

A dense tropical forest with sunlight filtering through the trees. The scene is filled with various shades of green, from deep forest greens to bright highlights where the sun hits the leaves. The trees are tall and slender, with many vines and smaller plants growing in the undergrowth. The overall atmosphere is lush and vibrant.

# Amazonia Archaeology Discovery

DATA\_FAIR 2026

Maja Somrak

An aerial photograph of a dense Amazon rainforest. The forest is a vibrant green, with a large, circular clearing in the center. The clearing is surrounded by a thick ring of trees, and the ground inside appears to be a mix of sand and sparse vegetation. The surrounding forest is a dense canopy of green trees, with some lighter green areas indicating different types of vegetation or perhaps a clearing. The overall scene is a lush, natural landscape.

# Designing AI for Sustainable Research

## Amazon rainforest

> 6 million km<sup>2</sup>

75% densely forested and largely inaccessible

10,000 ancient structures remain undiscovered \*



# Designing AI for Sustainable Research

## Amazon rainforest

> 6 million km<sup>2</sup>

75% densely forested and largely inaccessible

10,000 ancient structures remain undiscovered \*

---

Can AI help us discover ancient history  
... without cutting down a single tree?

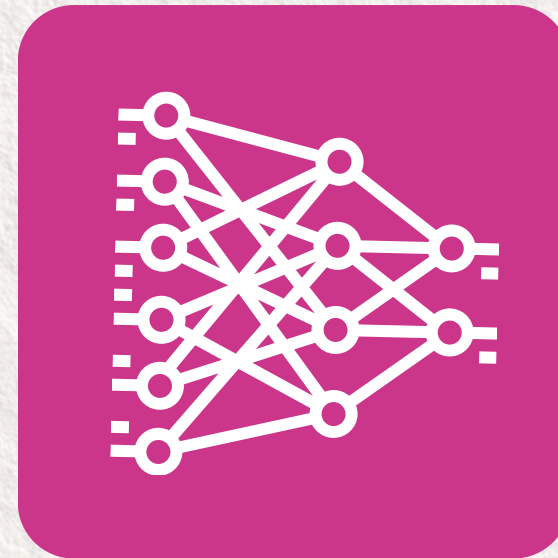
# Designing AI for Sustainable Research

 OpenAI to Z Challenge

“ Use technology to identify possibly hidden archeological sites, using available open-source data ”

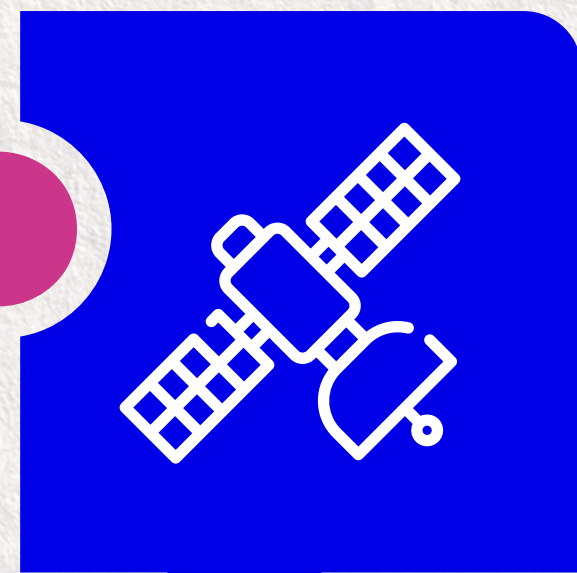
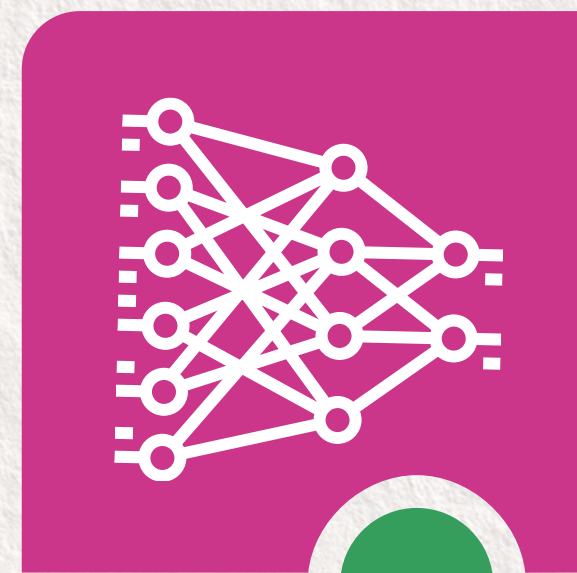
# Approach

**MACHINE  
LEARNING**



# Approach

**MACHINE  
LEARNING**



**REMOTE  
SENSING**

**ARCHAEOLOGY**



**CONTEXT**



# WHAT are we looking for?

Types and characteristics of archaeological features in Amazonia

- Amazon Dark Earths
- Geoglyphs
- Mounds
- Circular villages and other settlements
- Infrastructure earthworks
- Raised fields
- Water reservoirs
- Fish weirs
- Shell middens



