis, the concepts and categories should be grounded on the interview quotes. Biased analysis means the memos do not reflect what the data says. A good way to validate your analysis is by comparing the memo of a concept with the quotes under the code. If the concept's memo and the quotes do not fit together, there are two approaches that you can do. First, you can modify the concept's name and the memo. Second, you can move the unfit quotes under a different concept.

2.7 Produce the report

Once you have finished the analysis and checked whether every insight is complete, the next step is writing the report. The report should tell a story about how the AI solution will be useful. Specifically by explicating existing processes, opportunities, and barriers.

It is important for the report to tell a story about the relationships between categories. After that, the report should provide sufficient details about the codes and concepts that make up a category. A report that lists facilitators and barriers is generally not considered to be a desirable qualitative report [6].

Below is an example report of a concept that was identified by Saksono et al.'s evaluation of the a digital neighborhood map that shows families who use fitness sensors [32]:

> *Normative behavioral information* is a piece of information about the kind of behavior that is valued by one's peers (n=7). Caregivers learned this information by observing other families' behavioral practices (e.g., from the stories from other caregivers). In turn, when caregivers learned that their internal standards (i.e., what the caregivers value) align with the social standards (i.e., what their peers' value), caregivers felt validations of their behavior. For example, P12 explained why hearing other families' stories felt good:
>
> > **P12**: Just hearing the stories, you were just like, "Oh, okay. So, I'm not the only one." [. . .] When you sit back and you're like, "Okay, I'm not the only one doing this. I'm not the only one trying to find ways to keep my kids busy and happy." It's good. It works out.

The example above begins with a topic sentence explaining the concept (i.e., normative behavioral information). Then, the topic sentence is followed by additional explanations about the concept. Finally, the first paragraph is concluded with a sentence pointing to an example quote followed by the quote.

A good qualitative analysis should show the conditions that facilitate a phenomenon as well as the inhibiting exceptions. Exceptions should be explained after the main phenomenon has been explained. Your reader could get confused if you prematurely explain an exception while the reader has yet fully understood the main idea.

Another trait of a good qualitative analysis is when it shows the multivocality of people's experiences. Human's life experiences are rarely a monolith. Thus, analyses that show how people are reacting differently to a same (or similar) situation will demonstrate the strength of the interview and the analysis.

3 Tutorial: Paper Prototyping and Lo-Fi Prototyping

When building paper prototypes, you are creating the envisioned AI tool on a piece of paper (Rettig 1994). After that, you share the paper prototype with the stakeholders and the beneficiaries to get their feedback. Like other throwaway prototypes, it is meant for testing your ideas and getting feedback as early as possible. The goal is to avoid the risk of producing AI tools that are not useful. Rettig describes some of the benefits of paper prototyping [27], which include:

1. Developers will spend less time building a complete tool that could end up being useless. A single issue in a complete AI tool could mean the tool must be reworked from ground up.
2. Using easily-made paper prototypes encourage stakeholders to share more feedback on the key functionalities of the envisioned tool. In contrast, sharing a more complete and polished tool suggests that changing the tool is more complicated. This perception could inhibit the stakeholders to pinpoint fundamental issues that could render the tool useless. Instead, the stakeholders could end up being more focused on the color, layout, and the fonts of the design.
3. Throwing away a paper prototype design that doesn't work is easier than throwing away a more complete tool where a lot of resources have been poured into.

For instance, Saksono and Parker were developing an app for facilitating family conversations around fitness sensing data [33]. The goal was to encourage families to reflect on their fitness data and promote physical activity. They envisioned the app to include digital storybooks with prompts for supporting health reflections. However, this kind of approach has not never done before. It was unknown what is the benefit of such an app. Therefore, spending resources to build a fully-fledged app then testing its benefits is risky: the app could be completely useless.

Instead, Saksono and Parker evaluated a paper prototype of the idea with 13 families. They created a paper prototype in the form of a paper storybook (Fig. 4).

In the storybook, they put animal stickers on the pages. Then, they evaluated the paper prototype with 13 families with young children. They ask families to read the book together. If the child finds the animal sticker, they are asked to pick one of the "mystery" envelopes that contains a question for the family to answer together. In the interviews with the family who used the prototype, Saksono and Parker found that asking and answering health questions during the storybook reading encourage parents to be more aware of their family health behavior and identify opportunities to be healthy [33]. This finding solidified the decision to include storybooks and reflection prompts in the Storywell app (Fig. 1) [32].

There are two steps for conducting paper prototyping, which we will discuss below.

3.1 Develop the Paper Prototype

On a piece of paper, start drawing how the AI tool will look like. This is an opportunity to imagine how the AI will be used.

When developing a paper prototype, it is important to show the main functionality of the tool. The lines, drawing, and the writing does not have to be perfect or pretty. The predictions can be "simulated", meaning that you can use

