Human Computer Interaction (In English)

Prof. Rajiv Ratn Shah

Department of Computer Science and Engineering
Institute IIT Madras

Data Requirement, Gathering, and Analysis

 hello everyone welcome back to the class of human computer interaction so let's get started this week we will be talking about data requirement gathering and analysis before that in the previous lectures we have talked about interface interaction and their types we have also discussed about conceptual modeling aspect of the interfaces and cognitive aspects of interfaces. We have also discussed the importance and emerging trends of interfaces. We also discussed about tutorials on cognitive aspects, and we had hands-on experience with Adobe Express and Behance. So this week, we'll be talking about what is data requirement, Subsequently, we'll be talking about what is data gathering analysis. And with the help of one use case, we'll be discussing all these. We will be also having a tutorial on data gathering and analysis, followed by a panel discussion on IRB.

 That is called Institutional Review Board. So why data is important in SCI? So data is a fundamental in understanding user needs and informing design decisions in SCI. Because if you don't know about users, how you are going to build something useful for them? And what are the ways to know about their needs? You need to have some data, some information. So that's why you need data requirements.

 to identify the type of data needed to meet the research objectives, the problem you are solving. And there are several techniques to collect the relevant data. So that is something we are going to discuss as part of the data gathering techniques. And once you have the data collected, how to get useful insight from it so that you can build something useful, what your users are looking for. And that's what we do in data analysis.

 So in data analysis, we'll be talking about different methods for interpreting and extracting insight from the data. And there are several key considerations we need to keep in mind while collecting data or analyzing the data and so on. So that include ethics, bias, fairness, transparency, and accountability. For example, you are building an autonomous vehicle. If it makes an accident, who is going to be accountable? At the same time, if a person is driving a car and made an accident, then you make the person accountable and probably do the law of course.

 Similarly, other big consideration that we need to keep in mind, how much data do we

need? Often it requires too much data, which is beyond what you can handle, that is more towards big data. Crowdsourcing, when probably it is not enough to probably get the data from only few people around you. Can we get the help of more people who are scattered all across the globe. So that is using the technique of crowdsourcing, we can get more data from all over the world. Real-time data, often we need real-time data rather than some old data which you can dependent on.

Similarly, we need to include diversity and sometimes we need to perform the field study. to get the data in the real environment where the user is going to use. So user research is the fuel for innovation, driving design decision with real-world insight, what Don Norman mentioned. So the benefit is it helps you in accurate insights so that you have a better understanding of the user needs, their requirement, and a better insight how you can help them. So it help you improve design decision because once you have the informed choices during the design phase based on the insight you got about the user needs, you can help them in better way.

Iterative improvements, so you have the ability to re-find design based on the real feedback you are getting from the user. So again, feedback in terms of data that you'll be having at different stages. Stakeholder buy-in, so data-driven justification for design decisions rather than just assumptions that we discussed in the earlier lecture. Because what if your assumption is wrong? You are going to probably have a system which is not going to solve the actual needs of the users. So what is data requirement? So it is nothing but identifying what information is necessary to inform designer to make the designs or for the solutions.

that is set by SEI pioneer Ben. So you're going to deal with different kind of data. So one category you can think of more user data that may include their demographics, their preferences. For example, an e-commerce website like Amazon Collect user data including demographics, age, location, and preferences . This data is used to personalize product recommendations and tailor marketing campaigns.

companies like this it's not just limited to the individual data but they probably they can also consider the insight from a group of people's data so that probably they can make a global decision and probably for example they can probably think of a better supply chain solution based on forecasting the demand based on the interaction a group of people are having could be task data so that is primarily about a particular task for example task complexity and the frequency an example would be a bank app collects data on tasks like transforming funds checking balances paying bills and so on so it analyzes the complexity of each task and what are the different steps involved to complete these and how frequently user perform them to streamline the interfaces and reduce user efforts.

Often you have seen that the different example we have discussed earlier lectures. So there are many frequent functions or tasks user perform. It could be, for example, on the different payment app we are using, just scan and pay through UPI. So that is more frequent as compared to probably the checking balance.

