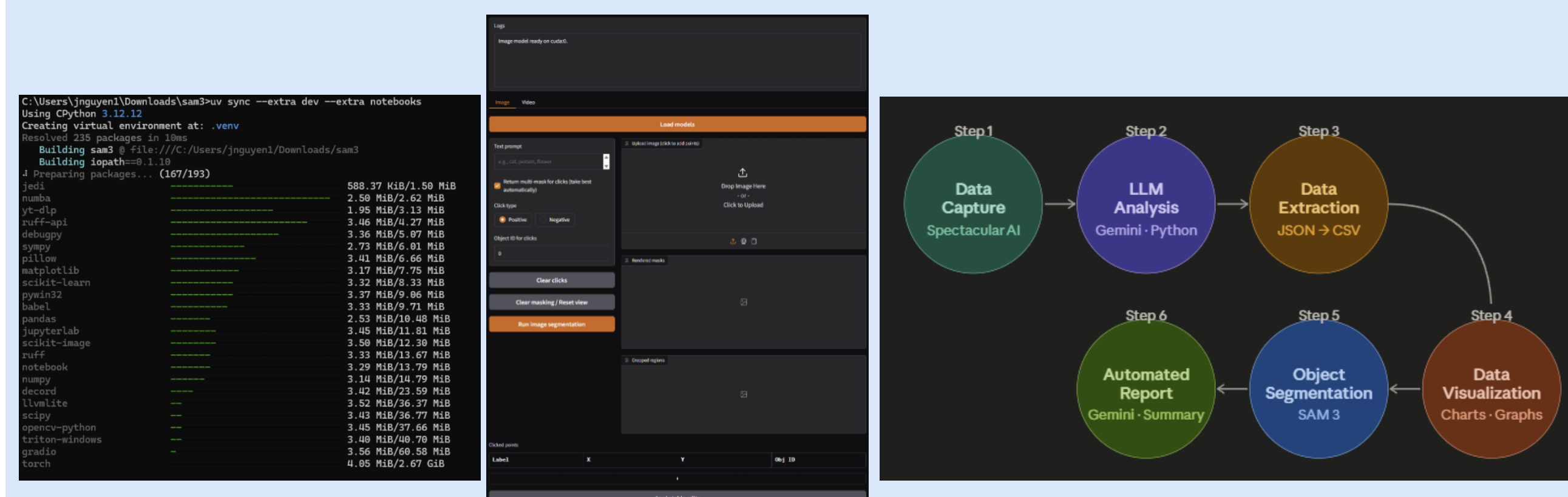


ABSTRACT

This is a deep dive into developing an algorithm for an automated framework to analyze construction site progress using temporal image data and artificial intelligence. Images collected from a residential construction site at multiple time intervals were processed using a large language model to summarize how construction has progressed over time. Specific investigations into using VLMs completed as part of this work include: generating structured JSON outputs, estimating object frequency distributions, revealing temporal trends, and corroborating VLM-generated conclusions with open-vocabulary-based image segmentation. The results demonstrate that integrating AI-based object detection, frequency analysis, and segmentation provides an efficient and scalable method for monitoring construction progress. This approach supports more accurate, data-driven decision-making for construction stakeholders and highlights the potential of AI in automating progress tracking.

METHODOLOGY

- Collected and captured data using Spectacular AI
- Ran Python script, in terminal, to output a results JSON file (object frequency) using LLM Gemini 3
- Extracted results JSON file into a CSV for object frequency count
- Created histograms and graphs using the results JSON file
- Used SAM 3 by inputting key word from the JSON file to highlight the object
- Used Gemini 3 to create an automated report, highlighting progress from start to finish



PHASE ANALYSIS

Phase I: Site Preparation & Foundation – Generated by SAM 3

- Heavy earthmoving with excavators and dump trucks managing soil piles
- Foundation walls using rebar, concrete, and wood forms
- Concrete mixer trucks deployed for foundation pours