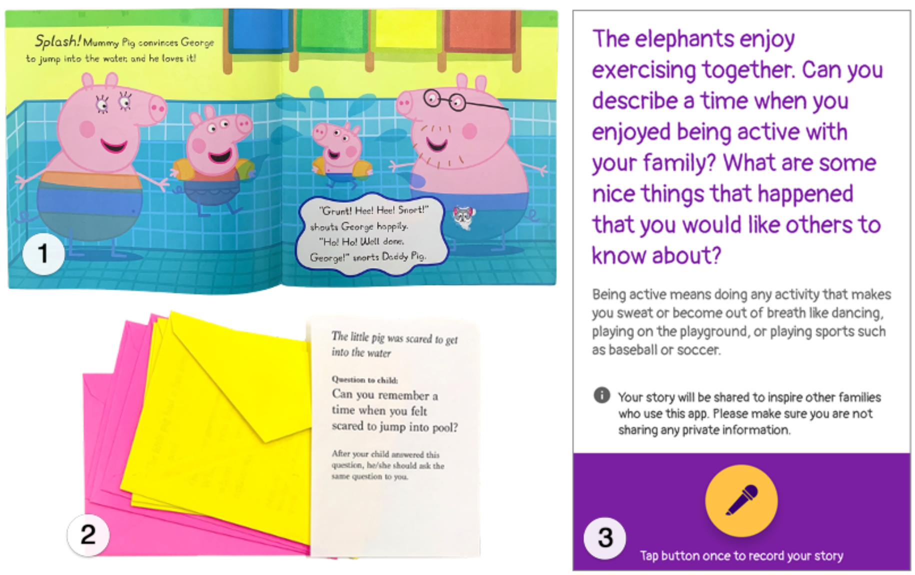


Fig. 4. Paper prototype in Saksono and Parker study [33]. The paper prototype consists of a paper storybook (1) and reflection questions (2). The goal is to emulate a digital app that invite families to reflect on their fitness data while reading a digital storybook. By evaluating the paper prototype, Saksono et al. can build the Storywell app (3) [32]

.

simulated data to show different cases of machine learning predictions in the AI tool.

For example, if you are developing an AI tool to help decision-making using a digital map, then the paper prototype would be a hand-drawn map on a piece of paper. The prediction of the AI could be a circular cutout of a paper with the confidence score on it, then you put the circular cutout on top of the paper.

If you are developing an AI tool to help clinical diagnosis, you can build a paper prototype that shows the confidence score of an ML prediction and the variables being used to explain the prediction. Jacobs et al. evaluated a digital low-fidelity prototype that uses simulated data to show the predictions and the features being used to make the predictions [16]. While the prototype is digital, it is relatively simple without the complexity of a complex prototype.

3.2 Evaluate the paper prototype

The rest of the paper prototype evaluation is very similar to conducting in-depth interviews. The only difference is you show the prototype to the interview participants and ask for their positive and negative feedback.

For instance, evaluating a paper prototype of an AI tool to help decision-making using a map, you can ask whether the map interface is useful for the stakeholders to achieve their goals. Then, you can ask in what ways the map is not useful. Therefore, you will get a broad understanding of the efficacy of the tool in solving human problems.

If you are evaluating an AI tool to help clinical diagnosis, you can ask whether the explanations that accompany the ML predictions are helpful and not helpful for the stakeholders in making clinical decisions. Then, you can examine organizational or structural barriers that inhibit effective clinical decisions if the AI tool is being used.

At the deeper level, you should do paper prototype evaluation if you have some ideas on how the AI tool would look like. Otherwise, to explore different kinds of solutions, you would conduct in-depth interviews first and develop the paper prototypes later.

When evaluating paper prototypes, it is useful to video-record the participants' interaction with the prototype (with the participant's consent). Observing the recordings will be useful to understand how a stakeholder would interact with the tool.

4 Tutorial: Quantitative Experiments

Quantitative experiments are best suited when we already have some understanding about the problems (exploratory) or we already develop a prototype that we believe will address the problem (confirmatory). In exploratory experiments, we want to test whether our AI tool prototype will be useful before we actually develop a more complete AI tool. In a confirmatory experiment we want to demonstrate whether a more complete AI tool will address the problem.

This approach is in contrast to in-depth interviews where we often approach the problem with a more blank slate.

This section will be focused on a randomized controlled trial approach of a quantitative experiment. Specifically, we will focus on AB testing. There are other approaches, such as cohort experiments, time-series experiments, factorial experiments, etc.

There are many uses of AB testing. First, you can use it to examine whether design A is better than design B, or vice versa. Second, you can use it to test whether an innovative solution (B) would perform better than a conventional solution (A) in supporting humans. For instance, suppose you want to test whether giving a 3D prediction map is more effective to help field workers to make their decision than a 2D prediction map. Then A would be a 2D map and B would be the 3D map. When you have figured out which design is better, then you can start allocating more resources to develop the AI tool.

Similarly, you can also test how users react in two situations. For instance, suppose that you want to test whether AI explanations are effective to prevent users from making inaccurate decisions when the AI is giving incorrect predictions [17]. Therefore, A would be AI with incorrect predictions and B would be AI with correct predictions.

In both examples, experiment participants will be randomly assigned to condition A or B.

4.1 Develop the research questions or hypothesis

At this point, you have developed sufficient insights into the problem and the personal, social, and organizational factors that influence the way users address the problem. You have also learned the domain knowledge of the problem. Finally, you have also synthesized a couple of ideas to help users tackle the problem.

However, you might be unsure about which approach to choose. Instead of spending time, funds, and resources to develop a complete AI tool with one approach with a risk that the approach is not useful, you can conduct an AB experiment to inform your decision. Once you have identified A and B, you can formulate a hypothesis. First, you will define a null-hypothesis first, then conduct an experiment to reject or accept the null hypothesis.

H_0: Approach A is just as effective as approach B for ....

Then the statistical tests we conduct at the end of the study will show whether H_0 can be rejected or not.

4.2 Identify the independent variable and develop the experiment's prototype

Independent variable is the variable that you manipulate during the experiment. This is the variable that you want to test. For instance, if you want to test the efficacy of 2D prediction map versus 3D prediction map, then you could develop a web app that shows 2D map or 3D map. Here, the independent variable is the map type (e.g., 2D or 3D).