So in that case, probably they can probably put more frequent tasks in the front page or something rather than probably keeping them behind different clicks so that you can facilitate user in completing the frequent task. And you can also analyze whether the users are facing any difficulty in achieving this, which are more frequent because that end of the day, I mean, it is again related to the user experience, the more tasks you'll be doing and how you are doing eventually that contributes to the user experience. So next kind of data you can think of is like contextual data. So more in the environment of use that you can say that. So for example, a fitness tracking app like Fitbit collects contextual data whether the user is indoor or outdoor, probably their location at the time of day.

This information help suggest exercise, track outdoor activities and provide tailored fitness recommendations. So a lot of can be done based on the context, based on the environment the users are going to use your solutions. Next is system data. So that is more about the performance or error metrics you can call that. So for example, video conferencing tools like Zoom monitor system data including call quality connection stability and error occurrence.

Because often during online meetings, we realize that probably there is lag of voice or probably the system freezes or probably you unexpectedly thrown out of the call and so on so so these kind of data basically helps identify the bugs optimize the platform for different internet speeds because when you are having an online call especially if you recall the days of covid there are students who are attending the online lectures from metropolitan cities where probably they may have very good internet connection. At the same time, we have students from the rural areas where they are hardly getting any connections. So how to build accessible solutions to accommodate all kinds of users depending on their bandwidth they have. So improve overall system reliability. So when we're talking about the data, as we discussed, we need to know what user want.

And accordingly, we have to collect the data which will help us in getting the insight, what kind of solution they are looking for, what kind of pain point they have in the existing solutions, and where actually we can help them in building a better system so that probably we can make them make their life better so a very important activity that is called need finding so it help in understanding the underlying needs challenges behaviors of users is critical for gathering relevant data and the purpose is as mentioned here identify what data is needed and based on the research questions or problem to be solved

and the focus area here is user behavior goals pain point preferences that we can know more about the users so there are different real world constraints such as time budget and accessibility So a good design begins with a deep user understanding set by the normal. So we have to know more about users and for that need finding help us getting know more about what they want, what challenges they are facing, what all the behavior they have while using the system what are the usual way of doing something so that we don't try to change that way so that it will be easier for them to onboard on the system and be a one of the loyal users for your system or applications so examples you can think of fitness app where the data requirement might include user activity level health metrics like how many steps you have walked today what are the your heart rate and motivations for the exercise based on your activity your goals that you have set your for yourself so similarly mobile mobile health apps where need founding would help involve understanding the specific challenges user faces so that is related to the tracking exercise diet stress and so on because different users have different needs. And that is why we need a better way of understanding different kind of solutions based on the requirement they have. So these guides the data requirements such as the need for the health metrics, user activity logs, or feedback data to support the apps functionality. Again, so different tools are there.

 Some of them we'll be discussing in more details in the upcoming So, for example, user interview, survey for the qualitative data. Again, Google Analytics app uses data for behavior and so on. So, one thing about interview or survey and so on, if you recall in the previous lectures we discussed, it is very important to ask the right questions. Ideally, the number of questions should be minimal in nature and it should broadly cover everything you want to know about the user. So considering that, I mean while building, constructing the user interview or the survey for qualitative data or any other ways basically you are going to collect the data, it is very important to be sure that you don't ask irrelevant question.

 You need to be focused, you need to ask the question basically around the solutions they are going to have, around the problems they are facing, around the expectation they have and so on. So often we also note that are some kind of system which requires very minimal data there are some kind of system which require huge amount of data in order to build the solutions so that's where basically we are talking about the big data so it's a massive amount of user generated data so here the user generated data could be of any type it could be text it could be images video and so on so these days all of us have the smartphones right so you keep generating multimedia data multimodal data and so on in fact there are A lot of sensors are also installed at many places and you are also contributing towards that and so on. So what is the uses of big data in SCI? So that help you in personalization. So understanding user behavior to provide personalized content,

recommendation tailored experience and so on. So often these recommendations or personalizations methods are based on deep learning models and so on.

So which requires a significant amount of data of different users and so on to help you in building such system. Similarly, the pattern recognition so that analyze large data sets to identify the trends and the pattern inform design decisions so again that require a good amount of data to build something and which can provide you kind of more accurate solutions real time decision making. So use big data to support real time decisions and adapt interfaces dynamically based on the live user input. So for example, Google search provides autocomplete suggestions you might have seen in your real life. Do you know how does it come from? So often these display results based on the trending queries and real time search data.

