# DETECTING LANDCOVER FROM SATELLITE IMAGES USING ELLMER

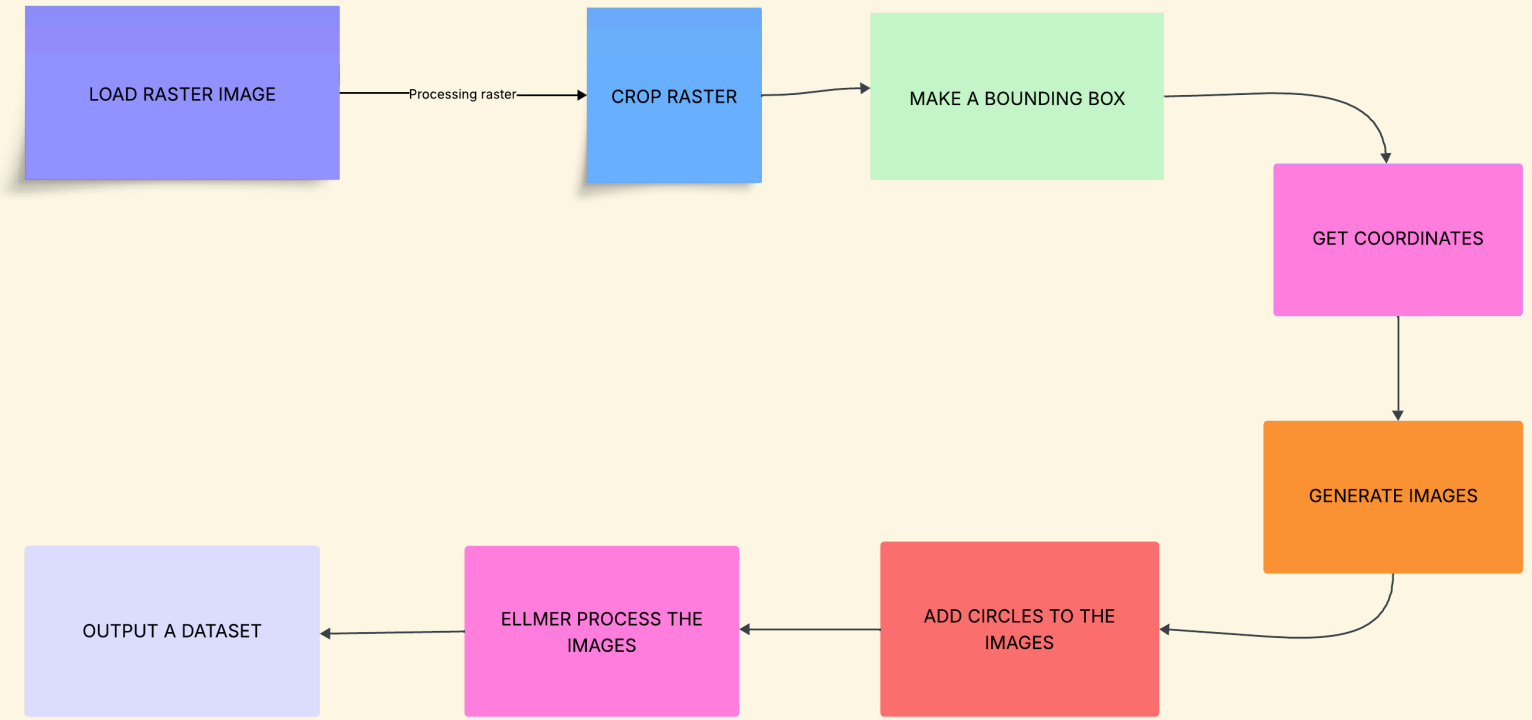
Odilon Ligan
Jean Marie Ngabonziza
Kelly McConville
Grayson White
Rebecca Beneroff
Andrew Lister

# TABLE OF CONTENTS

# THE PURPOSE OF THE PROJECT

The National Agriculture Imagery Program (NAIP) have satellite images.