In an AI explanation experiment, participants will also be randomly assigned to condition A and B. People in condition A will get incorrect AI predictions, whereas people in condition B will get correct AI predictions. Here, the independent variable is the AI correctness (e.g., incorrect v. correct prediction). Both participants will be asked to make several decisions based on a 'simulated' AI and get the same explanations.

Once you have identified the independent variables, you can start creating the prototype of the AI. The prototype can be a dummy tool with simulated data that are scripted to perform in a certain way during the experiment. It does not need to be a full-fledged AI.

4.3 Identify dependent variables and measures

Once you have identified the independent variables, you can continue identifying the dependent variables. These are the human-centric metric for measuring the AI's performance. They can be attitudinal or behavioral variables.

Attitudinal variables are the things in people's mind, typically measured using surveys. Such variables include the user's perceived usefulness, perceived ease of use, mood while using the tool, etc. For example, in the 2D v. 3D prediction map, you can use any of the aforementioned variables to assess people's perception towards the map.

Behavioral variables are about participants' actions. Such variables could include participants' accuracy, engagement, actions, etc. For example, in the effectiveness of AI explanations [17], the dependent variable would be the accuracy score. This is measured by dividing the number of times each participant made the correct decisions by the number of decisions they have to make.

Additionally, behavioral variables can also be in the form of surveys. These are called recall surveys, which ask people to remember their behaviors.

When using attitudinal surveys or behavioral recall surveys, you must use validated measures. Measuring human attitudes and behavior is very complex. You could end up measuring incorrect concepts if you do not use validated measures. Below are some examples of validated attitudinal and behavioral surveys:

1. Perceived usefulness (Technology Acceptance Model (TAM) [10])
2. Ease of use (TAM [10])
3. Predicted usage (TAM [10])
4. Mood (Self-Assessment Manikin (SAM) [4]).

Once you have identified the measures you can start putting the survey items into your survey software, such as Qualtrics, SurveyMonkey, or Google Forms.

You should include demographic and socioeconomic status surveys in your study, because they will help you explain the experiments' results. In the United

States, demographic surveys include age, gender, self-identified race, and ethnicity. Socioeconomic status surveys include income, educational level, employment status, how many years in the position, etc. Consult with the domain knowledge experts which of these variables are relevant for inclusion.

Important: Demographic variables often do not just represent biological reality. For example, in clinical decision-making algorithms, differences by race is often reflects the effect of racism (i.e., "*the experience of being black in America rather than being black itself*") [40]. Hanna et al. provided many conceptualizations of race that can be used depending on the problem you are trying to address [15].

Gender can also represent many meanings: bodily attributes (sex), a person's felt gender identity, how a person's gender is perceived by others, or the gendered social role that a person occupies [18]. Therefore, a person's gender as a variable is often a proxy of the power they carry in our society.

Measuring gender as a binary man–woman concept is also risky, as it will erase the experiences of people who do not fit this binary classification. To avoid erasing marginalized people's experiences, we recommend Scheuerman et al.'s guideline on how to measure gender [34].

4.4 Determine the sample size

In a quantitative experiment, the sample size (e.g., number of participants) determines whether the experiment is powered to make claims found in the results. Larger sample size means you will make more efforts to recruit the participants.

To determine sample size, you need to conduct a power analysis based on your expected effect size d. Effect size is typically determined by the community and the numbers are different in every research domain. There are several well-established effect size descriptors, including small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$). Table 1 shows sample size requirements for a two-tailed AB testing with an alpha of 0.05 required to achieve a power of 0.8:

Table 1. Sample size requirement based on the expected effect sizes.

Effect size	d	Sample size per group
Small	0.2	394
Medium	0.5	64
Large	0.8	26

Note: every experimental design requires a different power analysis.

4.5 Determine the study design and plan participant recruitment strategies

You can do a within-subject or between-subject experiment. In AB testing, a within-subject experiment means the participants will be exposed to both conditions A and B. The ordering will be counterbalanced, which means some participants will use condition A followed by B and the rest will use condition B and A. In contrast, a between-subject experiment means the participants are randomly exposed to either condition A or B.

The benefit of within-subject experiment is you can reduce the number of participants to recruit. For example, if the experiment calls for 26 participants for each group, then you will only need 26 participants in a within-subject experiment. In contrast, for a between-subject experiment, you will need $226 = 52$ participants.

However, you should only use within-subject experiment design if you are sure there is no learning effect (or carryover effect). For example, if a participant can learn something after being exposed to condition A, then what the participant has learned will influence the way they perceive condition B. If you believe that learning effect might happen, then a between-subject experiment would be more appropriate.

Once you have determined the study design and the number of participants, the next step is determining how you will recruit the participants. If the task can be done with lay people without domain knowledge, you can recruit participants from the general population. However, you have to make sure that the participants' demography matches the users of the study. You can directly recruit people, either in person or use mails, email, fliers, or social media.

If the prototype must be evaluated by someone with domain experts, then you must work with the target users to recruit participants.

Important: It is critical to make concrete steps in diversifying the participants demographic and socioeconomic status. Otherwise, the results will not sufficiently show social impact, because the evidence could not show the benefits are equitably distributed across all social groups. Diversifying the participation can be done by collaborating with community leaders and making the study more accessible.

4.6 Write the study protocol

Having a clearly-written study protocol will help you and your team run the experiments consistently. The setup of experimental studies are often very complex, thus it is very easy to miss a step or miss the ordering of several steps, which in turn could invalidate the results of the experiment.