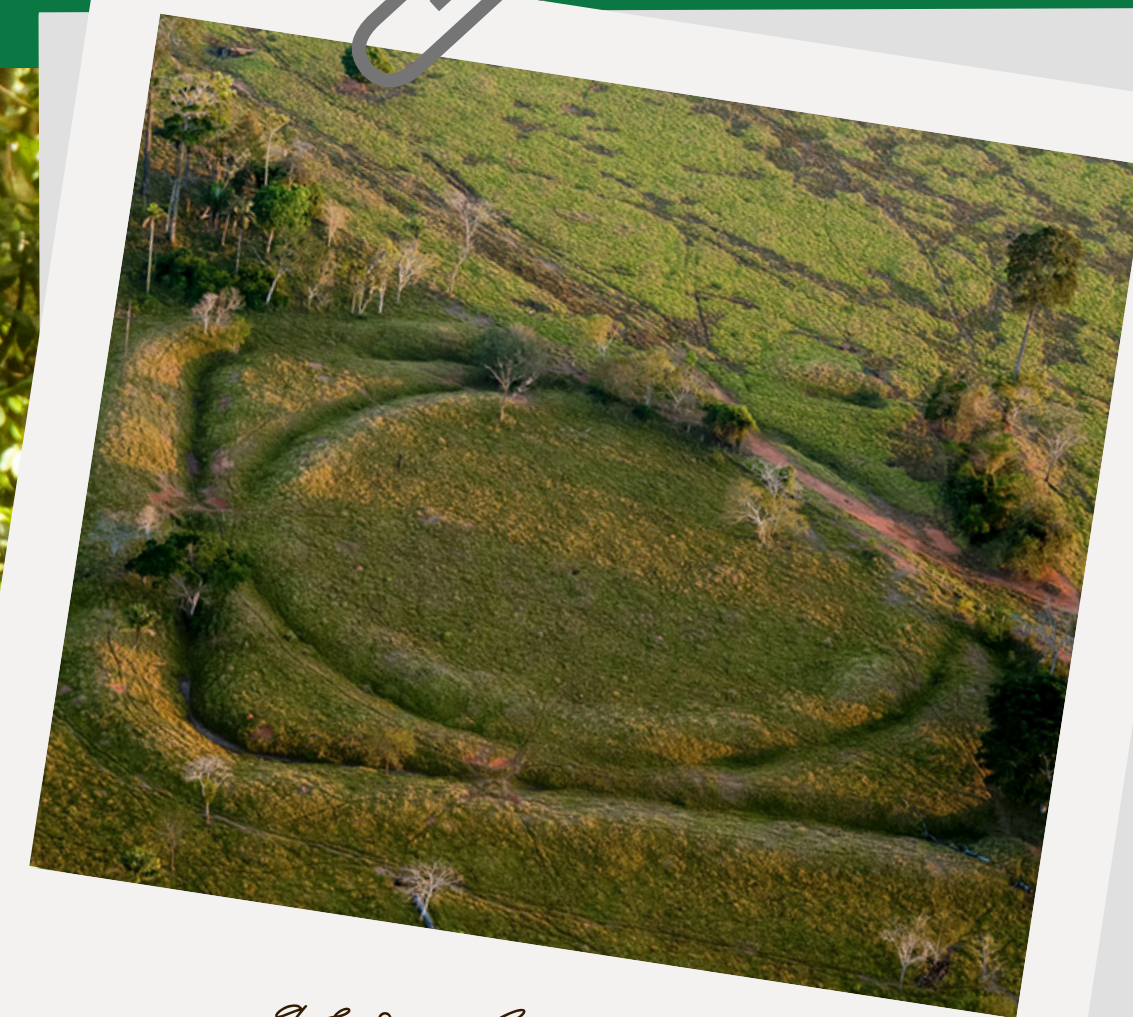
*geoglyph*



# WHAT are we looking for?

Types and characteristics of archaeological features in Amazonia

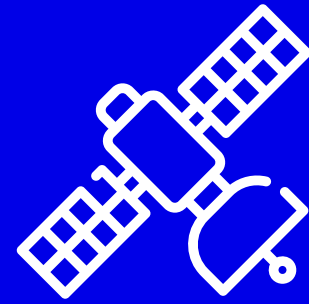
- Amazon Dark Earths \*
- Geoglyphs
- Mounds
- Circular villages and other settlements
- Infrastructure earthworks
- Raised fields
- Water reservoirs
- Fish weirs
- Shell middens



*geoglyph*



**EARTHWORKS**



# HOW are we looking for earthworks?

And why is it so difficult?

## **Optical Imaging (Sentinel-2, Landsat)**

10-30m resolution, full Amazon coverage  
Limited by cloudiness + 74% vegetation cover  
Only effective in 20% deforested areas (already studied)

## **Multispectral Imaging**

### **Hyperspectral Imaging**

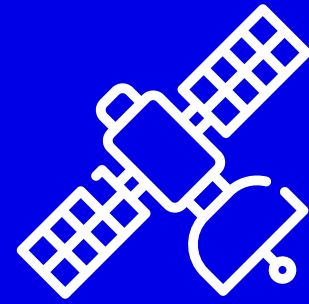
Detects anthropogenic soils (terra preta, shell middens)  
30m resolution adequate for large features only  
<10m needed for small features

## **SAR (Radar)**

All-weather, 1-3m resolution  
Limited vegetation penetration for earthworks

## **LiDAR**

Unparalleled vegetation penetration  
Gold standard for forested archaeology  
Requires 1m resolution (not 30m ANADEM)  
Uses .laz point cloud → DTM conversion



# HOW are we looking for earthworks?

And why is it so difficult?

## Optical Imaging (Sentinel-2, Landsat)

10-30m resolution, full Amazon coverage  
Limited by cloudiness + 74% vegetation cover  
Only effective in 20% deforested areas (already studied)

## Multispectral Imaging

### Hyperspectral Imaging

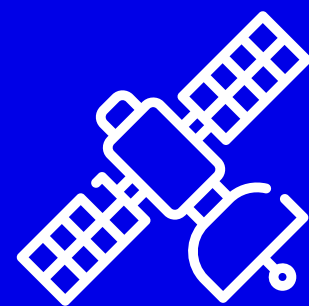
Detects anthropogenic soils (terra preta, shell middens)  
30m resolution adequate for large features only  
<10m needed for small features