- Leverage Large Language Models(LLMs) to assist in interpreting those imageries.
- Prototype a fast, flexible workflow

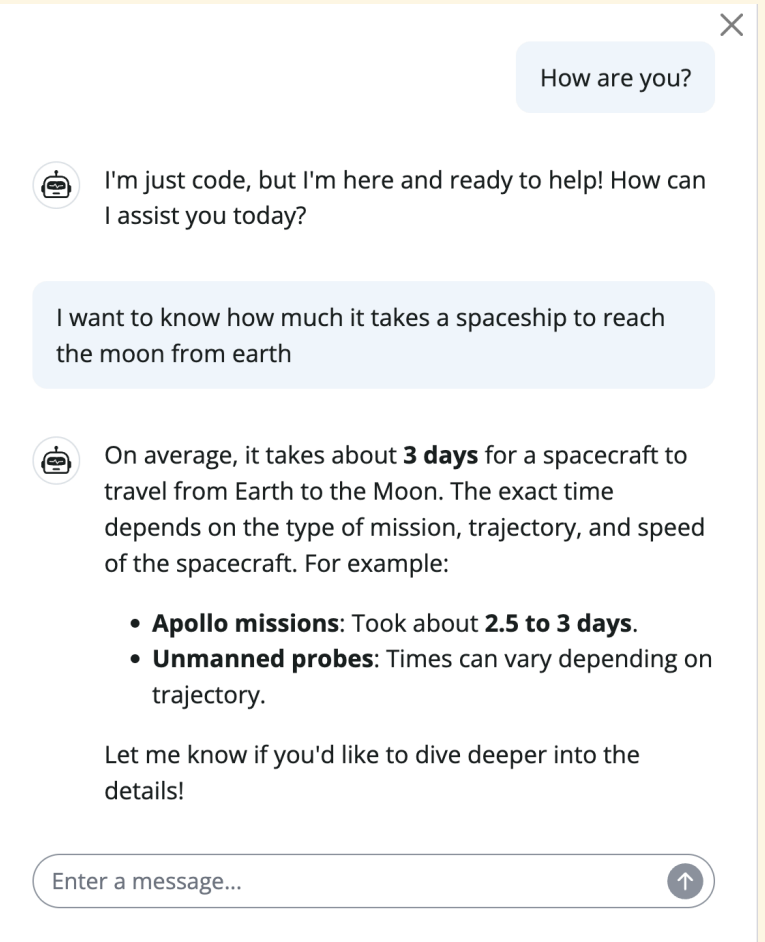


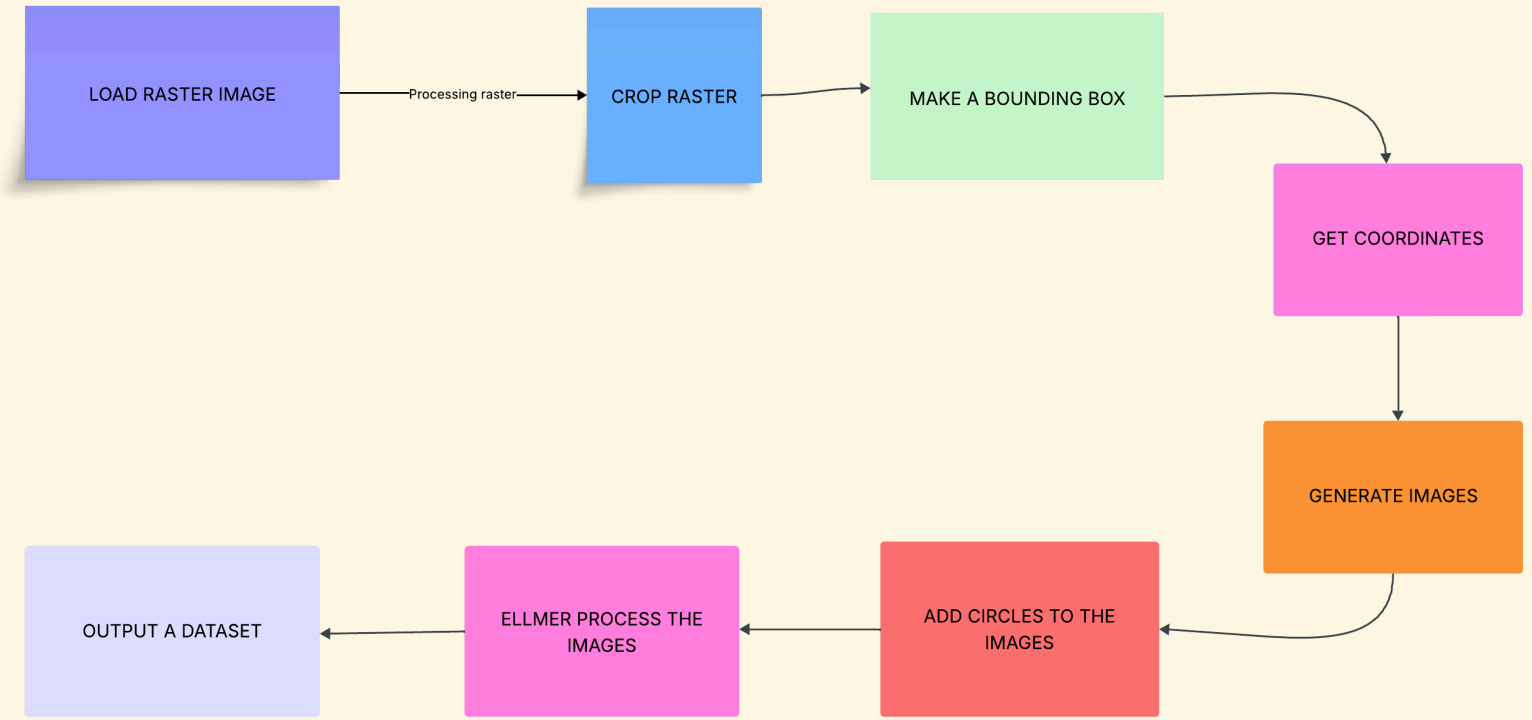

# Workflow Overview

# WHAT IS ELLMER?

Logo

- R package to interact with LLMs
- Prompts and responses directly into R
- Enables us to control the output we want
  - Dataframe