Phase II: Structural Framing – Generated by SAM 3

- Skeleton built with lumber, wood framing — most observed material
- Plywood sheets are used extensively for wall sheathing and flooring
- Roof trusses installed, completing the structural height



Phase III: Exterior Finishing & Closing – Generated by SAM 3

- Structure enclosed with siding, windows, and doors
- Exterior finished with brick accents and white siding
- Driveway, curbs, and utility poles finalized



GENERATED SCENARIOS

Infrastructure & Utilities – Generated by Gemini 3

- Utility poles and utility sheds installed in the mid-to-late stages
- Survey stakes used to mark property lines and landscaping boundaries
- Driveway poured and sidewalk connected to street infrastructure

Exterior Architectural Finishes – Generated by Gemini 3

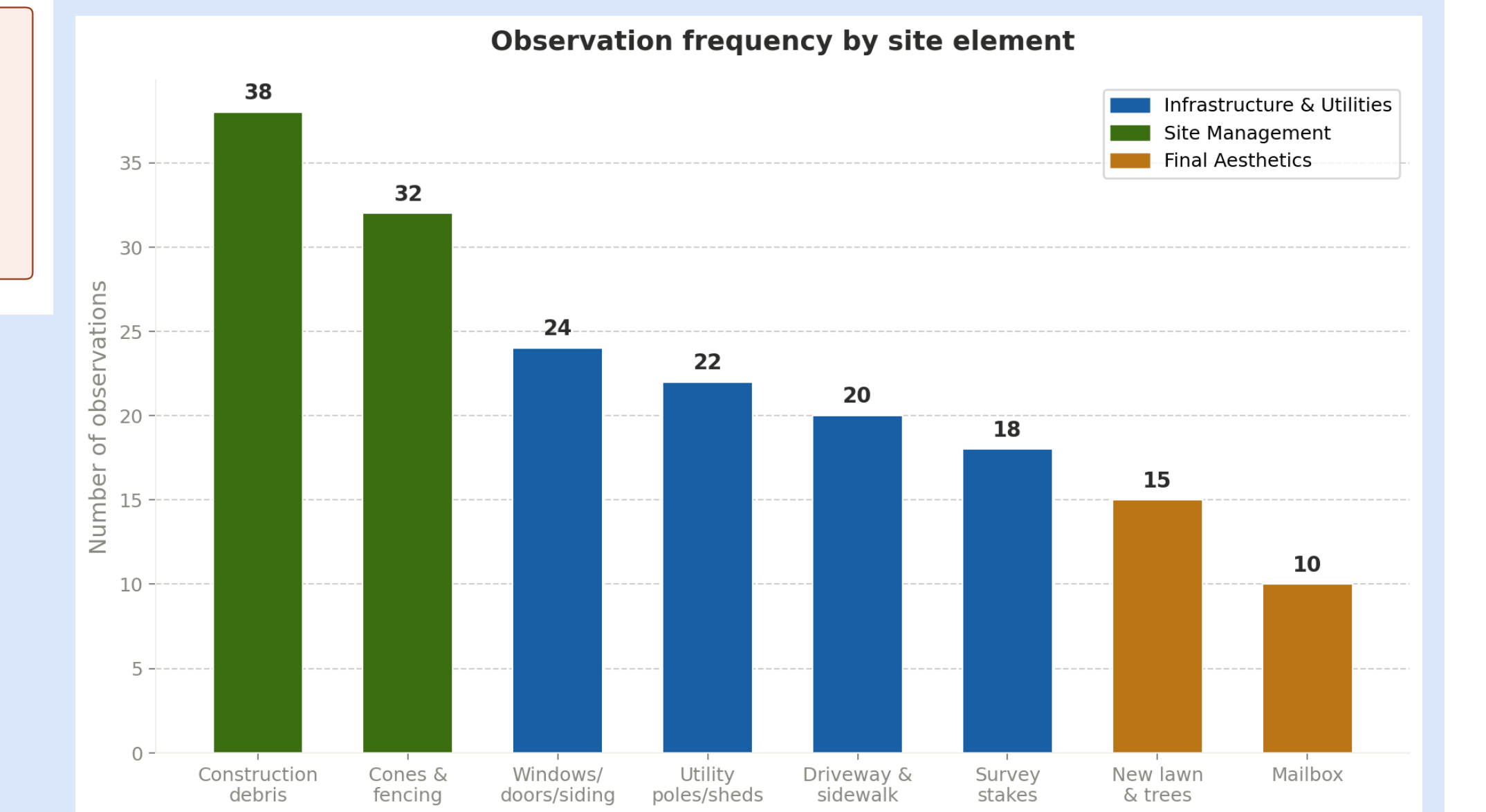
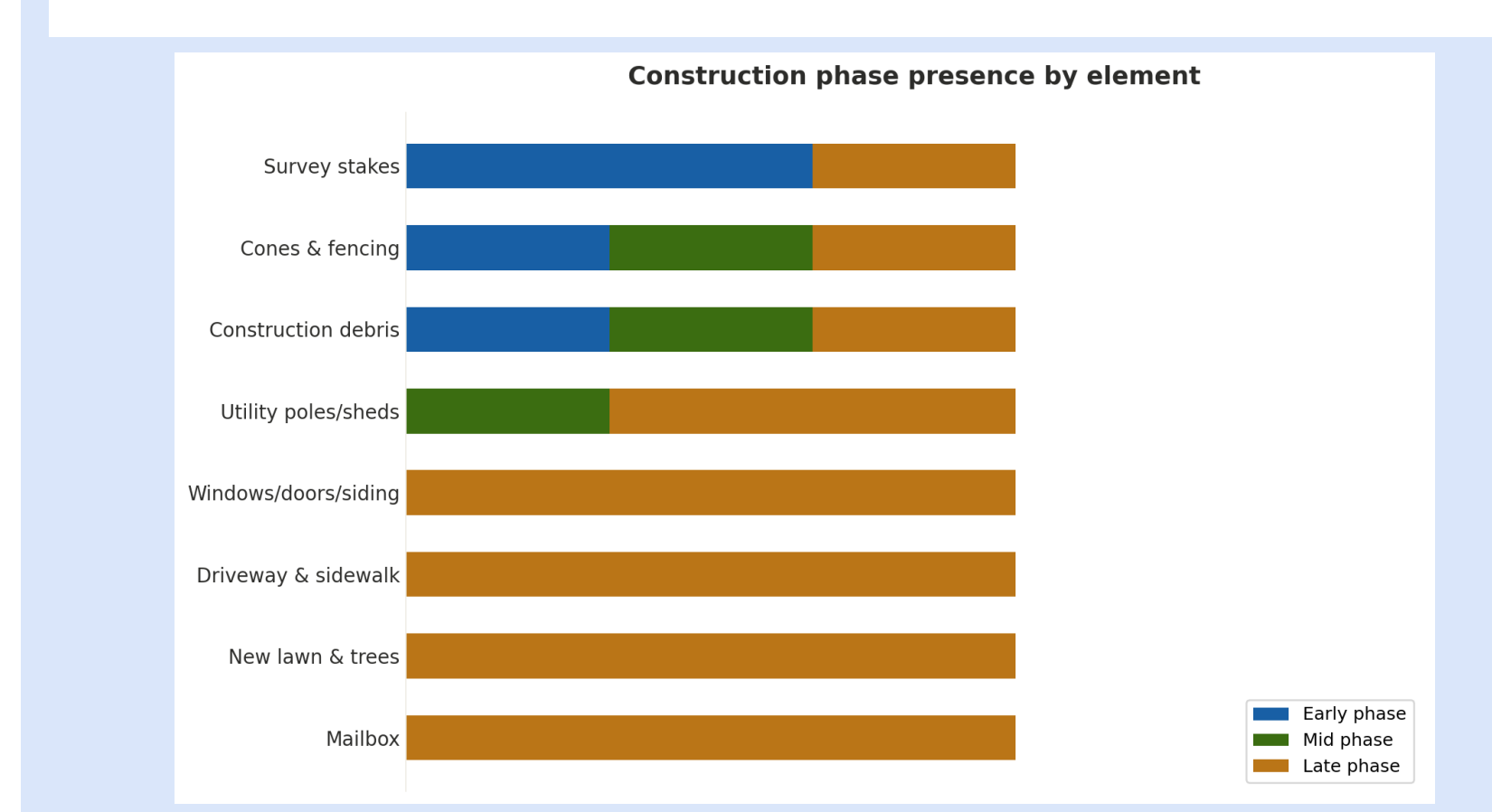
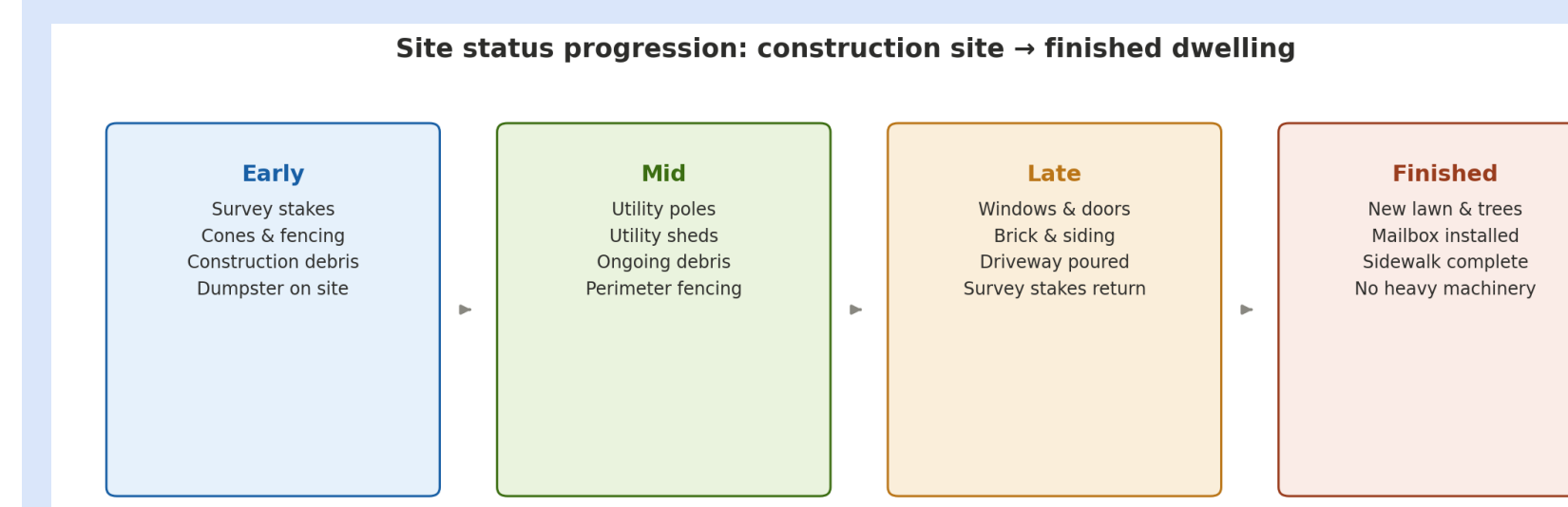
- Windows, doors, and siding complete the building envelope
- Brick accents and white siding give the final exterior character
- Driveway and sidewalk access infrastructure completed late-stage