## SAR (Radar)

All-weather, 1-3m resolution  
Limited vegetation penetration for earthworks

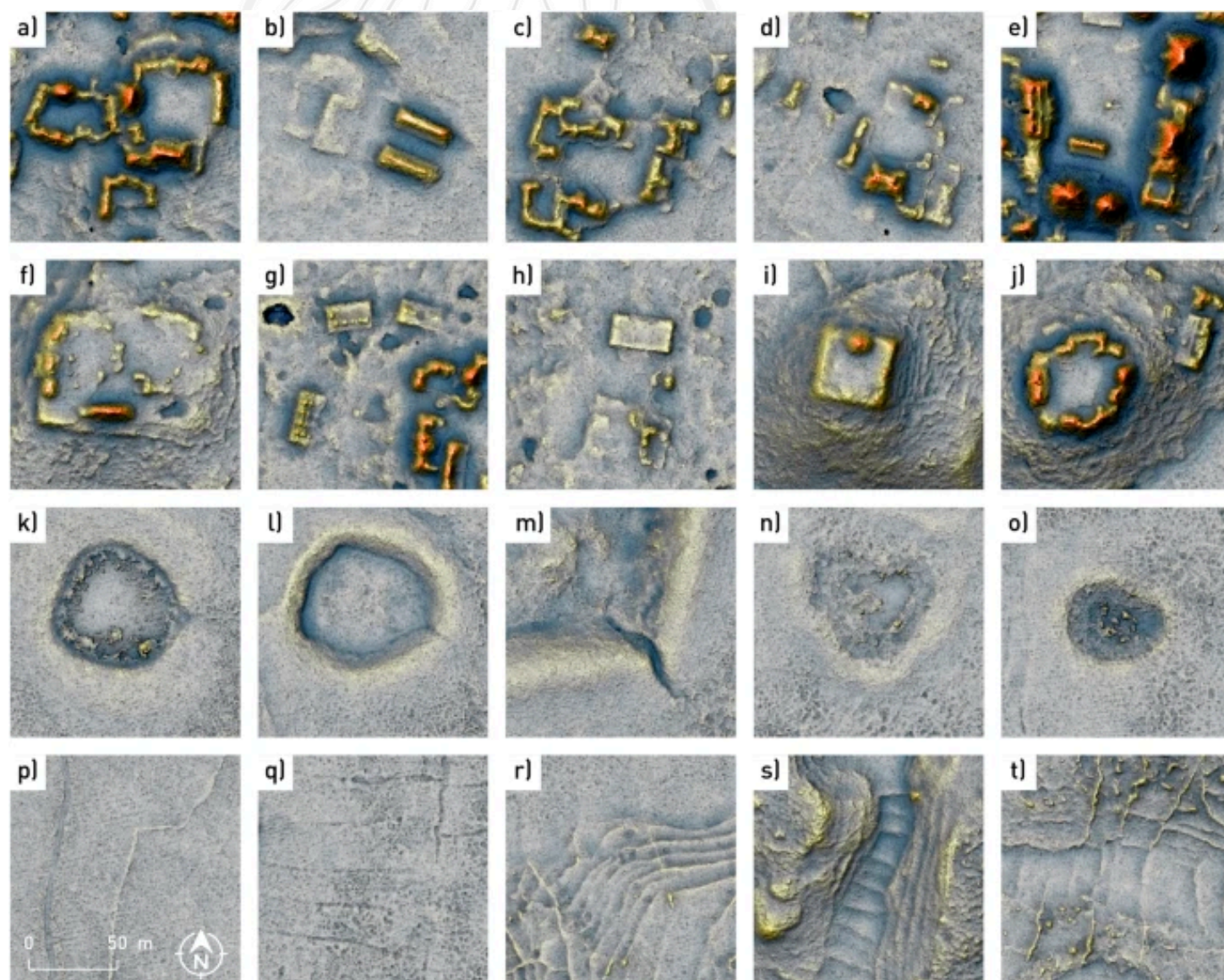
## LiDAR

Unparalleled vegetation penetration  
Gold standard for forested archaeology  
Requires 1m resolution (not 30m ANADEM)  
Uses .laz point cloud → DTM conversion



# Maya Chactun Dataset

Cross-domain learning that overcomes the scarcity of labeled Amazonian training data by leveraging extensive Maya archaeological research



## Total Dataset

2.000+ tiles with DTM visualizations  
high resolution: 0.5m per pixel

## Features

- **buildings**
- **platforms**
- **aguadas**
- **terrain** (everything else)

## Data Quality

Manually identified unlabeled features (e.g. terraces)



# WHERE are we looking for earthworks?

What are the geographic criteria and which areas should we focus on?

## PILOT AREAS

Data samples (Tapajos)

## DISCOVERY AREAS

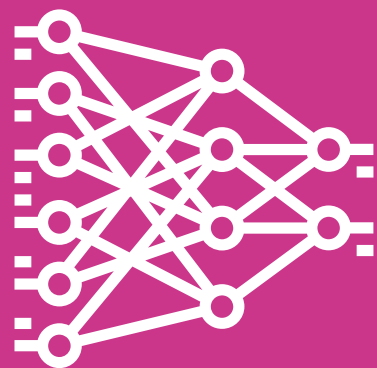
High-resolution LiDAR dataset (NASA):  
*Sustainable Landscapes Brazil*, that covers 3125 km<sup>2</sup>