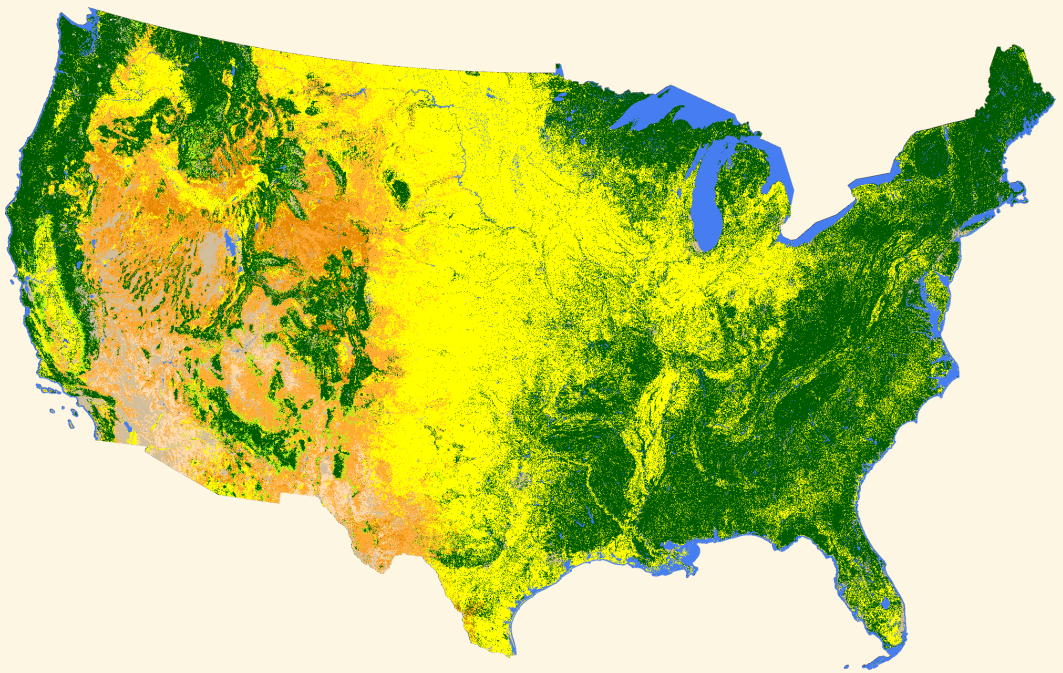


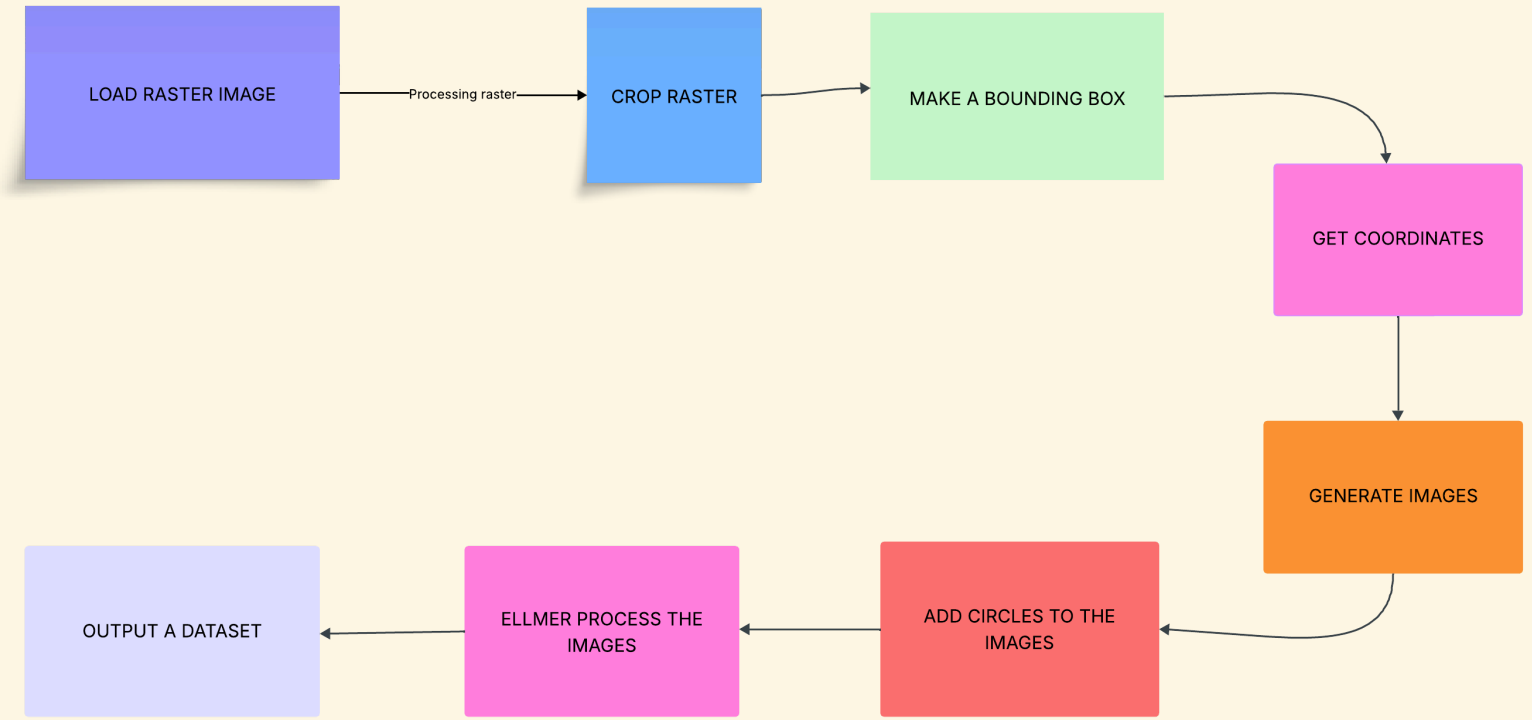



# Workflow Overview

# PROCESSING RASTER IMAGE

```r
# Load raster interactively
raster.img <- raster(file.choose())
```

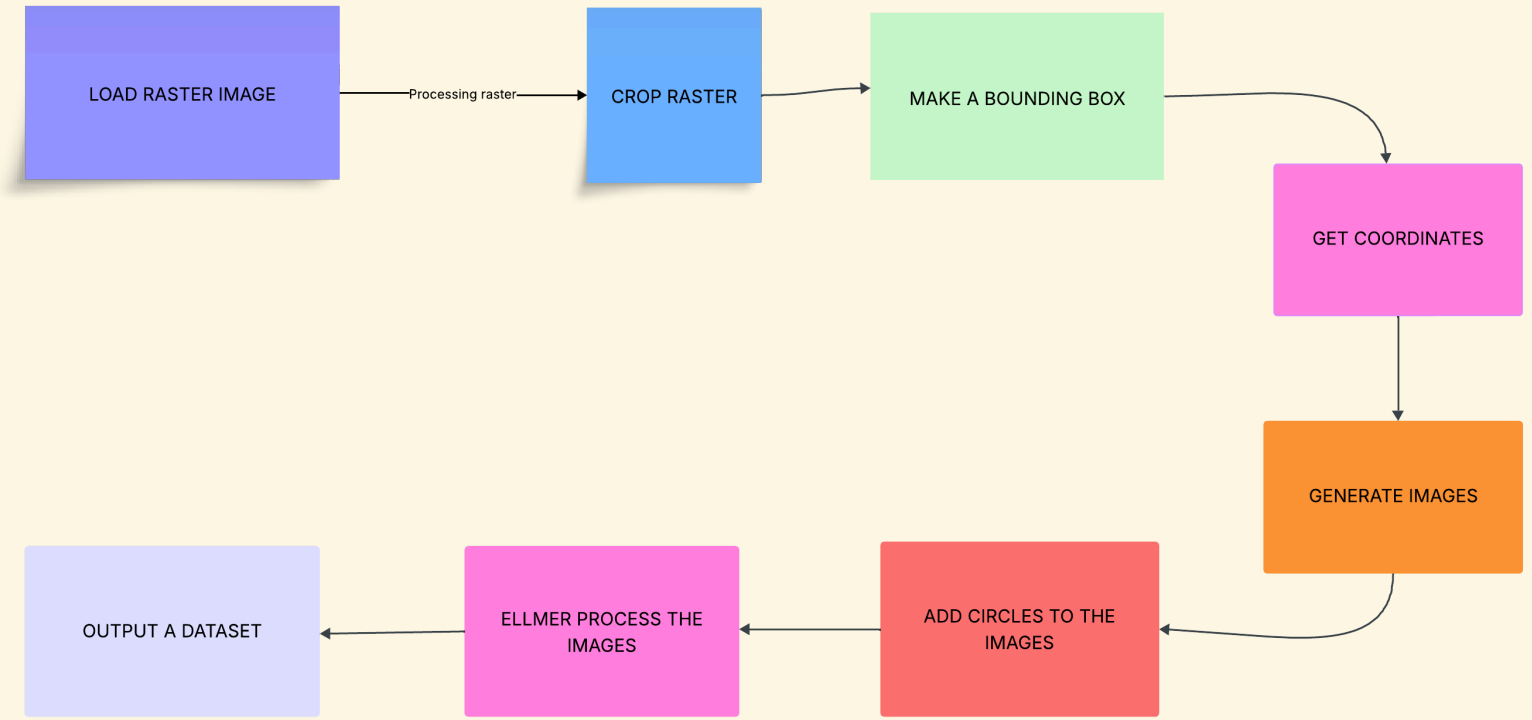