So in a way there are crowdsourcing kind of user input across the globe or probably your vicinity to help you probably answering more relevant and latest information. So for example, there is match is happening between say India and probably some other country if you ask what is the score so probably based on the context because there are probably many other people might have written for example India and Australia match score India and Australia test math score and so on so probably it may be able to just answer your question based on like math score based on probably even your past history and so on it can eventually tell you the score even you did not give in a complete input. So the key concept in any data collection that we are talking about or any data we are talking about, often they are biased in data because many systems that we are building these days, they are based on a very simple concept, garbage in, garbage out. If you have biased data, then probably you may have biased model. If you have noisy data, then probably you'll have a noisy model and so on.

Same holds true for fairness. you need to make sure that your model works fairly. And in that case, basically you also have to ensure that the data you collected is not biased and so on. Similarly, transparency. So you should have a clear explanation of how data is collected, used and processed. Because many, many times we see in our real life, probably we are surprised to see that how the different system we are using are working so often some of you might have noticed that so probably the based on any discussion that you are having with your friend you might see that some probably your smartphone give you a suggestion okay so probably you might be interested in an electrical vehicle car or bike and you'll be surprised to know that probably earlier in the morning or probably yesterday i was talking with my friend about a ev car or bike how did the system know about it same happens true for some of the e-commerce websites probably you get some suggestions which probably you were just thinking or probably talking about with your friends probably or some social media platform or you're just speaking

about it and so on so often there are there are basically the data leakage and that violate your privacy and happening so it's not like all the companies are doing but there is a quite high chance that probably also due to the settings that you may have your phone they probably sense what you're talking about what you are looking for and they can basically can manipulate and probably influence you to buy something, probably from some specific company.

For example, there is e-commerce website X or Y or Z, whatever, and probably if it says that you're looking for, say, an electrical vehicle, say bike, there are N number of companies basically selling that. So it may, if the company is basically getting a kind of good commission, from a company X, then probably it will promote you the content from X rather than Y and Z. Doesn't matter, the quality of Y or Z is better than X. So it is happening, right? So yeah, we'll have to be careful about it. So, Amazons, one of the example you might have seen that, Amazons recruitment tool found to be biased against women due to the biased data it has.

And that is the reason basically they scrapped their secret AI recruiting tool that showed bias against the women. So there at the same time there are many tools which are fair and doing showing the fairness indicator like AI fairness 360 to assess fairness in machine learning model so these are the way basically you can also find. So similarly there is a recruitment tool called Chiron. So Chiron basically helps in conducting fair online examination with the proctoring. So fairness is not a luxury it's a necessity said by Feifei Li.

Often in your requirement of the data, you are looking for some real-time data. And these real-time data basically help you in making some dynamic and immediate decisions. So it helps you in personalization, monitoring, and so on. So for example, personalization can be achieved by adjusting interface or content based on the user sections.

and similarly monitoring. So continuous observing user behavior to improve interaction in real time. So for example, Uber, Ola, and probably many other cab services that you are using. So it uses real-time data to match riders and the drivers quickly. So wearable devices and apps like Fitbit, Apple Health continuously track user physical activity, heart rate, and sleep patterns to probably find out how good or bad you're doing so systems like early warning network for natural disaster like earthquake or hurricanes say they are uses real data from sensor to alert people you might have know that i mean earlier what happens earthquake or probably such cyclone comes and probably thousands of life has gone in just probably few minute or something earlier we do not have enough technology or significant technology that can help us in finding that out in advance. But we are

fortunate that the advancement in technology has take us to a level where probably we are able to find all these in advance and we can evacuate the people from there and probably take them to some safer places.

So there are many tools like Socket AI, Firebase for real-time data handling, which are there. As I said, often we are in need of data in the real environment so that we can understand the real uses of the data. We can understand the users in the real environment. How do they play with the system? How do they work in the system? and probably using the system in different ways. And we can know more about their understanding, their challenges, their requirement, and probably the expectations.

