Mid-Term Proposal
UX and AI

Phoebe Pang

What is my main problem?

This project was inspired by me noticing how much visual content is prevalent in academics. Many college students struggle with understanding complex visual content in academics. Sometimes traditional learning tools don't cater to non-visual learners and those who are visually impaired.

My main goal:
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

Two Surveys:

- First survey: focused on current or newly graduated students from college.
- Second survey: targeted the visually impaired community

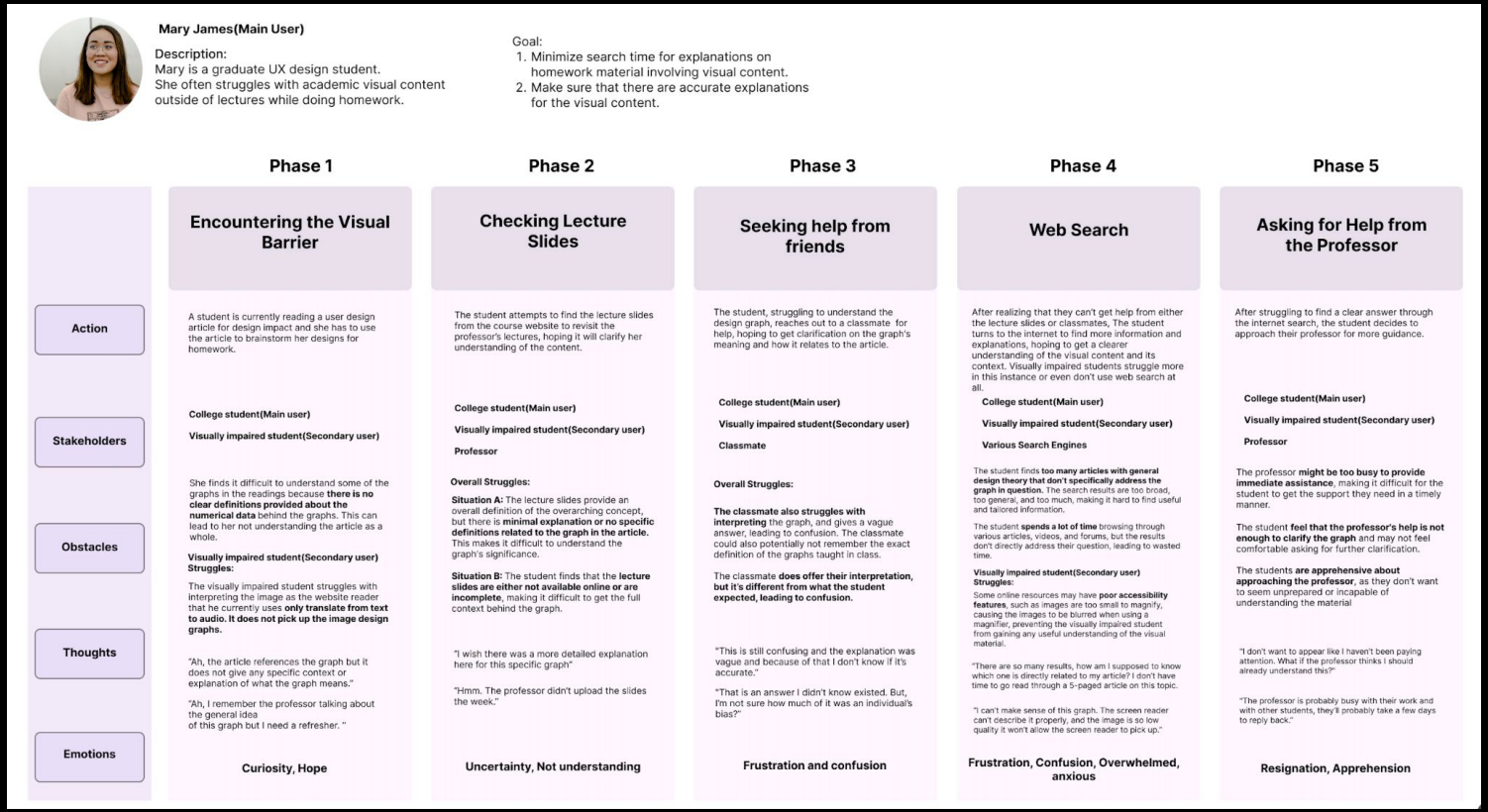
Two Interviews:

- One who is a regular student and one who is a visually impaired individual.