# Workflow Overview

# CROPPING RASTER TO SELECTED REGION

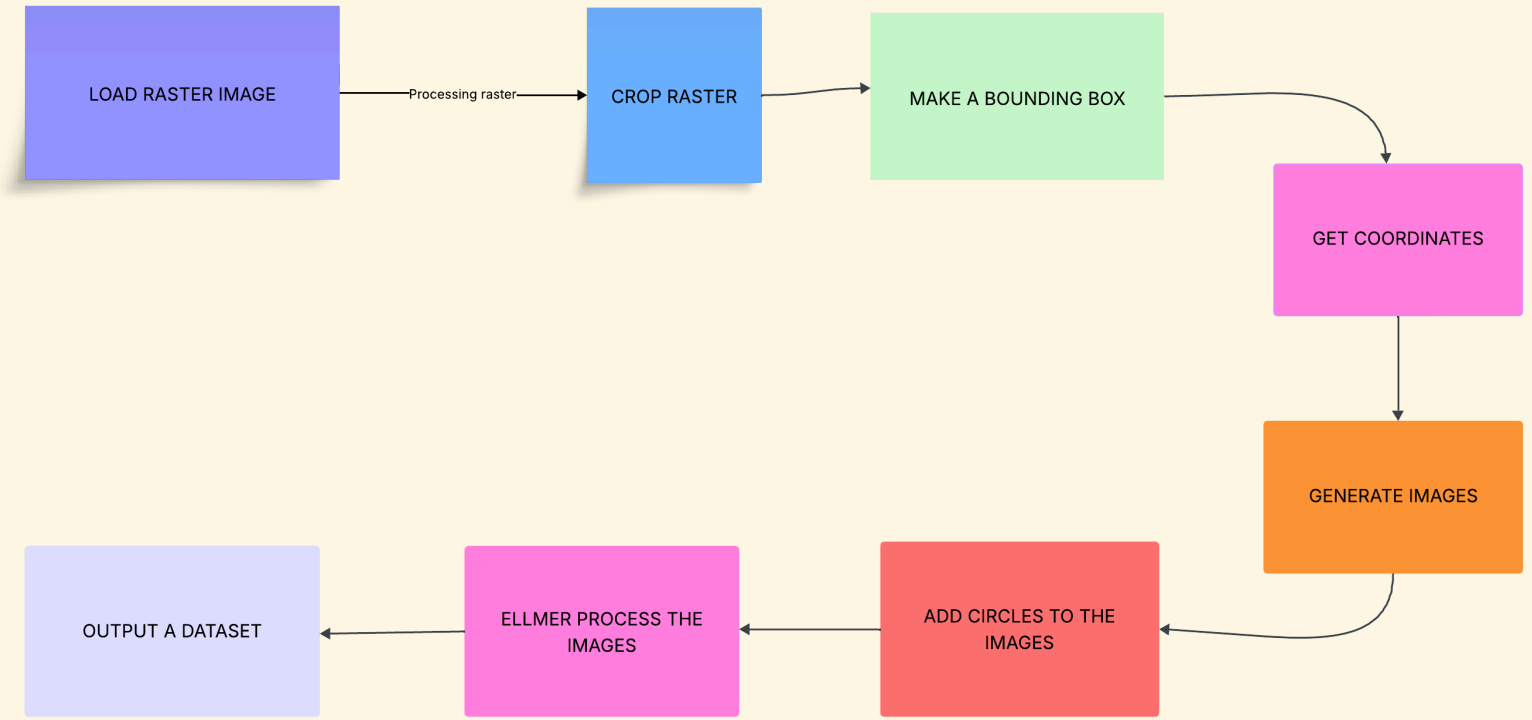
```
# 1. Define a bounding box (area of interest) in geographic coordinates
bbox_wgs84 <- st_bbox(c(xmin = -80.52071966199662,
            ymin = 39.720265072840725, xmax =-74.69529551145511,
            ymax = 42.269954234532335), crs = 4326)

# 2. Reproject the bounding box to match the raster's coordinate system
bbox_albers <- st_transform(bbox_wgs84, crs(raster.img))

# 3.  Crop the raster using the reprojected bounding box
raster.crop<- crop(raster.img, bbox_albers)
```