So the benefit is it provides a deeper understanding of user context and the behavior. and it identifies challenges that may not emerge in some virtual setting, lab setting and so on. So because if you assume that you can understand everything about the user just probably on a phone, on a video call or something, I think it's not going to work in that way. We have to we have to basically make our shoes dirty by going into the field and probably get to know more about our real users. So for example, Apple conduct field study to test the usability of the new iPhone features in everyday settings.

So again, there are different tools like field notes, video recording for documentation and all which are there. Another few points that we have to be a bit careful about when we are working with the data, we need to ensure the diversity. So for example, ensuring diverse representation of user and data collection to avoid biased or narrow insight. Because if you don't have the diverse set of users in your system, probably your system may work quite well for the kind of user it has seen, the kind of behavior it has seen. but it fails immediately when it sees some kind of new users or probably new kind of behavior it encounters.

So another next thing is importance. Accounts for different user demographics, behavior, and experience in design. So for example, Google ensure their search engines are tested across a wide range of cultures and demographic backgrounds. because if it won't do so, so probably it will start showing you some results which are offensive or probably inaccurate based on the culture or demographic that you have. So there are several tools you can think of like diversity metric and survey and data analytics tool to ensure that your data is diverse, having enough representation from the different stakeholders you have, different behaviors are captured, and so on. And you have basically a sample space which is representing all kind of users, all kind of behavior, all kind of functionality that they are going to have and so on.

So enough about data requirement, why do we need data, what are the importance, what

all we have to keep in mind while finding out what kind of data we have and so on. So let's talk about data gathering, how to collect the data. So gathering data is basically from the user through various methods to gain insight into their need, preference, and behaviors. So it could be, most of them basically can be, you can categorize into two kind of data.

 For example, you may have qualitative data. So that comes from interviews, some focus groups, ethnographic studies, field studies, the natural environment and so on you may have some quantitative data so where you can collect the same using surveys some user metrics some analytics some user logs and so on And regarding the tools, many of you might have seen the Survey Monkey, Google Form are most common tools for survey. There are many others. Similarly, for user testing, Lookback.io for remote usability testing people are using.

 So that is something are quite popular. So some real world example you can consider here like Spotify. So it's using crowdsourcing data to refine music recommendation based on different users interacting with the system. Similarly, mobile banking apps track system data such as server response time, transaction errors. make sure that we are logging everything. It's not just what is happening right or wrong, in fact both, because it help you in getting a better insight of the user behavior, where probably they might be making a mistake, how is there any probably intelligent way of interacting with the system, which probably most other users have not considered.

 And then probably again, you can help with the larger set of user by providing a shortcut or probably a better way and so. So as I mentioned that often you may need a big data in your system and how do you can collect those big data? because you may not have enough resources or probably reach out to the people to get the data. So that's where basically you need to take the help of crowdsourcing, where you gather data from a large number of people, often online, to get insight from a diverse user base and so on. For example, imagine Mechanical Turk human protocol to gather feedback on multiple interface designs or probably get the labels on different tasks and so on.

 So the benefits are scalability, diversity. So you can quickly gather feedback from the large user groups. Similarly, you can collect data from diverse user worldwide, which is crucial for inclusive design. So we discussed just few slides earlier as well. We need to have diverse set of data, diverse set of users. again for example being settled in India it may not possible for us to probably go to US or probably Europe or probably say Singapore to probably get data about or data from the users there so that is very time consuming that is very costly process so what if I can get the same in probably very short amount of time and probably less price that is something is now feasible but often one of

the challenge often we face in the crowdsource data how to control the quality so that i'm going to consider in the upcoming slides so some real world example of crowdsourcing you can think of is wikipedia is one of the world's largest crowdsourcing encyclopedia where volunteers contribute and edit articles Google map again relies on the user contributed reviews, photos and location data to provide accurate local business information.

So often we see that I mean when you are at some place you ask what are the places to see here you see a lot of recommendation based on the with the photos with the attraction and so or the even the reviews often we go through the reviews about any restaurant before going there. So all those are crowdsource information. So Airbnb uses crowdsource views from both host and guest to maintain trust within its community and everything is in public. What a host has written about a guest, what a guest written about a host. So you can get a better sense of about a host or a user and accordingly make a decision.