**Focus on** areas within 10km of water sources



**Exclude** areas with modern civilization presence



# BUILDING a ML model

Teaching a model to recognize ancient landscapes

1

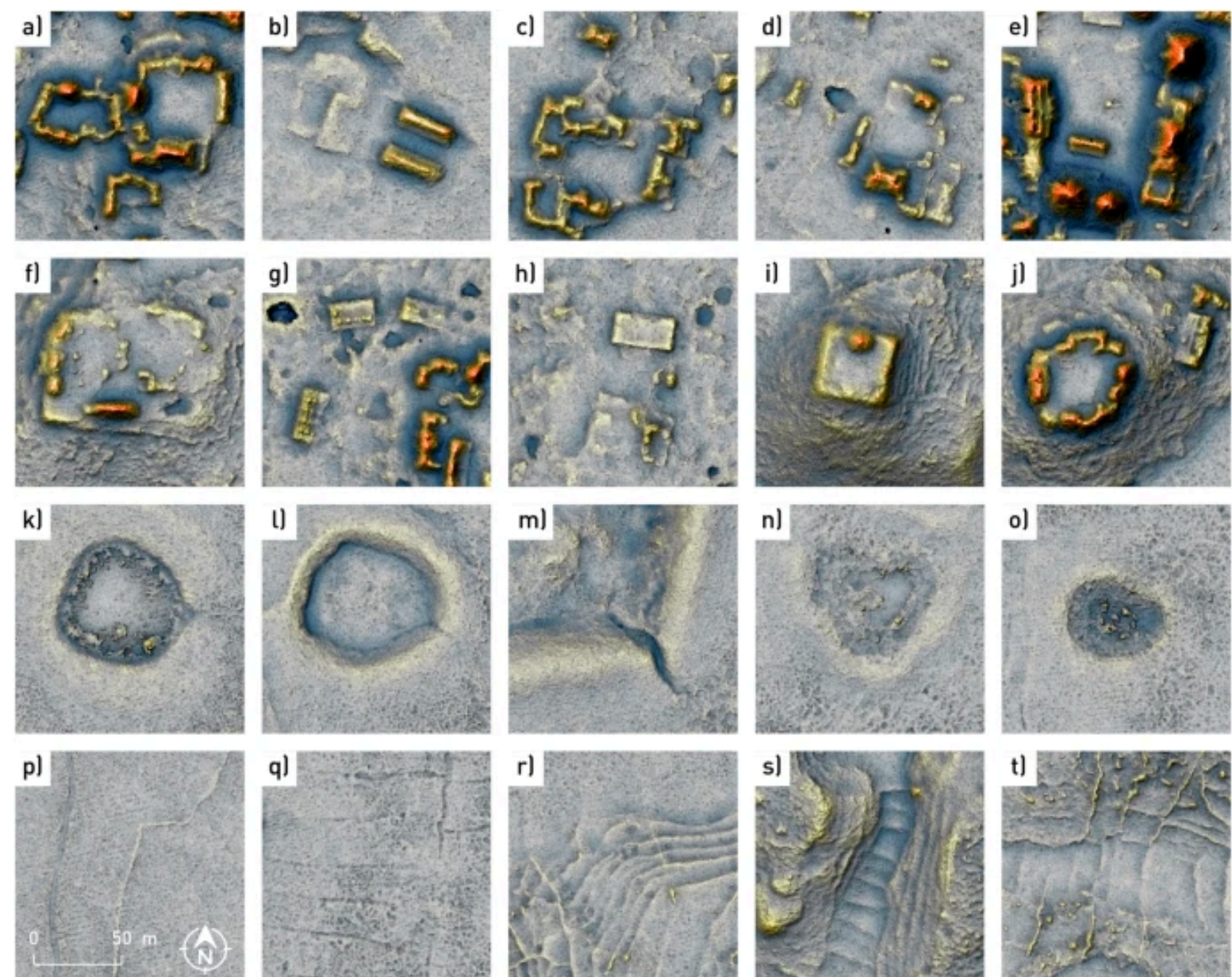
**Develop a segmentation model for Maya archaeology**

2

**Improve the model with data samples of known Amazon archaeological sites**

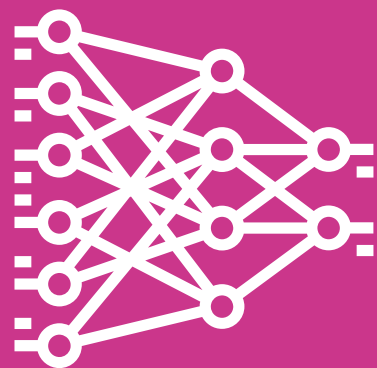
3

**Search for new discoveries in Amazon regions with high-res LiDAR coverage**



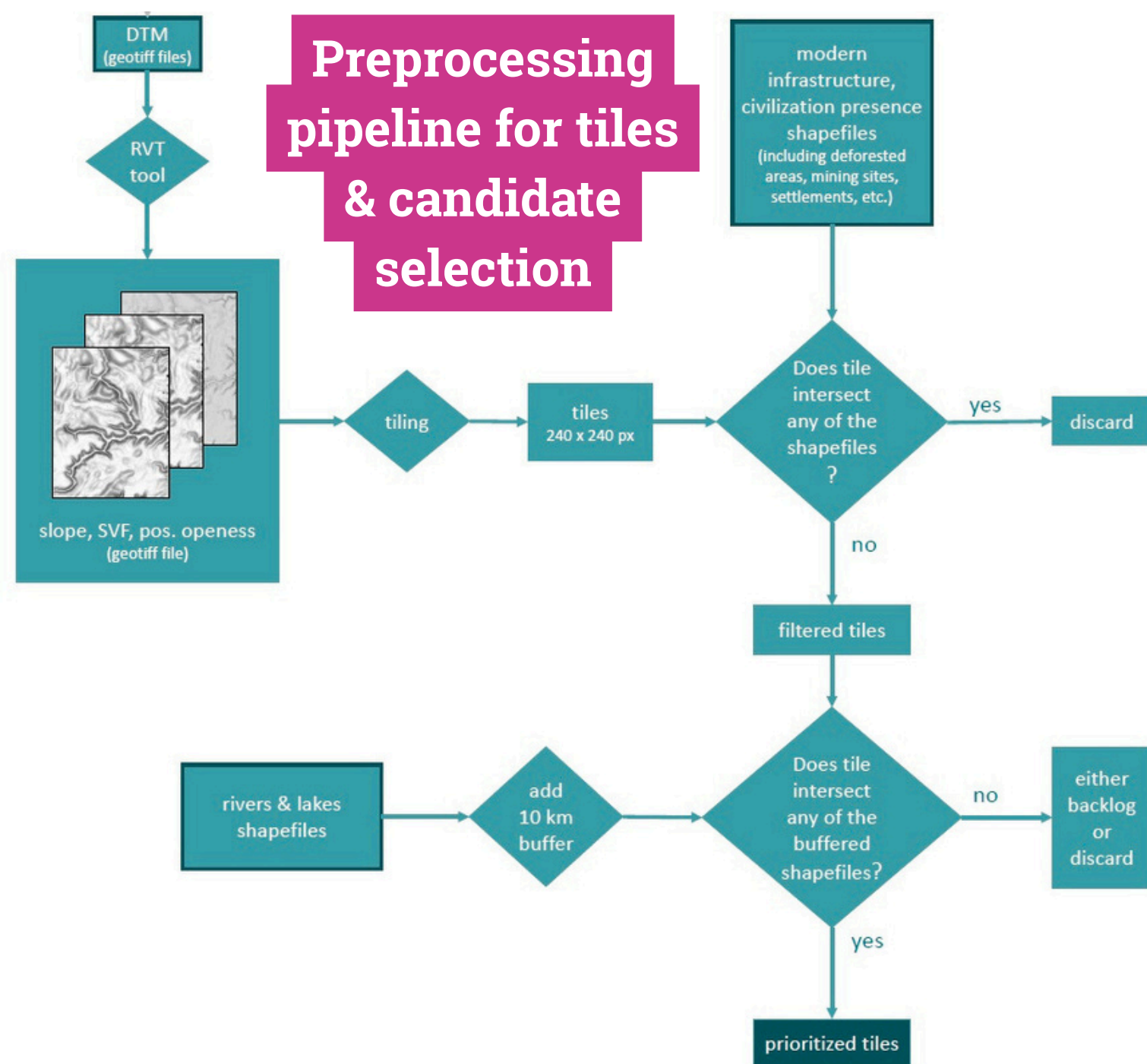
# Mapping Amazon earthworks to Maya archaeological categories