# Workflow Overview

# OBTAINED COORDINATES NEEDED BY NAIP

```r
#4. Stratified sample
  #obtain the classification of the pixels
  pixel_values <- getValues(raster.crop)

sample_pixels <- sampleStratified(raster.crop, size = 10,
                  strata = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14))

#5. Extract the pixel values at the sampled cell indices.
sampled_values <- values(raster.crop)[as.data.frame(sample_pixels)$cell]

#6. Get the spatial (x, y) coordinates of the sampled cells
sampled_coords <- xyFromCell(raster.crop, as.data.frame(sample_pixels)$cell

#7. Converting a point to a bbox
from_point_to_bbox <- function(x, y) {

  # Get CRS from raster
  raster_crs <- st_crs(raster.img)
```

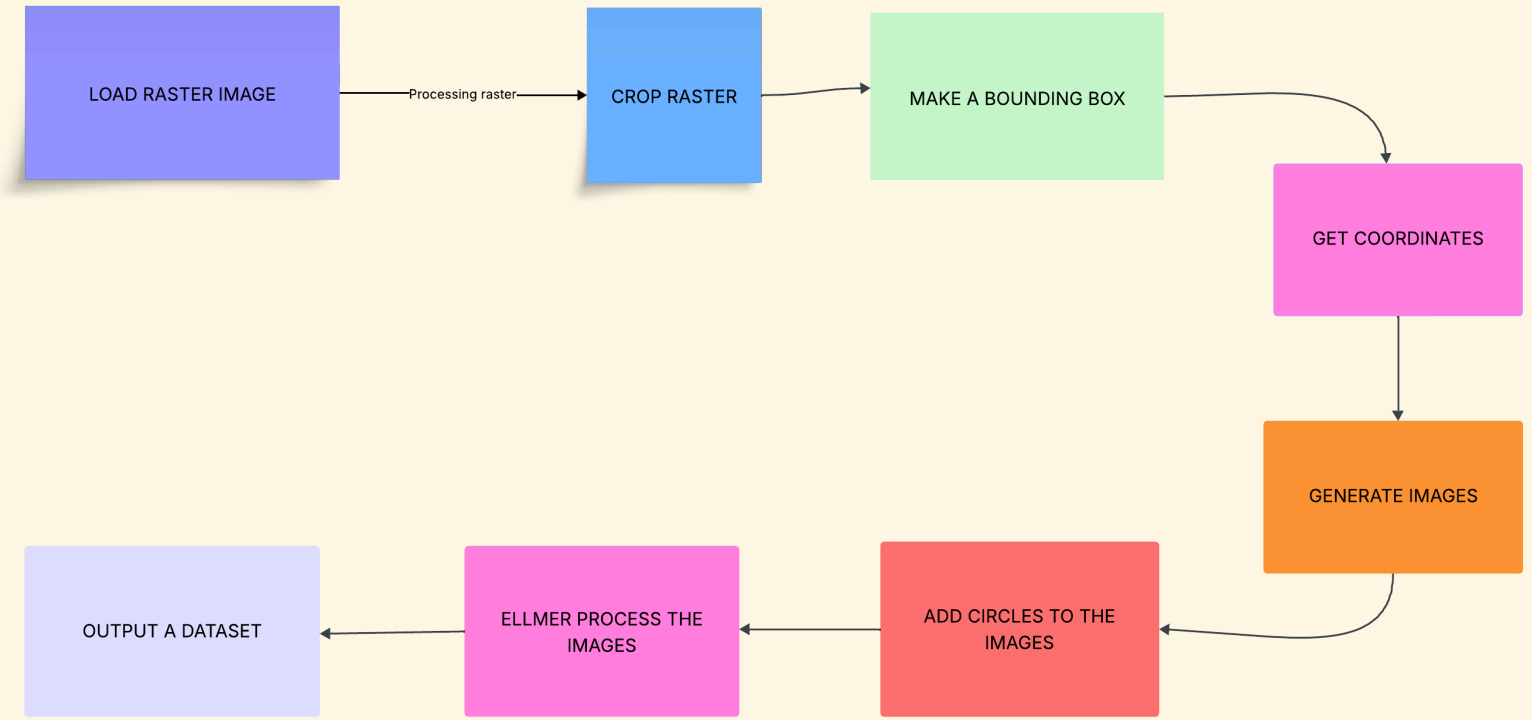



# Workflow Overview

# GENERATE IMAGES

- A function acess_image to build NAIP API URLs using bbox values and save PNG images locally.
- A function add_circle_to_image to mark a territory of interest on the images.

Original Image 




# Workflow Overview

# SETTING UP ELLMER CHATBOT

- System prompt to guide the LLM on what to look for
  - Detailed, output format, and identifying classes.

```
# Initialize chat with system prompt
image.analyze <- chat_github(
  model = "gpt-4o",
  system_prompt = "
 You are analyzing satellite images to determine the type of land cover in

 Classify the land cover in the circle using the following categories. For
 level using one of these options: 'Confident', 'Somewhat confident', 'Some

 Land Cover Classification Categories:

 Water
 - 11: Open Water – Areas of open water, generally with less than 25% cover
 - 12: Perennial Ice/Snow – Areas characterized by a perennial cover of ice

 Developed
 - 21: Developed, Open Space – Mostly vegetation (lawns), with impervious s
```

# DETECTING LAND COVER WITH ELLMER

- Convert output to JSON
- Then make it a data frame

```
# Define the land cover analysis function
analyze_land_cover <- function(image_path) {
  # Structured chat interaction
  response <- image.analyze$chat_structured(