So tools are a Michigan Mechanical Turk, Human Protocol, Audino, Crowdflower for crowdsourcing tasks and so on. So as I mentioned, so we need to make sure the data labeling should be of high quality and that's where you need to control the annotation quality so the importance of quality control in data labeling basically ensure you the high quality reliable and accurate level data for training ai and machine learning based models it minimizes errors and biases that can lead to flawed prediction or poor user experience, it affect the performance and generalization of data driven system and which is crucial for user centered design and the system. So some techniques through which basically ensure that the data is of good quality and you can discard the data which is of inferior quality so some gold standard you can make so what you can do you can use a small set of pre labeled high quality data as a benchmark and you can compare new level data against the gold standard data for accuracy So in that case, if from some user, if you see the data which they have labeled is not as of good quality as compared to the gold standard you have created, it means probably the user may not be or probably the annotator is not able to doing the annotation well so that is of not good quality otherwise we are good to go similarly enter annotation agreement so in this case what we do we employ multiple annotators for the same data set and It measures the consistency of the label using the popular matrix called Cohen Kappa and the place Kappa to assess the agreement. Suppose this is the task. So you give the same task to be annotated by two people, people A and people B.

So in that case this majority weighting and weighted consensus help you in determining the kind of right data. Similarly, FTA learning, where you use machine learning models to identify uncertain or difficult cases. It prioritizes these cases for manual review to ensure

that they are correctly labeled. Another technique for quality control is regular quality audit, where periodically sample labeled data to check for accuracy and consistency. so please provide feedback and probably retraining for annotators if quality falls below some set threshold so in a way you require some trained annotators or probably reviewers who are going to do the annotation for you.

Similarly automated quality checks so you can implement algorithm to detect inconsistencies or probably missing labels or outliers. So you can also use machine learning methods to flag some potential errors for manual review and so on because often when you're dealing with big data it's not possible for you to manually verify all the data samples that you have or annotation which are being done. So you need some assistance to probably check which one may be correct, which may be wrong. And then probably you go for reviewing the samples by using different techniques, be it like some algorithms or machine learning and so on. So there are different tools for data labeling and probably which support the quality control.

So there is LabelBox, a data annotation tool with built-in quality metrics. review flow similarly there is something called prodigy a tool that support active learning and quality checks for label data supervised.ly so it includes review and feedback system to ensure high quality labels audino a speech data notation tool with built-in metrics and probably review workflow and so on so example real world example you can think of amazing mechanical talk is one of the quite common and popular tool similarly Google's recaptcha that uses gold standard approach to comparing user's answer to some known set of correct labels to improve the machine learning accuracy so quality data labeling is a backbone of any successful AI project and what you feed is in what you get out so that's what I said garbage in garbage out and again there are so many people have contributed towards this learning another thing can you collect data from anybody can you collect data from anywhere isn't there any guideline to do so ideally there should be a guideline ethical guideline to inform that how you are going to collect the data who your user would be how you are going to use the data how you are going to store the data and what you are going to do with that data. So all these things should be informed. So that's why in real life, I mean, usually we go for ethical data collection, so that ensure privacy, fairness, transparency.

in gathering and analyzing user data so for that we need an irb approval so that is called institutional review boards that ensure that the research follows ethical guidelines protecting participants and so on and it is one of the very standard practice followed all over the world so for example facebook's ethics review conducts ethical review of experiment involving user data So the moral of the data collection is to collect with responsibility and not just about the results. So when we're talking about ethical

consideration, so data privacy is one of the key regulation that we need to keep in mind. So data privacy is a critical aspect of designing user centric system, focusing on how user data is collected, how it is stored and how managed securely. So compliance with the data security regulation is crucial to maintain user trust and avoiding legal consequences. Often we know in real life there are different companies that doesn't follow these compliance regulations, they face a lot of penalty.