Insights:

- There is a high demand for audio-based assistance
- People want detailed step-by-step explanations of the image.
- Interactivity can become a feature.
- People might prefer an AI-assistant tool because they believe they are not as biased as real-life people.
- People want a quick and easy way to upload PDFs, images, or real-time captures to the AI assistant.

User Experience Map

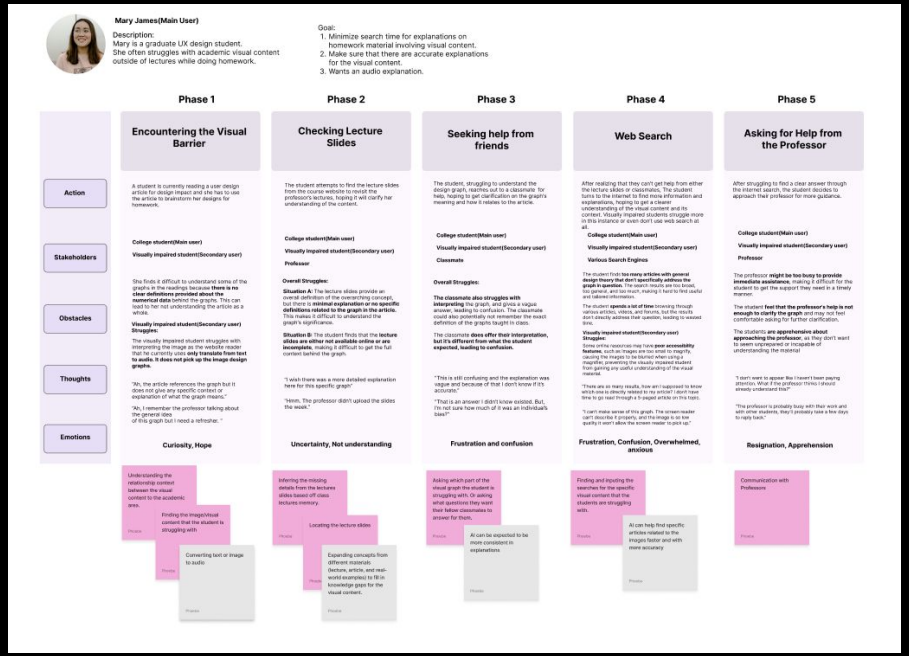


3 Pain Points

1. Lack of Accuracy & Vague Explanations
    a. This is mostly centred around peer/fellow classmate feedback in the user journey where the student asks around to figure out visual content.
2. Difficulty in Finding Specific & Relevant Explanations
    a. When searching for visual content, search engines often give results that are not specific to the exact image or graph in question, making it difficult to find relevant insights.
3. Time-Consuming
    a. Students spend a lot of time re-reading lectures just to understand a single visual graph/content.
    Students waste time manually trying to filter through multiple sources (articles, videos..etc) to find the right explanation.

Cognitive Offloading

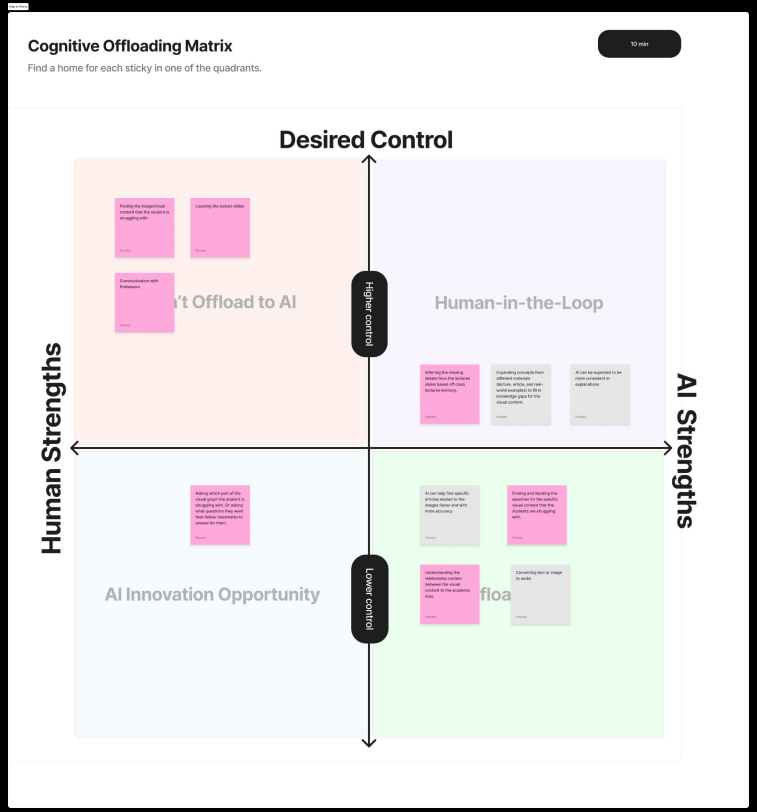
- I decided to focus on each phase of the user experience and brainstorm which part of the phase needs AI offloading.
- Human Offloading:
    - I wanted them to still make the decision-making process and I wanted students to keep their “freedom of choice”



