Mid-Term Proposal

How it began:

While I was brainstorming ideas, I wanted this project to be personal and reflect on my daily struggles. So, I decided to focus on the difficulties college students have when navigating academic visual content. I feel that, at the higher education level, students often face greater struggles, particularly because the subjects are harder and more complex. This is even more pronounced for students who major in subjects that require a lot of visually intensive documentation and homework.

These graphs, diagrams, and images, unless specifically taught in class, are often left for the students themselves to figure out. However, these visuals may not be the most effective learning tools for every student, especially for those who are not visual learners or those with visual impairments. This led me to think of an AI audio assistant that can help explain these complex visuals to the general student population. I specifically chose audio because I want this to also be accessible to the visually impaired, but my main users will still be the wider non-disabled student body.

My main goal was then to evaluate whether my hypothesis about students struggling with visual content and wanting an audio-based tool was accurate. So, I set out my research with three main questions:

1. Do students struggle with visual content, and if so why?
2. Is there a demand for audio-based assistant tools?
3. How do students believe AI can assist with interpreting visual content, and what are the limitations of current technology and techniques in this area?

User Research Summary:

With the three main questions in mind, I decided to choose two design processes for my user research: surveys and interviews. For the two surveys that I conducted, I focused on exploring people's challenges with visual content and their preferences for AI tech. The user base of the two surveys was different, one was aimed at the general student population while the other one was aimed at visually impaired individuals. Similar to the surveys, I also conducted two interviews, one with a current IDM graduate student and another one conducted with a visually impaired pastor who has experienced college life with his visual disabilities. In both the surveys and interviews, my goal is to understand and gain insights into my key research questions and also get specific details on the types of visual content students struggle with, the challenges they face, and also, the tools they currently use.

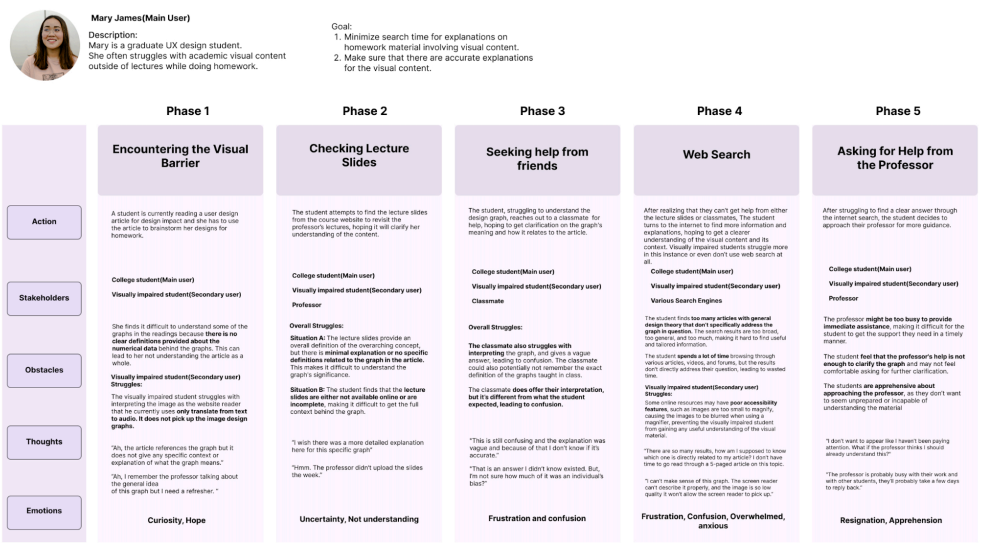
From my surveys and interviews, there was an overwhelming majority who had struggled at some point with visual content, 72.2% in the general student population survey, all participants

in the visually impaired population survey, and both my interviewees responded yes. The top three current techniques and tools they use are re-reading lecture slides, using online search engines for explanations, or asking peers/classmates/professors for their input. Of the visually impaired population, 50% of them use a physical magnifier while the other half use some sort of screen reader for their digital devices. When asked what type of assistant they want, 72.2% of the general student survey and 75.5% for the visually impaired population responded that they want an audio-based explanation. However, while an audio-based tool is preferred, it seems that a lot of students also want the assistant tool to have either an interactive element or more in-depth explanations of the visuals in the program.

As I explored the insights I've gained from my user research and experience journey, I realized that many of the three main pain points stem from the current limitations of the tools and techniques the students use. The three pain points are:

1. Lack of Accuracy & Vague Explanations when students are asking fellow students for answers.
2. It is time-consuming.
3. Difficulty in Finding Specific & Relevant Explanations, primarily in the web search process.