One example that comes to my mind is Paytm, right? So they were alleged that they sell the user data to probably someone in China. I don't know how truth or wrong it is, but based on those, basically, the RBI has taken actions against them to basically they put some penalty and so on again there are similarly such thing as happened on google meta and many other big companies as well and often they have they have been at least that they are selling the data they are not using the data they are not storing the data well and so on often we also got to know that i mean there is a company which somehow revealed the Aadhaar number of so many Indian users and so on, PAN card number of so many other users and so on. So again, so these are something which should not happen. So key, and that is the reason there are some regulation policies. So like key data privacy regulations are most important popular one I would say that GDPR, so General Data Protection Regulation and that is in European Union.

Similarly we have CCPA California, Consumer Privacy Act in USA, we have HIPPA, we have IPADIA and so on. So now different countries in the world now following similar data privacy regulations and often most of them are inspired by GDPR and so on. So what is the best practices for under data privacy regulation? We should have transparency. So we should clearly communicate the data practices in simple language to the participants or the users we have. We have security measures to implement the encrypt and data anonymity when whatever we are having.

Similarly, user control. So we should provide user with a tool to manage their data like privacy setting, opt-out option and so on. Data minimization, so where we collect only necessary data. It's not like you just collect everything about the user even you don't require or even which probably user has not agreed to provide. So collect user only necessary data and avoid storing it longer than required. And often, for example, if the data is only required or probably you are allowed to use only for a year, five years and so on, don't use beyond that.

So ethical consideration. So prioritize privacy and ethics in the design process to build user trust. So that's why someone said rightly, in the world of the design, the best data is the data that is collected ethically and use responsibly. So example is Facebook face GDPR penalties for not providing clear data users information and lacking proper user

content mechanism leading to major redesign of its privacy settings. Because privacy is not an option, it should not be the price wish except for just getting on the internet. another news probably you might have aware of Russia finds Google more than money there in the world again due to something related to data and account and so on so if you can't protect it don't collect it so how you can ensure privacy and data privacy so there are several tools which are there so one trust trust arc which are GDPR compliance and data privacy management platforms.

Similarly, we have data anonymizing platforms like ARClock and ARX for data and so on. Similarly, consent management tools like Cookiebot, Usano for managing user content and so. So by understanding and adhering to these regulations, designers and developers can ensure that their product not only comply with the law, but also prioritize user privacy leading to most trustworthy and user-friendly systems. So that's all about data gathering. So what you do once you gather the data? So the next thing is how to analyze it so that you can derive some useful insight from it.

So data analysis is a process of interpreting the data to derive some actionable insights. And the purpose is to identify the patterns, trend, insight from the collected data, turn raw data into some actionable design improvements or feedback because often you collect data which is huge probably often may not be in structured form there may be probably too much noisy data in it how to get something useful out of it and so on so there are several techniques as we discussed primarily we have kind of data quantitative data and qualitative data so correspondingly we have quantitative analysis and the qualitative analysis so for quantitative data we perform like statistical data analysis right for example regression analysis t-test mean max and so on. Similarly, we have data visualization for representing trends and patterns. So for qualitative analysis, so you recall in the previous lecture, we discussed about affinity mapping.

Similarly, we have thematic coding for categorizing the interview data. You try to identify the patterns by looking into the data that you have. And similarly, you synthesize the findings, so where you develop themes and insight out of it. You also prioritize the issues, so focus on the most critical user problem because you may come up with probably so many user problems, which one is probably faced by so many users that you have, so that you can prioritize them and probably focus them on first. Similarly, translate it to design.

So use finding to inform design solutions accordingly. So data analysis in SCI is about turning information into some actionable insight that improve the user experience. That said by Bill, one of the SCI pioneer. So this is example of data analysis. So in this case, you can see that an affinity diagram basically built during the design of web application

as shown in the book. So similarly, you can say that improving mobile app for food delivery, if you talk about all the steps we discussed, so you gather through some surveys or interviews.

Similarly, you can build some affinity mapping used to identify pain points like confusing navigation. Similarly, you develop persona to represent different types of users. So for example, you may have busy professionals or probably you have students, you may have probably housewife or probably anybody else. Similarly, design changes implemented to simplify the menu and improve search functionality based on the probably the user you have. Another kind of probably analysis that we perform, especially where there is some conversation is happening, So, we are basically conversational discourse, which is analyzing users' conversation or interaction with interfaces to understand communication pattern.