A typical study protocol is as follows:

1. Share recruiting information. Explain to the study participants about the goal of the study, why they are asked to do the study, what they need to do, the time they need to set aside to participate, costs to participate (if

there's any), and the compensation they will get. Also, indicate that their participation is voluntary.
2. Request participant's consent. Once you have explained the details of the experiment, request the participant's consent. This can be done in person verbally, in writing, or by checking a digital checkbox.
3. Ask participants to fill a pre-survey. Sometimes you need to measure attitudinal changes before and after the exposure to the prototype. In that case you need to do the pre measure at this point.
4. Randomly expose the participants to the study conditions. Your prototype can handle the randomization, if you prefer.
5. Ask participants to fill post-survey and demographic surveys.
6. Thank and compensate the participant. Indicate that the study is complete, thank the participants for their time, and initiate the compensation process.

Tip: Determining compensation is a nuanced process. Providing too much compensation could lead to coercion: you indirectly force people to participate because of the amount of compensation. Providing too little could be an exploitation especially if the participants came from low-socioeconomic status backgrounds. Should you decide to give some compensation, a pragmatic approach is to use the living wage based on the length of a typical duration of the study. For example, if the living wage is $15/hour, then a participant for a 15-minute study should receive $3.75.

4.7 Pilot the study

Piloting the study serves many purposes. First, similar to beta-testing, piloting allows other people to identify problems in the study design before you deploy the study. For example, during piloting you could notice a wording error in your survey or unclear instructions. You might also learn that the program for randomization is not behaving as it should. Second, piloting helps you determine how long the study will take, which will inform your study protocol. Finally, piloting helps to generate test data that you can use to prepare the statistical analysis, which I will describe next.

Typically, you pilot the experiment with your colleagues. They can participate in the experiment and share their notes later. Alternatively, if they prefer, they can participate in the experiment while you are observing them.

4.8 Set up the statistical analysis

You must determine how you will statistically analyze the data. For AB testing we will use variants of t-tests to determine whether there are significant differences between participants in A and B. If the t-test computed the p-value to be less than .05, then we can reject the null hypothesis (no difference) and accept the alternate hypothesis (there is a difference).

If you are conducting a within-subject experiment, then you will use paired samples t-test. Conversely, if you are conducting a between-subject experiment, then you will use t-test for independent means. These statistical tests can be done using statistical tools such as R, Python with SciPy, and SPSS. Check their documentations on how to run the statistical tests.

Note that t-tests are parametric tests. They assume that the means of dependent variables in both conditions A and B are normally distributed. You can check this by using a histogram or Shapiro-Wilk tests.

If the means are not normally distributed, then you will use the non-parametric variants. For a within-subject AB testing, the nonparametric equivalent is Wilcoxon Signed Ranks test. For between-subject AB testing, the non-parametric equivalent is Mann-Whitney U test.

Tip: Use the pilot study data to help you develop the script to clean the data and to run the statistical tests.

4.9 Run the experiment

Begin by running the experiment with a small number of participants. Although you have run a pilot study, the pilot study participants' backgrounds might be vastly different than the actual study participants. For example, the pilot participants might have a strong technical background who knows how to avoid technical issues. Therefore, by running the experiment with a small number of participants you can identify issues before recruiting more participants.

As you gradually collect the data, once you have enough data (as determined by your power analysis) you can stop the recruitment and start running the final analysis.

4.10 Analyze the data and produce the report

Since you have developed the statistical analysis before running the study, a this point you simply need to run the tests and get the results. For a within-subject study with paired samples t-test, you will report the results as follow:

We conducted an paired samples t-test to compare [the dependent variable] in condition A and B. There was a significant (not a significant differences)

There was a significant (not a significant) difference in [the dependent variable] for condition A (Mean = ..., SD = ...) and condition B (Mean = ..., SD = ...); t(...)=..., p = ...

You can get the Means, SDs, degree of freedom, and p-value from the output of your statistical analysis.

5 Ethical Considerations

5.1 Ethical considerations for the study participants

After the serious ethical violations in the Tuskegee Syphilis Study which harmed and killed Black men, the US created the Belmont Report in 1979 [26]. This document emphasized that any study that involves human subject must meet three criteria:

1. Respect for the person (e.g., protect their autonomy, ensure informed consent).
2. Beneficence (i.e., minimizing harm and maximizing benefit of the study).
3. Justice (i.e., equitably distributing the benefits to the study participants).

In the present day, every study that involves humans in the US must receive ethical approval by the Institutional Review Board (IRB), specifically if the institution received federal fundings. Your institution's IRB will be able to provide training on how to ensure your user-centered design studies are ethical.

However, the IRB has a narrow mandate to protect individual participants. It does not necessarily ensure the study is ethical nor it will protect a group of people or the society. Thus, when developing AI for social impact, we must consider ethical issues beyond the individual participants.

5.2 Ethical considerations for the societal impact

While the above ethical guidelines are indispensable for protecting individual study participants, we also need to think about how to protect our society from the harms that AI tools could produce.

Unintended consequences emerge when a technology unexpectedly harms a specific group of people. User-centered design is a useful method to mitigate this harm, specifically if you work with people who will be impacted by the AI tool.

Suppose that a cis researcher developed an AI that could detect trans people using their photo. The goal is to pre-emptively connect trans people with mental health services. By working with trans people and experts of trans studies, the cis researcher will soon learn how such a tool can be harmful. For example, bad actors in an anti-trans society could use the AI to identify and criminalize people that were identified as trans.