Cognitive Offloading

Two moments:

- Finding and analyzing academic content
  - Reduces time spent on manual searches
- Connecting & Expanding Explanations from Different Sources
  - AI processes large datasets and it can reduce the cognitive load of manually piecing together information.



Solution 1: AI-Interactive Explanation Tool

What: Image to Text interactive tool that helps students eliminate the issue of broad and vague statements or explanations. It help targets specific areas of the graph.
Who:

- Main users: College students in data-heavy majors.
- Secondary users: Visually impaired students

When:

- After lectures/during homework when students encounter a graph they don't fully understand.

How:

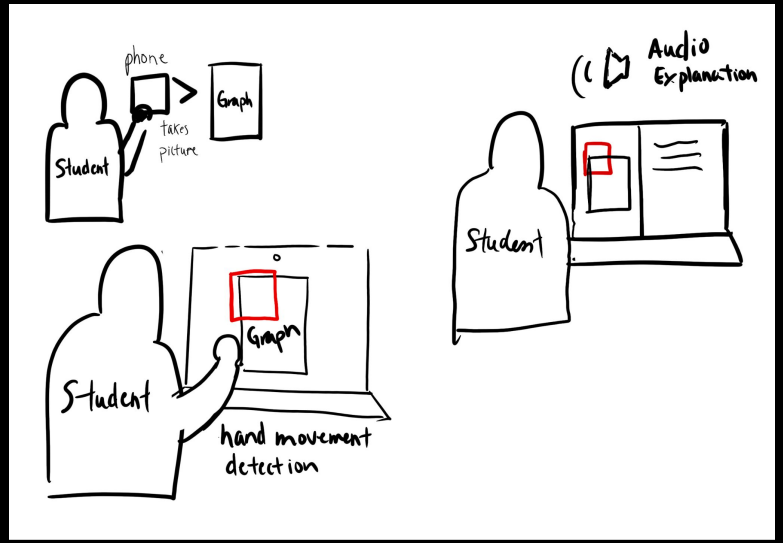
- Students upload their graph/chart they don't understand and then circle a specific area that they need AI to explain.
- AI will then analyze and give an audio explanation

Tech: Image-to-text Computer Vision Models, Text to Speech Models(Hugging Face Kokoro-82M), Machine Learning hand gesture detection

Solution 1: AI-Interactive Explanation Tool

System Level Diagram

User	Input	Server/ Computer	Interaction	AI Model	Output
College Student	Uploads picture of a graph. Active	Server to host the picture is a computer. Passive	Student highlights or circles a specific section using hand gesture detection. Active	The AI generates an explanation from the image to a text format. Passive	Audio explanation of the highlighted area of the graph. Passive



Solution 2: AI-Academic Synthesizer

What: Extracts keywords and finds direct connections between diagrams and past lectures, reducing the time spent searching for explanations.

Who:

- Main users: College students in data-heavy majors.
- Secondary users: Visually impaired students

When:

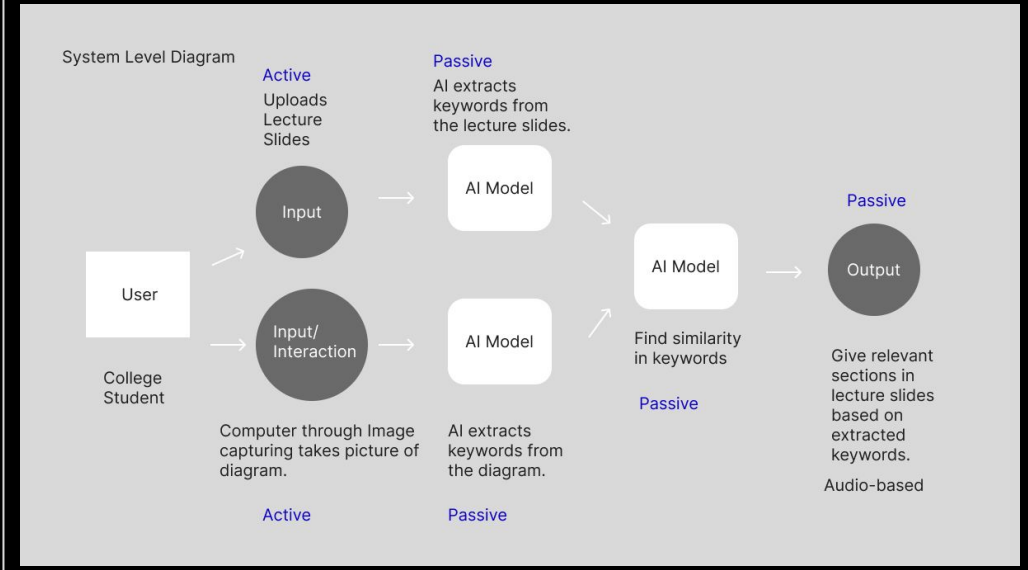
- When students are reviewing notes but struggle to locate or synthesize explanations from previous materials.

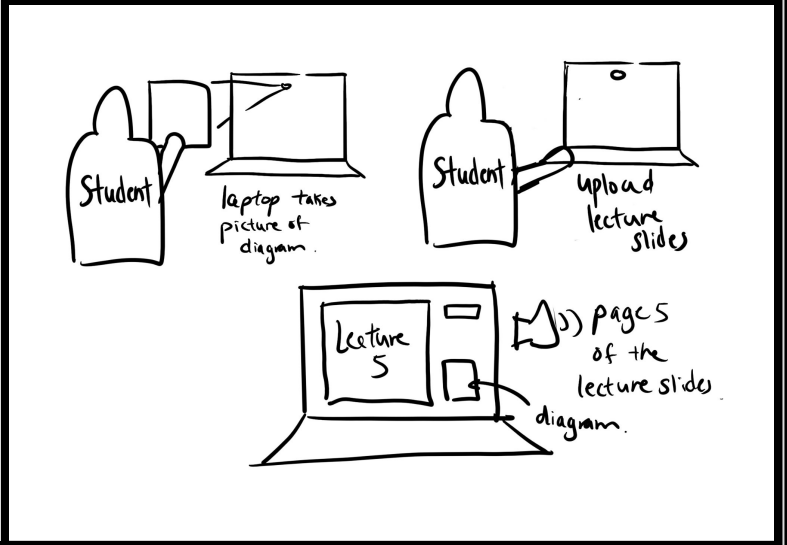
How:

- Students scan/upload a diagram into the AI program and also upload a copy of their class lectures
- The AI extracts any text and keywords from the diagram and also from the class lectures.
- If the texts and keywords matches, the AI will identify the page of the lecture through audio

Tech: Image-to-text Computer Vision Models(text extraction), Text to Speech Models(Hugging Face Kokoro-82M)

Solution 2: AI-Academic Synthesizer





Solution 3: Wearable Smart Eyeglass

What: A wearable smart glass system that provides instant explanations of visual content and retrieves relevant research articles.

Who:

- Main users: College students in data-heavy majors.
- Secondary users: Visually impaired students

When:

- When students who want instant explanations of visual content.
- When students struggle with manual searches for relevant academic sources.

How:

- Students wear smart glasses (Meta Ray-Ban) and use voice commands to capture an image of the visual content they need explained.
- Students instruct the glasses to take a picture and upload it to a connected computer program.
- The computer program scans the image, extracts keywords, and retrieves relevant research articles.
- The AI reads aloud the titles and authors of the recommended research articles.

Tech: Smart Glasses(Meta Ray-Ban), Image to text machine learning models, AI model that can search research articles(API's for scholarly articles)Text to Speech Models(Hugging Face Kokoro-82M)

Solution 3: Wearable Smart Eyeglass

System Level Diagram

User	Interaction	Server/ Computer	Input	AI Model	Output
College Student	Students trigger actions through audio: explanation and picture via smart eyeglass Active	Smart Eyeglass	Picture and text explanation of graph uploaded to computer program Passive	Finds academic papers based on extracted text. Passive	Title/website of the articles Can be audio based or text Passive