And given that any interactive system in any SCI-based system, conversation is the key. So without this, it's kind of incomplete to perform the analysis. So there are several methods you can consider. Conversational analysis where examining users' dialogue to improve voice interface or chatbots. For discourse analysis, understand how language and communication save user interaction.

So this is something you can see and read more about it. So example is Alexa analyzing voice command to improve natural language understanding, communication tools like Slack and Microsoft Teams that facilitate team communication through channels, direct messages and video calls and so on. So there are several popular tools like speech analytics software for voice data, LIWC, which is one of the quite popular data that we have for discourse analysis, you can have a look here. So often it is important to see the inside or probably figure out by looking at visually. And that's where basically we require data visualization. So graphical representation of the data to identify trends, patterns and insight, because often we say that, right? We see what we believe rather than probably what someone told me.

So importance is it helps designers to understand large amount of data and make informed decisions. So for example, Google Analytics, so it visualizes a website traffic pattern to improve user experience. So there are several popular tools like Tableau, Power BI, Google Data Studio for creating interactive visualizations. there are several statistical data analysis tools are available so that using statistical method to quantify and interpret user data so it can perform regression analysis, ANOVA, t-test and many others to identify the relation between different variables you have in your research so for example facebook uses statistical analysis to measure the effectiveness of the change in the newsfeed algorithm and and so on so rspss python Again, so with libraries like Pandas,

SciPy for statistical computation, computing that you use.

So let's apply all the things we have learned through a case study. So let's consider a case study of enabling independent learning for a BVI. So BVI is blind and visually impaired. So in this case, basically we are working on a problem, how BVI students face significant challenges in education due to absence of visual stimuli, which impact their ability to perceive, focus, retain and process information effectively. So if you talk about traditional education approaches, they primarily rely on braille and audio based materials.

And they often fail to meet the diverse learning needs of students with BVI. So in our study, which is being done with my student and my collaborator colleague Dr. Richa. allowing users to switch modalities based on their preferences and see what kind of behavior they have when they're given a choice to probably switch between the different modalities which one is more effective which one is more convenient where eventually they are going to have better user experience so limitation with the traditional approaches as i said so it limited availability of accessible educational materials restrict restricted reach of assistive technology and heavy reliance on human assistance. And especially everyone wants to be independent.

Doesn't matter whether the person is differently abled or normal. So in that case, basically, how we can make them more independent. So BUI students face challenges with inaccessible visual content, cognitive difficulties and high cost and impracticality of build material making inclusive education harder to achieve. So the motivation is a multi-modal approach can offer effective way to address the limitation of a single modality based education approach that is something we found in our research. As proposed by the dual code theory Integrating verbal audio and non-verbal tactile system can establish multiple pathways for memory storage and retrieval.

This can help in reducing sensory overload and facilitate deeper understanding of the material. So what kind of data is required because it's very important to understand what kind of data we need to have and later basically we go for the collection. So in our case basically we require activity wise details physiological data and learning and response time improvement to understand were there any improvement with switching modality which modalities contributes to better learning and so on so we capture time measurement pre-test learning activity and post-test so here basically we want to understand is there any what is the impact of learning they have which modality contributes to a better learning so we give a kind of topic to students to learn about it and before that we ask few questions to just understand their background knowledge about it we call it pre-test how do they perform in pre-test then there is learning activity happen with the different modality Once the learning happened with different modality, we ask

them to appear in the test again and see how do they perform, whether there is any improvement in learning or not. So that is called pre-test, post-test. So we see the time difference, time measurements. how much time did they take we also see the heart rate data again collected during the pre-test post-test and learning activity we use samsung galaxy 4 smartwatch for that again you can use anything others any other device learning improvement so we want to see how much score person has scored in the pre-test as compared to post-test Response time, so measured in second for each activity and so on.

So you can see, I mean, these are the different instruments, these are the different units and these kind of different data we have collected, we require for solving our problem statement. Of course, as I mentioned earlier, it is very important to have ethical approval for the research you are doing. So you need to have an IRB application that typically includes the following document study overview so that talks about title abstract purpose objective and the background of research research team so information about the principal investigator and the probably the team members who are going to perform the research, study design and methodology so more information about the research design methodology and the timeline that we have in mind participant details so and what is their inclusion and exclusion criteria so their eligibility criteria recruitment methods and sample size and the compensation it is very important to have the informed consent. So consent form and the process for obtaining consent form for participants that need to be mentioned in your IRB application. Risk and benefit assessments that potentially risk mitigation strategies and what are the benefits for the participant that is need to be informed.

