General overview:

The idea is to use a smartphone's camera for real-time object detection and send the detected information to an LLM model to generate natural descriptions of events. These descriptions are then converted to audio, providing visually impaired individuals with real-time awareness of their surroundings and enhancing their independence. This easily relates to some of the UN Goals.

Goals 3 & 10

General objectives:

* Medium:
* Languages:
* Platforms?:
* Quantify success?
* Any competition / precedents?: [OpenAI](https://platform.openai.com/docs/guides/vision) (prototype?)

Questions/Concerns (for the TA)?:

Tentative schedules:

Notes:

Potentially we can make it so that the person asks questions to the app and it responds with knowing detailed info of the environment. When a person asks a question, the program takes a screenshot and then sends it to an llm.

Tech Stack

* FE: Vue, possibly able to convert from webapp to mobile
* DB: Supabase or Neon -> should be easier to set up.

Possible tech stack by Daniel:

* Frontend: NextJS
* DB: Supabase
* LLM that searches and describes the image: openai
* Text to speech: Elleven labs or openai

Consider: Next.js (JS/React/Typescript, TailwindCSS) - full stack environment, Prisma ORM (for database) modeling

- we can use SQLlite, mySQL, or postgres with it.