FEATURE	B	P	A	T	RATIONALE
AMAZON DARK EARTHS (ADE)	-	-	-	✓	ADEs are anthropogenic soils that create terrain modifications but don't form discrete architectural features detectable through LiDAR segmentation.
GEOGLYPHS	-	✓	-	-	Large geometric earthworks with embankments that function as raised platforms for ceremonial purposes
MOUNDS	-	✓	-	-	Artificial elevated surfaces that serve as platforms for structures and activities, similar to Maya platforms
CIRCULAR VILLAGES	✓	-	-	-	Settlement layouts with individual house structures arranged in circular / rectangular patterns around central plazas
RAISED PLATFORMS	-	✓	-	-	Elevated earthwork surfaces for elite residences, rituals, and flood protection
CAUSEWAYS	-	✓	-	-	Elevated roadways that function as linear platforms crossing wetlands
TERRACE FIELDS	-	?	-	?	Agricultural terraces creating multi-level elevated planting surfaces *
WATER RESERVOIRS	-	-	✓	-	Engineered water collection basins serving the same function as Maya aguadas
FISH WEIRS	-	?	?	-	Stone/wooden barriers in waterways might register as water management features, but detection via LiDAR is uncertain
SHELL MIDDENS	-	?	-	-	Accumulated refuse piles creating subtle elevation changes that could register as low platforms, though irregular in shape



# BUILDING a ML model

Teaching a model to recognize ancient landscapes



## Segmentation model ensemble

DeepLabv3+ / U-Net / HRNet

**Combined multiclass + simplified binary classifier**  
**85% IoU** (5-fold cross-validation)

### Test-Time Augmentation

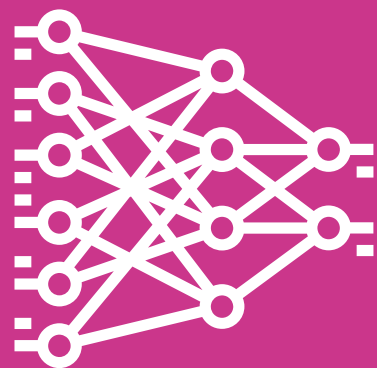
4 rotations, averaged – simple trick, big impact