The list could be extended to other domains. AI that generates images visualizing people's slim appearance could further stigmatize obesity, and further harm people with obesity. AI that predicts criminal behavior could inadvertently criminalize Black and Brown people.

Intervention-generated inequality when the technology actually works well but mostly for privileged people [39]. As a result, the technology preserves, if not widens, disparities. This gap is caused by five factors: baseline inequalities, access, adoption, adherence, and effectiveness.

Suppose that a company created a new patient-facing health technology. It turns out that the technology is generating inequality because there is an existing health disparities among low-income patients (baseline), the technology is not affordable (access), and too technical for people who are not early adopters (adoptions). Moreover, people with low-income may not have the time to use the technology (adherence) and may face many barriers to be healthy (low-effectiveness).

Veinot et al. argues that a true user-centered design and participatory design could mitigate issues around adoption, adherence, and effectiveness [39]. For example, by making the technology easier to use and more accessible as well as ensuring that the new technology could fit into people's existing routines and social practices.

Examples on how to mitigate unintended consequences and intervention-generated inequality shows that a user-centered design could play a major role in ensuring the positive impact of AI tools. Specifically by enabling domain experts (including the laypeople who will be impacted by the AI) to show the ways the AI tool will be harmful or less useful. During the interviews and paper-prototyping, these domain experts will be able to pinpoint any problematic approaches.

Intervention-produced power inequities arise when technological interventions preserve existing power inequities that create the social inequities in the first place. There are some examples of this.

Masiero and Das examined datafication in anti-poverty programs [21]. They suggest that powerful politicians who championed datafication benefitted from being seen as problem solvers and could get electoral advantages. However, the actual datafications could end up excluding marginalized people who are less experienced with technologies. Similarly, recruiting domain experts in user-centered design could also reinforce power inequities and create harmful effects. If an AI project benefits from the insights brought by domain experts (e.g., members of a marginalized community) without activating their power, then the AI project is at risk of being extractive. The benefits from the AI projects are not equitably distributed to the domain experts.

It should also be noted that the demand for data could end up exploiting fieldworkers. Bopp et al. found that the demand for data-data driven decisions could lead non-profit organizations to over focus their efforts in data collection rather than performing the original mission of their organizations [3].

Therefore, a better approach is participatory design and community-based participatory design.

With Community-Based Participatory Design, community members were involved during the entire process, including the problem definition to the interpretation of the user-centered design data analysis and the deployment of the AI tool. The ultimate goal of community-based participatory design is power sharing. That is, activating the community's power in addressing the problems that they think are important. As a side effect of this collaboration, you will be able to develop AI tools that effectively address social problems.

Wallerstein et al.'s book on Community Based Participatory Research provides a detailed approach for community-engaged research [41]. Unertl et al. provided some case studies on Community-Based Participatory Research projects for developing health technologies [37].

6 Theoretical Foundations for User-Centered Design

Theories are helpful for us to describe and infer social phenomena. These are a set of tools that help us think about the personal and social impacts of technologies [14], including AI.

The *descriptive* and the *rhetorical* powers of theories help us explain our social world. The *inferential* power helps us to develop ideas for the AI solution by hypothesizing their potential impacts. Ultimately, the *applicational* power of theories help us develop AI solutions by applying the concepts and processes in a theory into the AI design.

We will cover several theories briefly as introductions to the vast array of theoretical toolkit that will help our design. We include the main papers or books of each theory and we encourage you to explore the descriptive, rhetorical, inferential, and applicational power of the theories in depth.

6.1 Socio-Ecological Model

The socioecological model highlights the complex relationships between personal, social, as well as the cultural and societal factors. These layers are typically called the Micro, Meso, and Macro layer. The model was initially developed by Urie Bronfenbrenner [8] to explain the multiple levels of factors that influence a child's development. Subsequently, it was adapted to explain other phenomena as well.

For instance, Veinot et al. used the socioecological model to emphasize the importance of health informatics interventions to focus at the upstream level rather than just at the downstream level (Fig. 5) [38]. They referred this approach as multi-level interventions, where the health informatics supports people at the Micro level (e.g., psychology and behavioral level) as well as Meso (e.g., home, neighborhood, workplace, hospitals) and Macro level (e.g., policies and culture).

Fig. 5. A socioecological model for health informatics interventions, adapted from Veinot et al. [38]

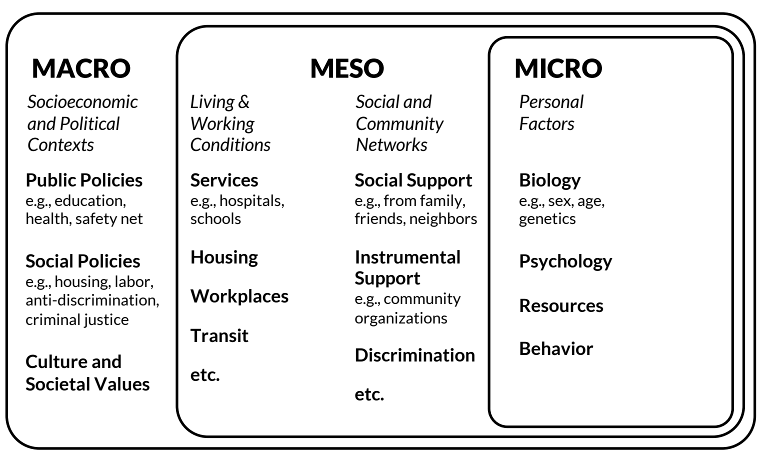


Fig. 5 shows an adapted Socio-Ecological Model for health informatics interventions [38]. This model is helpful for AI developers to target their interventions. For example, a developer with expertise in education could develop an AI tool that intervenes at the Meso level (e.g., AI for schools) or at the Macro level (e.g., AI for informing public policies). However, when developing Macro-level interventions, we should attend to the ethical concerns of Macro-level interventions that reinforces existing power inequities.