One of the pain points, lack of accuracy, comes from students not feeling fulfilled or satisfied with vague answers given by fellow classmates or professors. Another pain point revolves around searching for explanations through online search engines and re-reading lecture materials. Many students invest significant time in trying to locate accurate answers or relevant articles, which leads to feelings of anxiety and time constraints. And, even when they come across a semi-relevant article, they need to spend even more time reading it to fully grasp the concept. So students want a program that is efficient and straight to the point. This problem is worse with visually impaired individuals who would have to manually look through images with magnifiers. The final point is the difficulty in finding specific and relevant explanations, which stems from the limitations of search engine technology. Students often want to specify and explain a specific part of the graph, current search engines like Google reverse image search, offer a general description but it requires a lot of back and forth to get the specific information of the graph that they want.

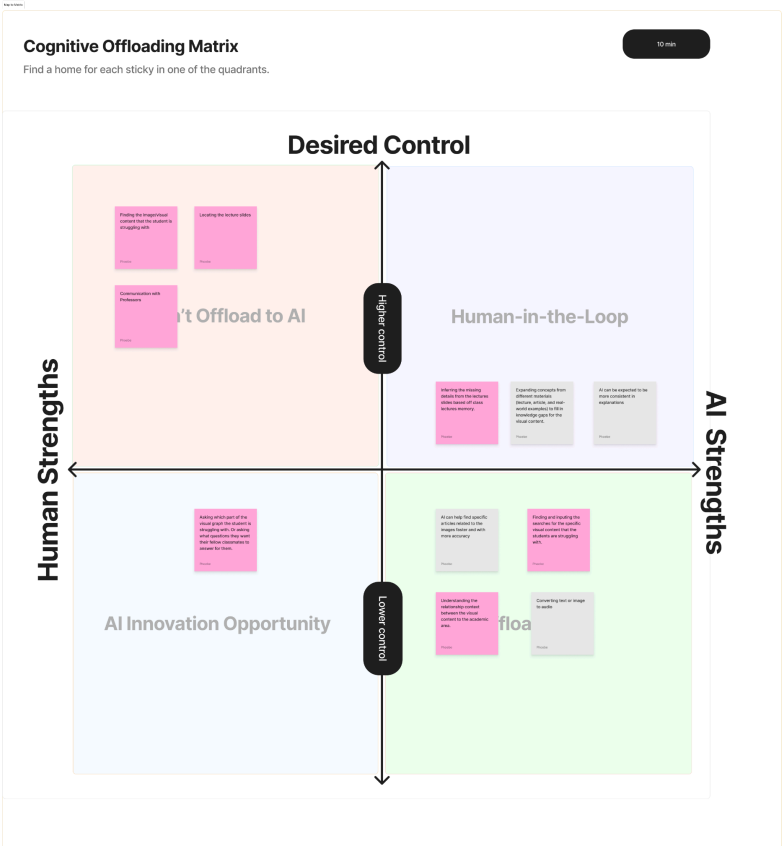


Cognitive Offloading:

For my cognitive offloading, I decided to focus on each phase of the user experience and brainstorm which part of the phase needs AI offloading. For cognitive offloading to humans, I wanted them to still make the decision-making process and I wanted students to keep their "freedom of choice". Students should be able to still choose which images they feel they struggle with over using AI to predict which image they think a student is struggling with. I also realized, as I was going through the process, that finding and analyzing academic sources, and connecting and expanding explanations from different sources are the two moments best suited for AI.