### Post-processing

Probability thresholding + spatial clustering + size filtering

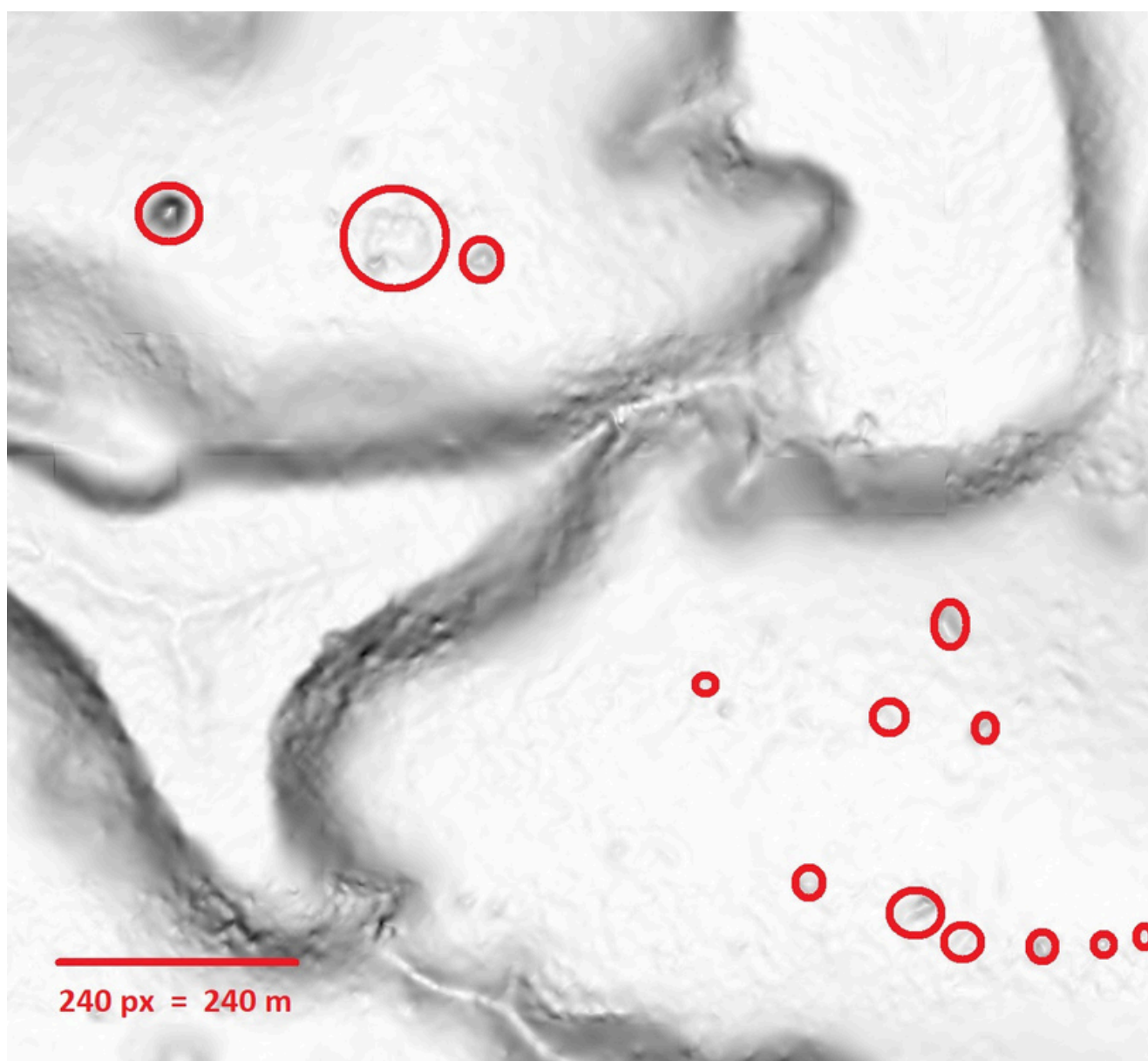
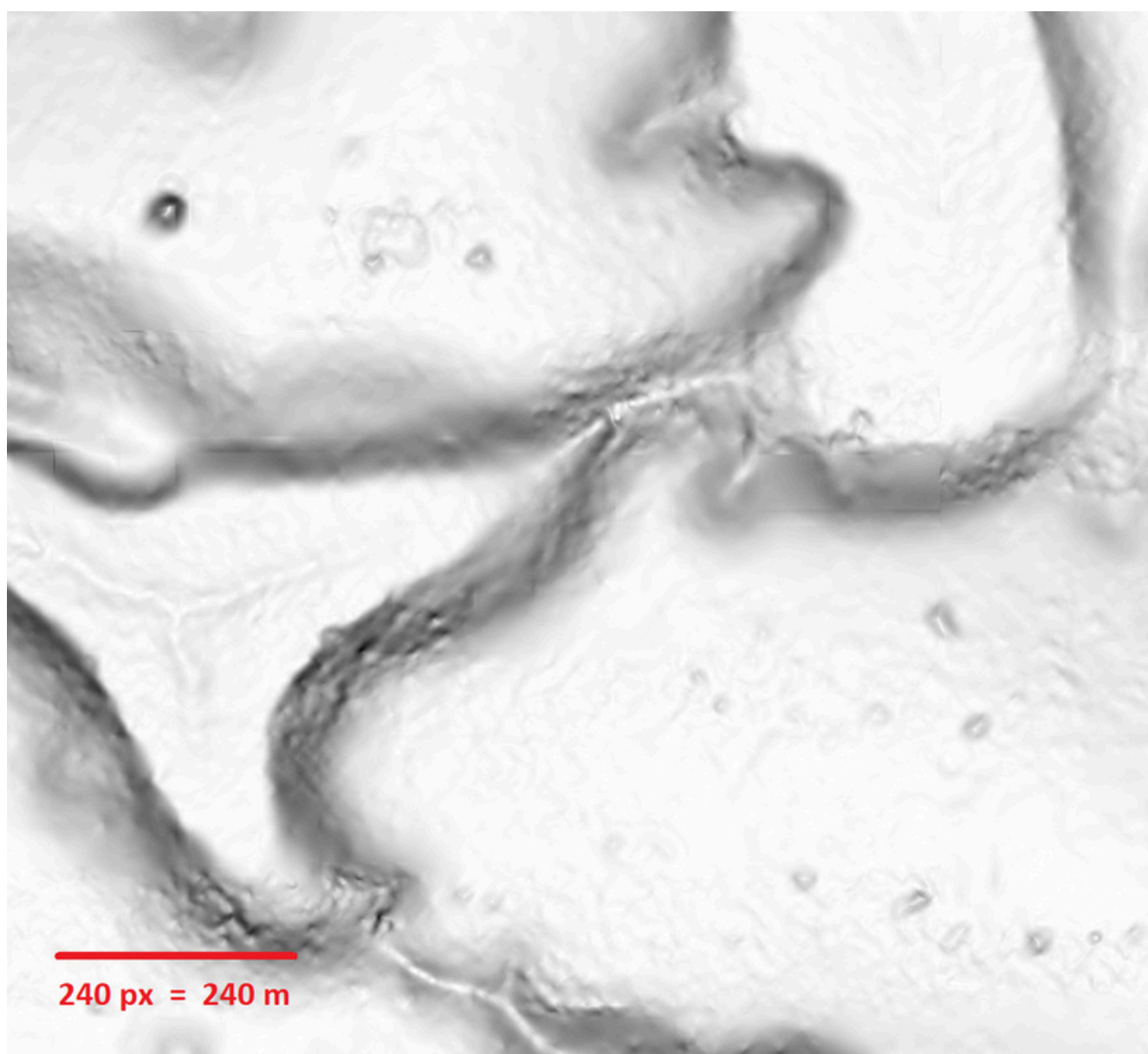
### Cross-domain transfer