6.2 Asset-Based Design

Asset-based design seeks to address disparity issues by activating the community power, so that the community can address the problems with their capacities rather than relying solely on external support [19]. Proponents of this approach argue that interventions focused on the community assets will make the interventions more sustainable because they are aligned with the community's aspirations and capacities.

In contrast, deficit-based design is over-focused on the illness, joblessness, poverty, crime, and the hopelessness in the community. Over-focusing on these deficits is problematic in two ways. First, it assumes that communities are passive bodies with ingrained deficits. Second, it assumes that communities do not have the capacity to make positive changes. In reality, many communities have aspirations and capacities that could be supported with AI-based tools.

The difference is indeed subtle. Deficit based-design is focused on problems and thus, perpetuating beliefs that the community members are the source of the problem and incompetent. Asset-based design acknowledges that problems do exist and are often created by greater systemic forces such as impoverishment, racism, and marginalization. Asset-based design also acknowledges that community members have the aspirations and capacity to enhance themselves from within. The role of technology is to help amplify these aspirations and capacities [36].

6.3 Social Cognitive Theory (SCT)

Social Cognitive Theory explains how human thinking and actions are driven by their cognition and social environment (1998). This theory is widely used in public health and behavioral interventions. This theory is helpful to understand why people use or do not use technological solutions.

Central to this theory is the idea that self-efficacy and outcome expectations, since these two attitudes or beliefs influence our actions [1]. *Self-efficacy* is our belief that we can complete a task and *outcome expectations* are what would happen if we complete the task. For example, suppose the task is jumping a 5-feet fence. Therefore, self-efficacy is our belief in whether we can jump the 5-feet fence based on our experiences in the past jumping fences. Additionally, our observations whether other people similar to us can jump similar fences also influence our self-efficacy. In contrast, our outcome expectation is whether we will get compliments, rewards, or self-satisfaction if we successfully jump the fence.

It can also mean negatively, for example, whether we will be hurt physically or embarrassed if we fail to jump the fence. Having positive self-efficacy and outcome expectations will induce us to perform the task.

Self-efficacy can be supported by four things: our previous experiences, our emotional experiences, support from other people, and social modeling (i.e., vicarious experience, learning by observing other similar people performing similar tasks). For example, a new AI tool deployed in an organization would be easier to adopt by leveraging prior experiences (e.g., the tool is designed so that users could use their prior expertise in using a similar tool) and also through social modeling (e.g., having a demonstration of the tool by senior field workers).

However, sometimes there are impediments that inhibit people's actions [1], specifically at the personal, social, and structural levels. When designing an AI tool, it is critical to examine the impediments and barriers that limit people from using the tool effectively. The socioecological model is a useful theory to identify whether the impediments are because of barriers at the micro, meso, or macro level. Knowing the level at which an impediment operates, we can appropriately address the impediment.

6.4 Self-determination Theory (SDT)

Self-determination Theory is a theory of human motivation which argues that implicit motivation is more sustainable that extrinsic motivation [30]. It also introduces three human needs that will support intrinsic motivation when they are satisfied.

People are intrinsically motivated when they do a task for the sake of doing the task. In contrast, people are extrinsically motivated when they do a task because (1) the task is seemed to be intrinsic of their self, (2) because they worry that negative outcomes will happen if they do not do the task, (3) because they receive rewards if they do the task, or (4) they will receive punishment if they do not the task.

Intrinsic motivation is best developed when the three basic human needs are satisfied: autonomy, competence, and relatedness. Autonomy is the need to do a task based on our own volition. Competence is the need to feel that we have the capacity to do a task. Relatedness is the need to be socially connected with other people.

Therefore, an AI tool would not be optimally used if the tools damages people's autonomy, competence, and relatedness. Conversely, an AI tool would be naturally adopted if people get a sense of autonomy from using it, they feel competent in using the tool, and get the opportunity to feel they are socially connected to their peers and the community members from using it. For example, an AI tool that is very effective in giving predictions but takes away field workers' opportunity to meet the beneficiaries are less likely to be motivating to use.

6.5 Dual-Process of Cognition Theory

Dual Process Theory argues that people's thoughts are driven by implicit (automatic) and explicit (controlled) processes [20, 35]. This view is in contrast to prior theories like SCT and SDT that assume explicit process as the salient driver of thoughts. Since the two processes behave differently, it could affect the way users make their decisions with AI tools.

Implicit processes happen automatically and quickly — they rely on heuristics and associations between concepts. Explicit processes happen effortfully and are more time consuming — they rely on a set of rules in our mind. In other words, implicit processes guide a person's responses in two cases: (1) when a person needs to make a response quickly, or (2) when the person is busy, distracted, experiencing limited cognitive capacity.

The associations in implicit processes are harder to change because they are learned incrementally over many observations. In contrast, the rules in explicit processes can be learned symbolically and socially through a single event (e.g., having a teacher telling us what is right, reading a book telling the right method, organizational policy telling employees what to do.