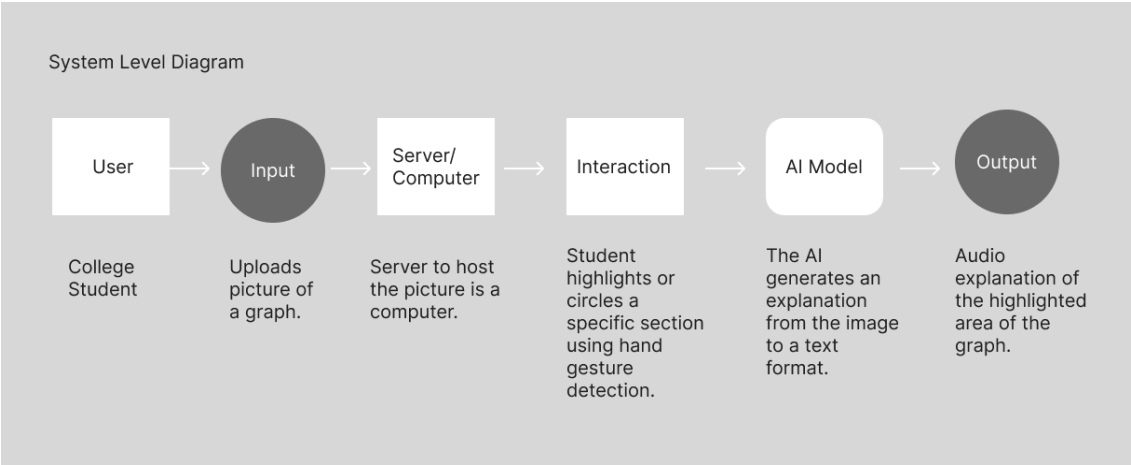
As mentioned in my user research, time constraints are one of the key pain points during the user experience of students trying to figure out visual content. Students, when manually searching and filtering articles and videos, find that it takes up too much time and often, they struggle with even finding the actual relevant information. AI can search through hundreds or even thousands of articles, images, or videos in large databases incredibly quickly. Students won't need as much control over the AI as they mainly want a more efficient and automated process. Even if the articles that the AI finds don't exactly match what the students want, it will still save students a lot more time. Students can also simply ask the AI to start the process over again.

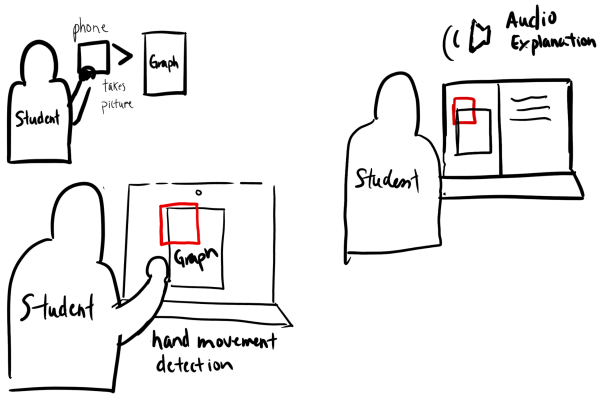
Another moment that is well-suited for cognitive offloading to AI is the process of connecting and expanding explanations from various sources throughout the user experience journey. While AI is not the best for remembering explanations in past lectures, it is, however, good for synthesizing information from those lectures and applying it either to graphs/diagrams or expanding the information. Again, I feel that this is the case because, similar to finding articles, AI is better suited for gathering a large amount of data with very little time.



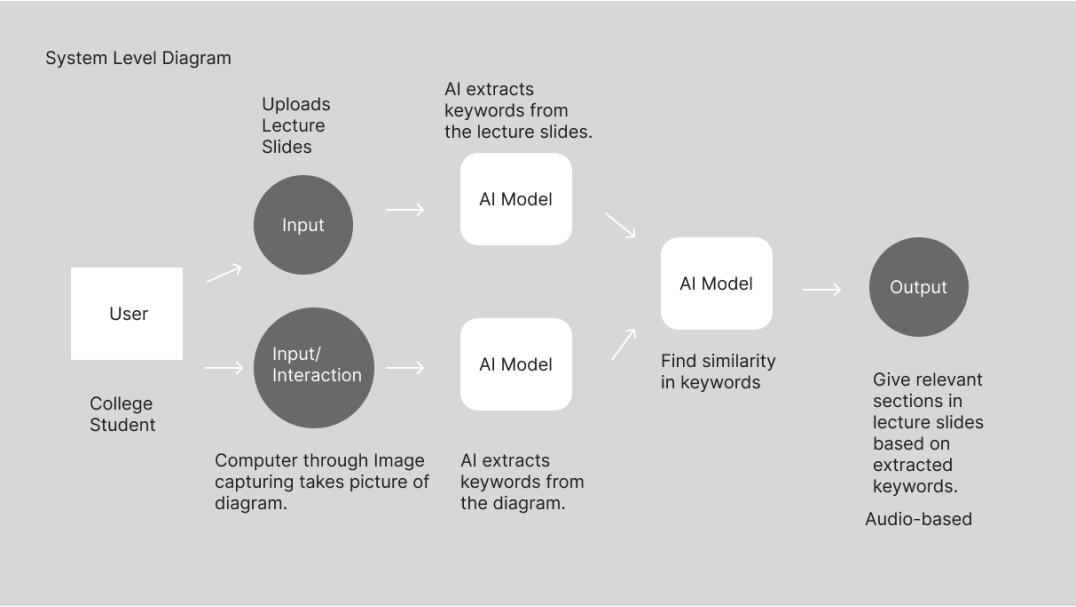
Three Proposed Solutions:

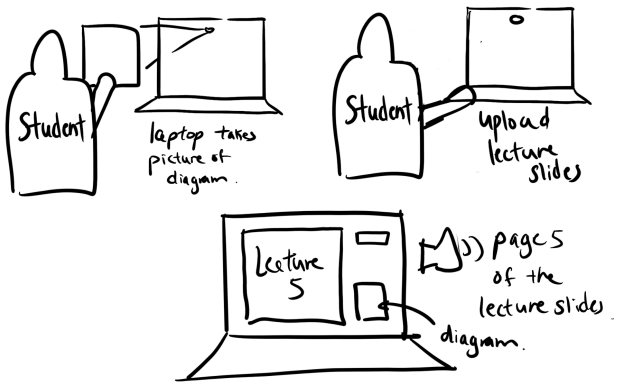
1. My first proposed solution is what I call an AI-Interactive Explanation Tool. This solution targets the pain point of lack of specific explanations. From my user research, and the user experience map, a lot of students commented on wanting to learn about one part of a data graph versus putting it through a Google reverse image search and getting a very general description of the whole image. This way, students are able to learn more about the data, in-depth and not just simply surface leveled. The AI-Interactive Explanation Tool is targeted specifically at college students who have a degree with large amounts of numerical data, and majors such as mathematics, biology, and physics. I had to narrow my idea down, because while I was testing different graphs from different subjects, I found that AI models, like ChatGPT, work better with something that is concrete, it limits the probability of bias in their explanations. For this solution, students, after lectures and while doing homework, can upload a graph they are struggling with to the program on the computer, and they can highlight or circle a specific section of the graph and the AI will provide an audio description of that specific part of the graph.
   The LLMs I need for the prototype would be an image-to-text or text extraction model that specializes in translating numerical data graphs to readable text. I did find one from hugging face,[microsoft/Phi-4-multimodal-instruct](), which, from the descriptions, transcribes mathematical graphs to text pretty well. I also need machine learning to detect hand movements for the circling interaction. I need a text-to-speech model to translate the image-to-text into an audio format.





2. The AI-Academic Synthesizer addresses the pain point of time constraints during the re-checking lecture note process. Students struggle with finding and connecting the relevant lecture notes to a specific diagram due to time constraints. AI is also better suited to expand and fill in knowledge gaps with different materials because it offloads and automates the action of students searching for explanations. When students need to re-check lecture slides and synthesize information, the laptop can scan diagrams into the program. The input does have a limitation where the diagrams have to be centered around text. At the same time, the students will also upload their past lecture slides. The program then extracts any text and keywords from the diagram while also doing the same thing with the lecture notes. If the program detects similar keywords/text, it will give the specific page of the lecture that contains these keywords and the text in the lecture slides and explain it through audio.
   If I were making a prototype of this, I would need a computer-vision image-to-text model that specializes in extracting text from images. For the lecture notes, I would need a NLP, text classification model, where it can label or categorize a page of the lecture. I would also need a text-to-speech AI model where it generates the texts into audio.





3. The most speculative of my three solutions is a wearable design that connects a smart eyeglass and visual content. It is designed for college students who need fast, real-time explanations of graphs and visual content after lectures. Students use voice control to allow the wearable smart glasses to pick up the image in front of them and the wearable glass can use audio to explain the images back to the students. Students can then ask the eyeglass to take a picture and upload them to a computer program, in that computer program where it scans the text and retrieves recommended research articles. If the student is visually impaired, the program will speak the name of the articles to them and if the student is not visually impaired, the eyeglass will give the website name through text. The pain point this solution tackles is the web search process of the user experience journey and specifically lack of accurate sources and explanations. Many times, students struggle with finding relevant and accurate information so this solution is created to help them quickly find research articles relevant to their images.
   To prototype this, I would need a smart eyeglass, something similar to the Meta Ray-ban smart glasses. The Meta Ray-ban glasses can take pictures and they already incorporate audio control and Meta AI. For the computer program I would prototype, I would need NLP text extraction and also an article retrieval AI system. I did find a list of APIS for scholarly research on the University of Calgary's website. Like the three other solutions, I need a text-to-speech model for the audio.