Similarly, data management. So the data collection methods, storage, confidentiality, analysis plan that you have. Privacy and confidentiality, how you're going to basically ensure the privacy and the data security of the participants data. Similarly, ethical consideration, compliance with ethical guideline and handling of vulnerable populations. Withdrawal process, you need to inform like information on participants withdrawal rights. So for example, while going through the data collection, the user feel, no, I don't want to be now part of the study.

Then you need to include, then you need to remove the user's data. Funding and conflict of interest. So source of funding and potential conflict is very important to be highlighted. Supporting documents, so copies of survey, recruitment materials, interview guides, questions, consent form, everything you need to provide a supporting document. Signature and assurance, signature from the PI and relevant institutional authority to inform that, I mean the data will be used as per the IRB declaration and so on. So these are some sample for example IRB application that requires the project details, how you want the application to go through, probably when is your anticipated start date, end date,

name of the probably PI, and then again all the other details I mentioned in the earlier slide that you need to provide and once it is approved by then probably you will get some kind of approval letter which kind again with some terms and condition in what case and all that you can use this data and after that in a way you are going in right direction and you can go ahead with collecting data so in this case you can see that So this is our experimental data collection setup where there are students, different signals and the data we are collecting and there are pre-test happening, there is post-test happening, this is learning activity happening and in the learning activity there are three group of users. like multi-modal activity similarly we have audio modality so we have just build modality and see how the different users are doing so these are the devices that we are that we are planning to use for probably collect the data and this is the setup as you can see that where in case of multimodal setup how the person is using in case of just audio how the person is using in case of braille how the the person is just using the braille keyboard to go through so these are some data analysis once you collect the data how to get the insight from it so you need to perform the data analysis and probably need to perform more experiment on it so in this case as you can see that So in case of audio and multimodal comparison, a t-test was conducted and the result shows that multimodal performing slightly better in terms of mean they have computed.

So with a p-value of this. Similarly, in case of braille and multimodal comparison is done with this particular U-test. And in this case, it shows that braille performing better and blah, blah, blah, that is something you can see. So at least you can see here, for example, in this case, it's showing learning improvement in case of multimodality as compared to audio. Similarly, here basically you can see that the study analysis has been done on average response time improvement by different modality. Through the result, they have shown that multimodal settings you have better time improvement in terms of pre-test and post-test as compared to the braille modality and the audio modality so in a way i mean it is confirming the hypothesis and probably the research problem has been started is the multimodal learning is useful or not can it support it in independent learning for blind and visually impaired students.

Again, so similarly heart rate analysis has been done where you can see that how the heart rate fluctuated in case of multimodality audio and the braille. And you might have seen that in case of braille, the heart fluctuation is probably quite high. And similarly, in case of audio and braille, it is correspondingly lower and so on. And we might have seen that in case of braille, the heart fluctuation is minimal.

So this is basically the heart rate variation result that you can see here. So they have performed across stages, across modalities and accordingly they produce the results. So the summary is data is the backbone of user-centered design. Proper data collection

analysis and ethical practices are foundational to the user-centered design principles. So data-driven design creates more intuitive and satisfying user experience. Effective gathering and analysis lead to better interfaces, balancing the need for big data, real-time feedback, fairness, and sure, equitable, transparent, and efficient system design.

Diversity and ethical principles are key to develop inclusive and trustworthy SAI systems, and iterative data collection refines the product continuously. So good data leads to good design, as Jared mentioned. So these are some additional materials. So we'll be covering a tutorial on data gathering and analysis by our brilliant TA Hrithik Bamba. Since IRB is one of the most important aspect of any experiment, any research, any data collection, it is important to have a discussion on it.

So that's where we have one panel discussion on IRB by Rahul Harmehar and Tony and subsequently we have a sign here. For further readings you can go through these resources which are quite quite useful for more information and gather deeper insights. With this thank you so much.