    content_image_file(image_path),
    "Analyse the image and give the answer in json format and I want just o
    Somewhat confident, Somewhat not confident, or Not confident. I don't w
    you can not identify the landcover just put NA. And give precisely, to
    type = type_object(
      Land_cover_type = type_string(),
      confidence = type_string()
    )
  )
  # Convert response to JSON and extract data
  json_response <- toJSON(response, pretty = TRUE, auto_unbox = TRUE)
  # Convert to data frame
  df <- data.frame(
```

# OUTPUT OF THE PROJECT

Sample of my_dataframe

| xmin | ymin | xmax | ymax | Pixel | NLCD_Class | landcover | confidence |
|---|---|---|---|---|---|---|---|
| -75.27140 | 39.94556 | -75.03496 | 40.12405 | 149223473 | Barren | Developed, High Intensity | Confident |
| -76.43660 | 41.66671 | -76.19433 | 41.84545 | 50042923 | Planted | Planted/Cultivated | Confident |
| -75.80918 | 41.49185 | -75.56751 | 41.67056 | 54955682 | Planted | Forest | Confident |
| -74.71331 | 40.33583 | -74.47559 | 40.51435 | 117879812 | Forest | NA | Not confident |
| -79.62203 | 42.31122 | -79.37755 | 42.49008 | 37464043 | Water | NA | Not confident |
| -75.45732 | 39.75493 | -75.22150 | 39.93341 | 163565875 | Water | Developed, High Intensity | Confident |
| -76.05469 | 40.32546 | -75.81697 | 40.50400 | 132981256 | Barren | Planted/Cultivated | Confident |
| -75.29682 | 39.01298 | -75.06341 | 39.19142 | 209804149 | Water | Forest | Confident |
| -76.39755 | 40.33389 | -76.15981 | 40.51243 | 135976510 | Herbaceous | NA | Not confident |
| -77.00169 | 40.27071 | -76.76417 | 40.44924 | 146169073 | Herbaceous | NA | Not confident |

# SETBACKS AND FUTURE WORK

- Limited on the number of images to proceed
- Being able to identify bad images
- Deeper investigation in the sub categories of the classifications
- Detect and interpret land changes across time

# THANK YOU

- Professor Kelly McConville, our mentor
- Andrew Lister, our stakeholder