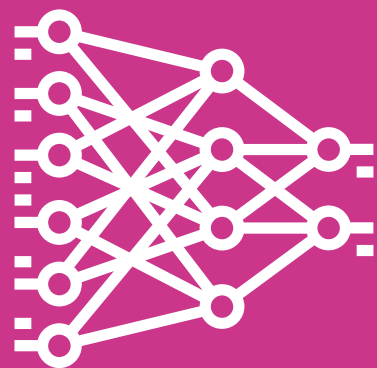
Maya → Amazon



# Fine-tuning with limited Amazon data

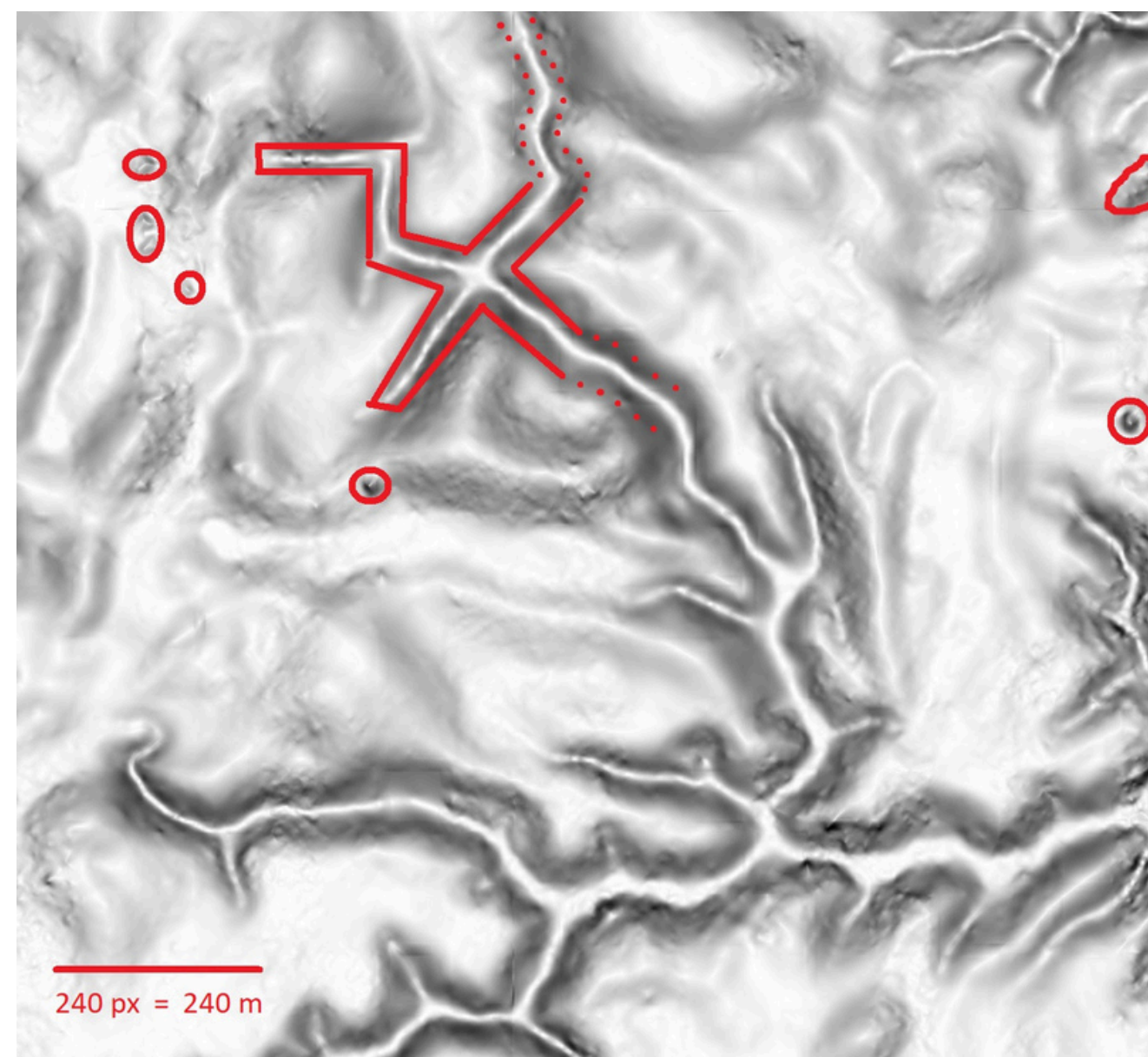
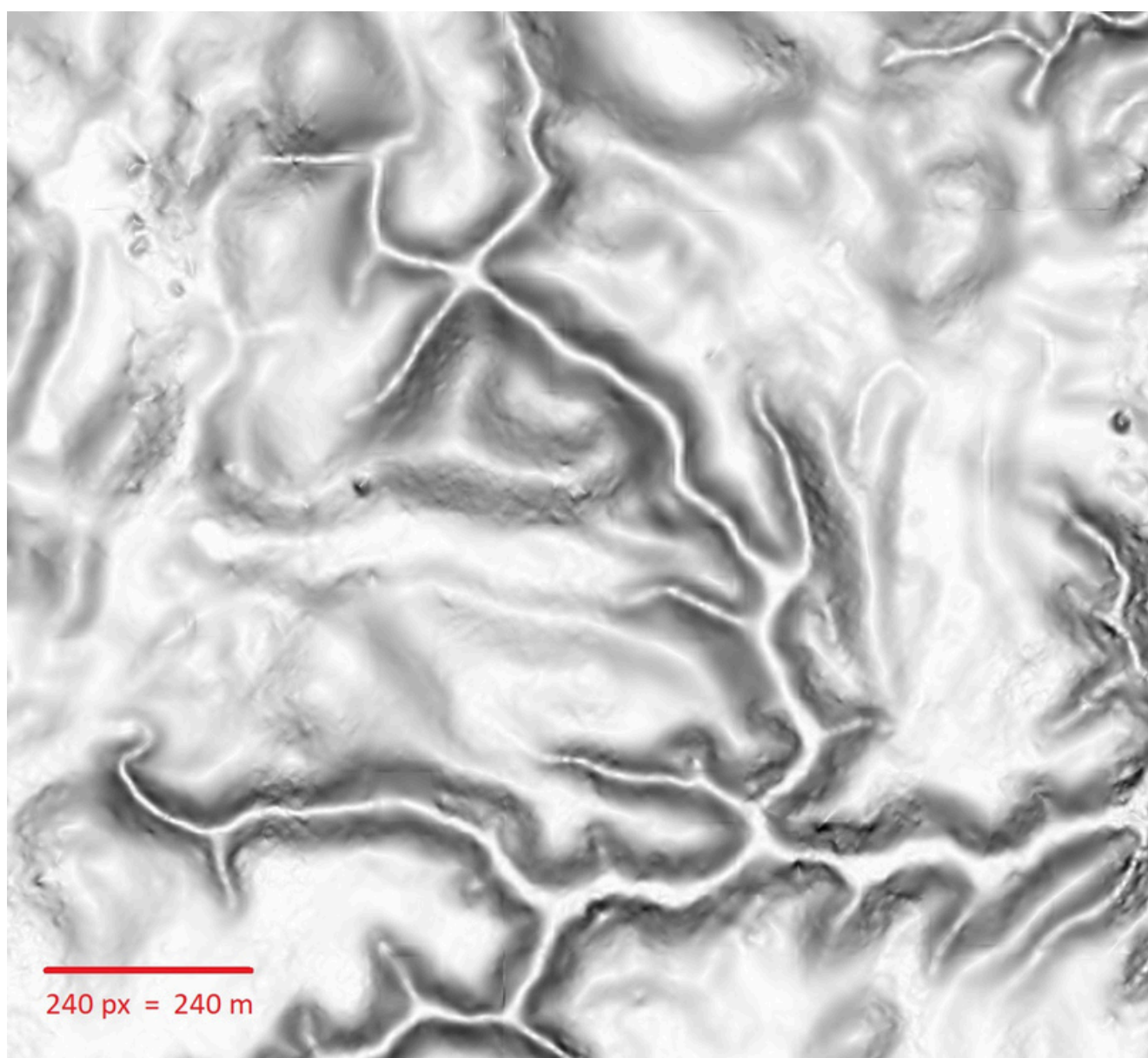
Issues of ground truth and interpretation