For instance, borrowing Rothman, Sheeran, and Wood's example [29], suppose that a person is asked to reduce sugar consumption. When the person encounters a cookie with a nutritional label, their implicit and explicit processes are activated. The implicit process will make positive associations with the characteristics of the cookie (for example, the buttery smell, the golden brown color). On the other hand, the explicit process will effortfully examine the sugar level in the label. Based on a set of rules, the explicit process might determine that it is best to not consume the cookie because of the sugar level. In turn, when the person has to make a quick decision or is cognitively exhausted (thus could not effectively perform the explicit processes), the person's implicit processes will guide them to eat the cookie. Studies have shown that implicit processes guide interracial interactions, consumer decisions, and close relationships [20].

In the realm of AI for Social Impact, the dual process theory informs us that AI users might overrely on AI predictions, which is risky when the AI is incorrect and is used under high-stake conditions [9]. Buçinca et al. also demonstrated that cognitive forcing functions can reduce AI overreliance by compelling people to form their own hypothesis and decision before seeing an AI's recommendation. For example, by requiring users to click a button to see the AI predictions and explanations, requiring users to wait to see the AI predictions, or by asking users to make a prediction first before being able to see the AI prediction.

7 Conclusion

We discussed a set of tutorials and guidelines for conducting an impactful user-centered design that will help the development of AI for social impact.

The tutorials guide you to do interviews for needfinding, to create paper prototypes for testing ideas, and quantitative AB testing experiments to test or confirm a prototype. In the end, the goal of user-centered design is to demonstrate that the AI tool will have positive impacts on people and society.

We also discussed several topics on how to engage with ethical issues. First, we touched on how to conduct user-centered design ethically with the study participants. Then, we discuss how to think about ethics from the perspective of social impact. More specifically, unintended consequences, intervention-generated inequalities, and power-inequities reproductions. These concepts should be a starting framework for developers to think critically about their AI tools. Furthermore, as our society evolves and adapts to changes, these concepts may become less relevant or very relevant. Thus, it is equally important to engage with the changes in our society.

Finally, we presented several theories that give some languages for AI developers to engage with social issues. These theories operates at the societal level (e.g., Socio Ecological Model, Asset Based Design), interpersonal level (e.g., Social Cognitive Theory), and interpersonal level (e.g., Self-Determination Theory, Dual Processing Theory). With these theories, AI developers can enrich themselves with concepts and processes that explain our complex social world, in which we intervene towards a positive social impact.

7.1 Recommended readings

As we covered the basics of user-centered design, we encourage you to explore the methods further. Below are some papers and books that could help you refine your skills in user-centered design.

1. **Interviewing techniques**: Svend Brinkmann and Steinar Kvale. 2015. Interviews: Learning the craft of qualitative research interviewing. Sage. [7]
2. **Thematic analysis**: Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology. [5]
3. **Grounded Theory analysis**: Juliet Corbin and Anselm Strauss. 2014. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage Publications. [11]
4. **Paper prototyping**: Marc Rettig. 1994. Prototyping for tiny fingers. Communications of the ACM 37. [27]
5. **Designing experimental studies**: Kenneth S. Bordens and Bruce B. Abbott. 2018. Research Design and Methods: A Process Approach. McGraw-Hill Education.
6. **Community-Based Participatory Research**: Wallerstein et al., eds. 2017. Community-Based Participatory Research for Health: Advancing Social and Health Equity. John Wiley & Sons. [41]

References

1. Albert Bandura. Health promotion from the perspective of social cognitive theory. *Psychology and Health*, 13(4):623–649, 1998.

2. Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M. Vardoulakis. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *CHI Conference on Human Factors in Computing Systems (CHI '20)*, page 1–12. ACM, New York, NY, USA, 2020.
3. Chris Bopp, Ellie Harmon, and Amy Voida. Disempowered by data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, page 3608–3619, New York, New York, USA, 2017. ACM Press.
4. M Bradley and Peter J Lang. Self-assessment manikin (sam). *J. Behav. Ther. Exp. Psychiat.*, 25(1):49–59, 1994.
5. Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. Citation Key: Braun2006.
6. Virginia Braun and Victoria Clarke. One size fits all? what counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3):328–352, 2021.
7. Svend Brinkmann and Steinar Kvale. *Interviews: Learning the craft of qualitative research interviewing.* Sage, Thousand Oaks, CA, 2014.
8. Urie Bronfenbrenner. *The Ecology of Human Development.* Harvard University Press, Cambridge, MA, 1979.
9. Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. To trust or to think: Cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. *Proc. ACM Hum.-Comput. Interact*, 5(CSCW1), 2021.
10. Wynne W. Chin, Norman Johnson, and Andrew Schwarz. A fast form approach and other acceptance to measuring technology acceptance. *MIS Quarterly*, 32(4):687–703, 2008.
11. Juliet Corbin and Anselm Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory.* Sage Publications, 2014.
12. Sheena Erete, Aarti Israni, and Tawanna Dillahunt. An intersectional approach to designing in the margins. *Interactions*, 25(3):66–69, 2018.
13. Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L. Mega, and Dale R. Webster. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410, Dec 2016.
14. Christine A. Halverson. Activity theory and distributed cognition: Or what does CSCW need to do with theories? *Computer Supported Cooperative Work*, 11(1–2):243–267, 2002.
15. Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. Towards a critical race methodology in algorithmic fairness. In *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, page 501–512, 2020.
16. Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. Designing ai for trust and collaboration in time-constrained medical decisions: A sociotechnical lens. In *CHI Conference on Human Factors in Computing Systems (CHI '21).* ACM, New York, NY, USA, 2021.
17. Maia Jacobs, Melanie F. Pradier, Thomas H. McCoy Jr, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. How machine learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational Psychiatry*, 2021.