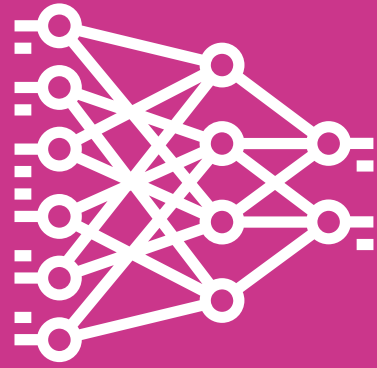




# Fine-tuning with limited Amazon data

Issues of ground truth and interpretation





## Zero-Shot Learning with OpenAI LLM

Instructed to look for key anthropogenic indicators



- **Bump-like features and elevated anomalies**

Mound-like formations and artificial raises that deviate from natural terrain patterns, ditches, embankments.

- **Geometric shapes and regularity**

Complete or fragmented geometric forms, straight lines, right angles, and bilateral symmetry

- **Parallel and systematic structures**

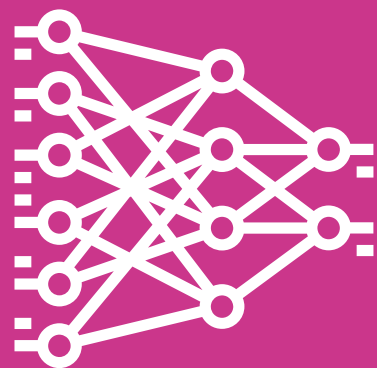
Terraces, walls, aligned features, and repetitive spacing patterns

- **Settlement patterns**

Circular and rectangular formations of individual features

- **Edge contrasts and boundaries**

Sharp transitions in elevation, slope, or terrain texture indicating intentional human landscape modification



# Zero-Shot Learning with OpenAI LLM

Instructed to look for key anthropogenic indicators



## Observations:

- The central area features a possible rectangular depression or enclosure (center-right) bordered by more pronounced edges—this is highly suggestive of an anthropogenic earthwork.
- There are multiple repetitive small mounds across the flatter zones with fairly regular spacing—potential burial mounds or post-platform remains?
- In the lower left quadrant, a pair of parallel ridgelines appears unnaturally straight and evenly spaced—this could be indicative of old terrace walls or ancient road embankments.
- Edges of some plateaus appear too straight and angular to be natural—especially in the upper-right quadrant.

## Assessment:

This tile shows strong indicators of human modification, especially geometric enclosure-like shapes and repetitive patterns. These are the kinds of features often associated with planned settlements or ceremonial earthworks in Amazonian archaeology. It has high anthropogenic potential.

# Limitations

## **What this solution does**

Shift effort from blind exploration to informed prioritization.  
Ground truth validation is still needed to 100% confirm the findings.

## **Data quality constraints**

LiDAR coverage is limited and results depend on data resolution, scan density and terrain and vegetation complexity.

## **Lack of universal tracking of sites discovery**

Without definitive record of all known archaeological sites, no team can fully verify if their discoveries are really new.

# Why this matters?

## **Cultural heritage preservation**

Estimated 10,000+ earthworks still hidden under the canopy

## **Environmental**

Fewer blind expeditions = less forest disturbance

## **Cost/time reduction for archaeological surveys**

## **Scalable framework**

Same approach could work for other forested regions

## **Indigenous land rights**



# Project Website



[amazonarchaeologydiscovery.com](https://amazonarchaeologydiscovery.com)