18. Os Keyes, Chandler May, and Annabelle Carrell. You keep using that word: Ways of thinking about gender in computing research. *Proc. ACM Hum.-Comput. Interact*, 5(CSCW1):1–23, 2021.
19. John Kretzmunn and John Mcknight. Assets-based community development. *National civic review*, 85(4):23–29, 1996. Citation Key: Kretzmunn1996.
20. Benedek Kurdi, Samuel J. Gershman, and Mahzarin R. Banaji. Model-free and model-based learning processes in the updating of explicit and implicit evaluations. *Proceedings of the National Academy of Sciences of the United States of America*, 116(13):6035–6044, 2019.
21. Silvia Masiero and Soumyo Das. Datafying anti-poverty programmes: implications for data justice. *Information Communication and Society*, 22(7):916–933, 2019.
22. Cormac McGrath, Per J. Palmgren, and Matilda Liljedahl. Twelve tips for conducting qualitative research interviews. *Medical Teacher*, 41(9):1002–1006, 2019.
23. Cade Metz. Google's AI reads retinas to prevent blindness in diabetics. https://www.wired.com/2016/11/googles-ai-reads-retinas-prevent-blindness-diabetics/, 2016. Accessed: 2022-08-12.
24. Penn State's College of Earth and Mineral Sciences. Active listening. https://www.e-education.psu.edu/marcellus/node/807. Accessed: 2022-08-09.
25. Ihudiya Finda Ogbonnaya-ogburu, Angela D R Smith, Alexandra To, and Kentaro Toyama. Critical race theory for hci. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2020)*, page 1–16, 2020.
26. National Commission for the Protection of Human Subjects of Biomedical Research and Behavioral. *The Belmont Report*. Washington, DC, U.S., 1978.
27. Marc Rettig. Prototyping for tiny fingers. *Communications of the ACM*, 37:21–27, 1994.
28. Yvonne Rogers, Helen Sharp, and Jenny Preece. *Interaction design: beyond human-computer interaction*. John Wiley Sons., 2015.
29. Alexander J. Rothman, Paschal Sheeran, and Wendy Wood. Reflective and automatic processes in the initiation and maintenance of dietary change. *Annals of Behavioral Medicine*, 38(SUPPL.):4–17, 2009. Citation Key: Rothman2009.
30. Richard M. Ryan and Edward L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, 55:68–78, 2000.
31. Herman Saksono, Carmen Castaneda-sceppa, Jessica Hoffman, Magy Seif El-nasr, Vivien Morris, and Andrea G. Parker. Family health promotion in low-ses neighborhoods: A two-month study of wearable activity tracking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, page 1–13, Montreal, BC, Canada, BC, Canada, 2018. ACM, New York, NY, USA.
32. Herman Saksono, Carmen Castaneda-Sceppa, Jessica A. Hoffman, Vivien Morris, Magy Seif El-Nasr, and Andrea G Parker. Storymap: Using social modeling and self-modeling to support physical activity among families of low-ses backgrounds. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, page 14, Yokohama, Japan, 2021. ACM, New York, NY, USA.
33. Herman Saksono and Andrea Grimes Parker. Reflective informatics through family storytelling: Self-discovering physical activity predictors. In *Proceedings of the 2017 ACM Conference on Human Factors in Computing Systems - CHI '17*, page 5232–5244, Denver, CO, USA, 2017. ACM. Citation Key: Saksono2017.
34. Morgan Klaus Scheuerman, Katta Spiel, Oliver L. Haimson, Foad Hamidi, and Stacy M. Branham. HCI guidelines for gender equity and inclusivity. https://www.morgan-klaus.com/gender-guidelines.html, 2020.

35. Eliot R. Smith and Jamie DeCoster. Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2):108–131, 2000.
36. Kentaro Toyama. Design, needs, and aspirations in international development. In Jyoti Choudrie, M. Sirajul Islam, Fathul Wahid, Julian M. Bass, and Johanes Eka Priyatma, editors, *Information and Communication Technologies for Development*, IFIP Advances in Information and Communication Technology, page 24–32, Cham, 2017. Springer International Publishing.
37. Kim M. Unertl, Chris L. Schaefbauer, Terrance R. Campbell, Charles Senteio, Katie A. Siek, Suzanne Bakken, and Tiffany C. Veinot. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. *Journal of the American Medical Informatics Association*, 23(1):60–73, 2016.
38. Tiffany C. Veinot, Jessica S. Ancker, Heather Cole-Lewis, Elizabeth D. Mynatt, Andrea G. Parker, Katie A. Siek, and Lena Mamykina. Leveling up: On the potential of upstream health informatics interventions to enhance health equity. *Medical Care*, 57(6):S108–S114, 2019.
39. Tiffany C Veinot, Hannah Mitchell, and Jessica S Ancker. Good intentions are not enough: how informatics interventions can worsen inequality. *Journal of the American Medical Informatics Association*, 0(May):1–9, 2018.
40. Darshali A. Vyas, Leo G. Eisenstein, and David S. Jones. Hidden in plain sight — reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, 383(9):874–882, 2020.
41. Nina Wallerstein, Bonnie Duran, John G. Oetzel, and Meredith Minkler. *Community-Based Participatory Research for Health: Advancing Social and Health Equity*. John Wiley Sons, Oct 2017.