Operational Observations – Generated by Gemini 3

- Construction debris is the second-most observed (38 times), requiring a constant dumpster presence
- Cones and fencing managed the site perimeter throughout all phases
- Wooden and survey stakes were used early and late for boundaries

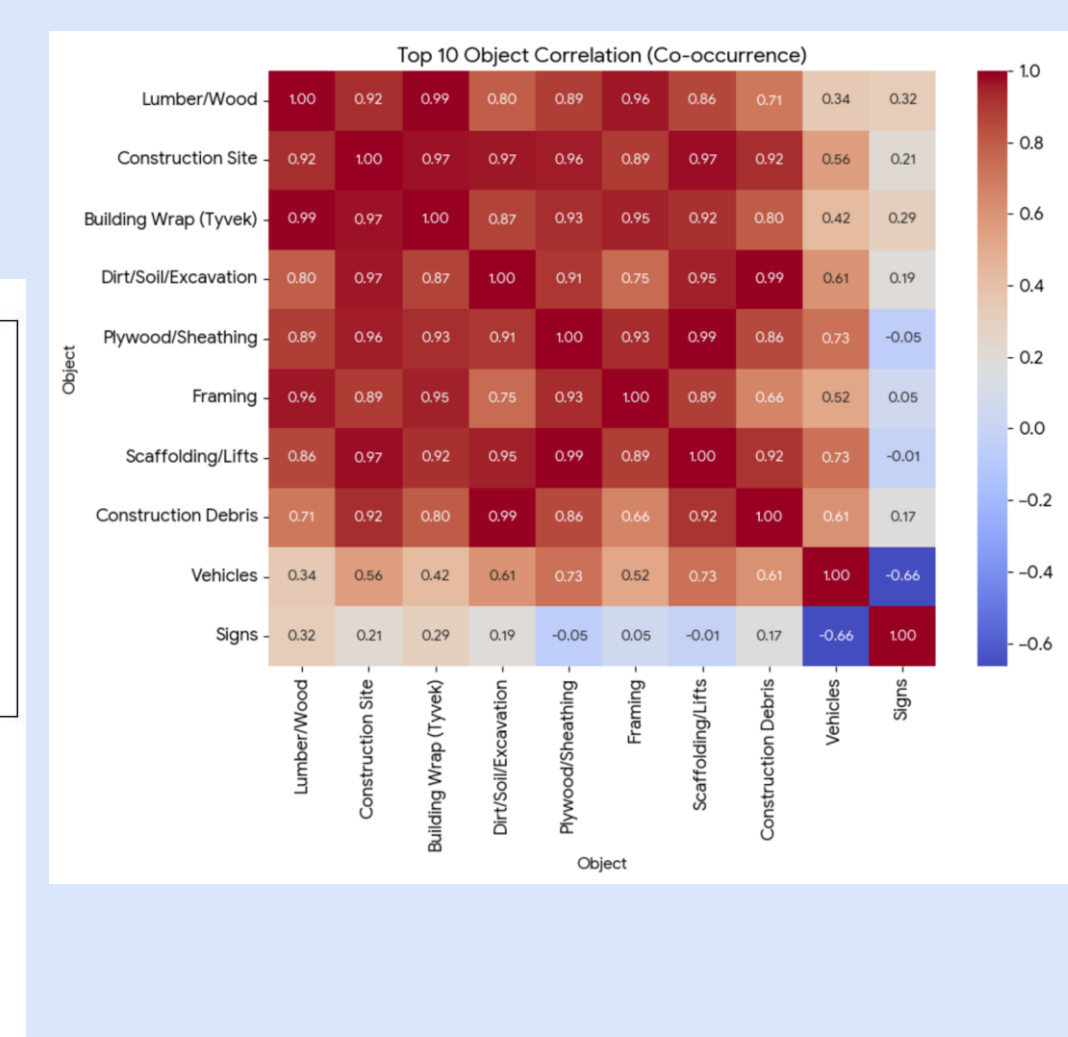
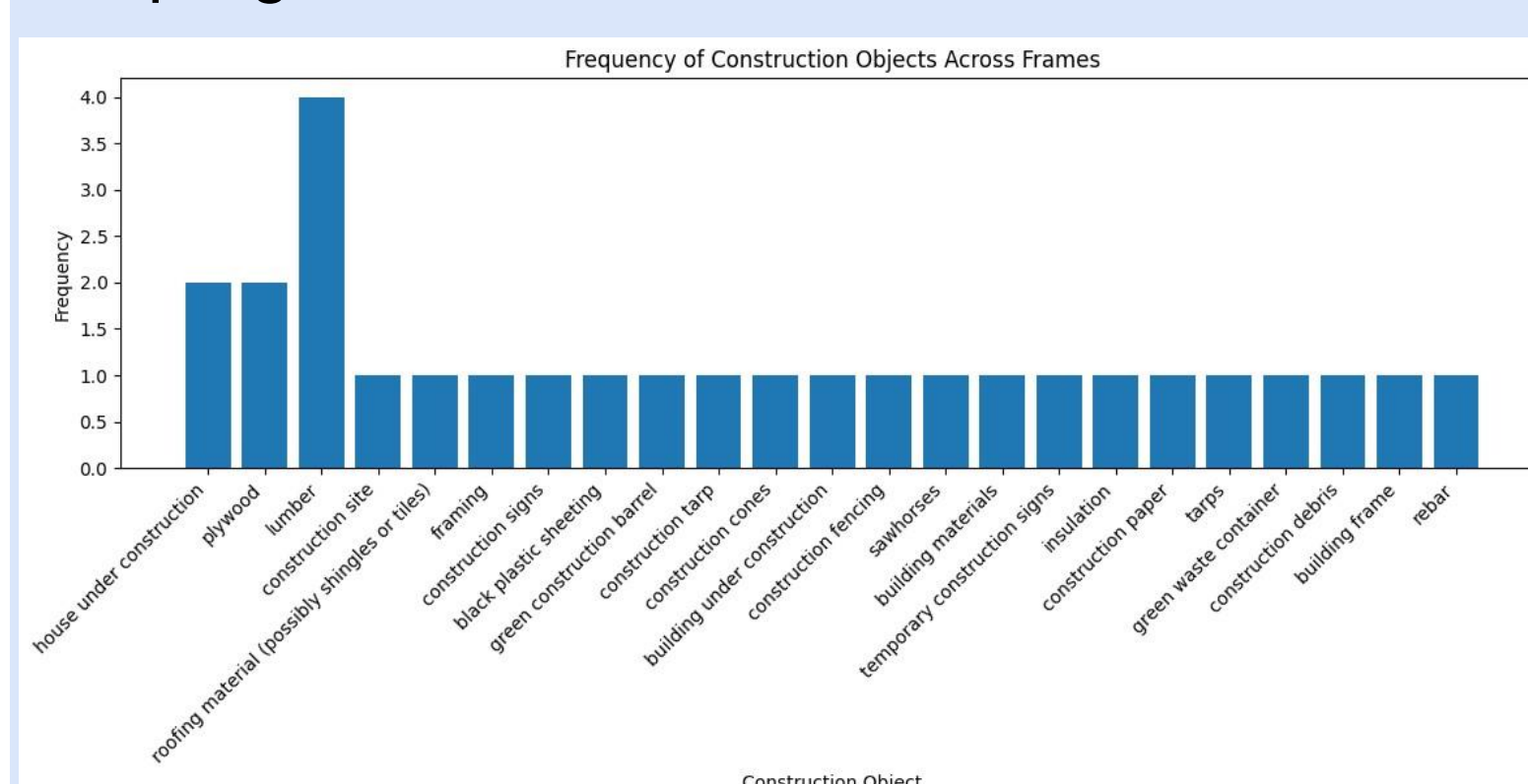
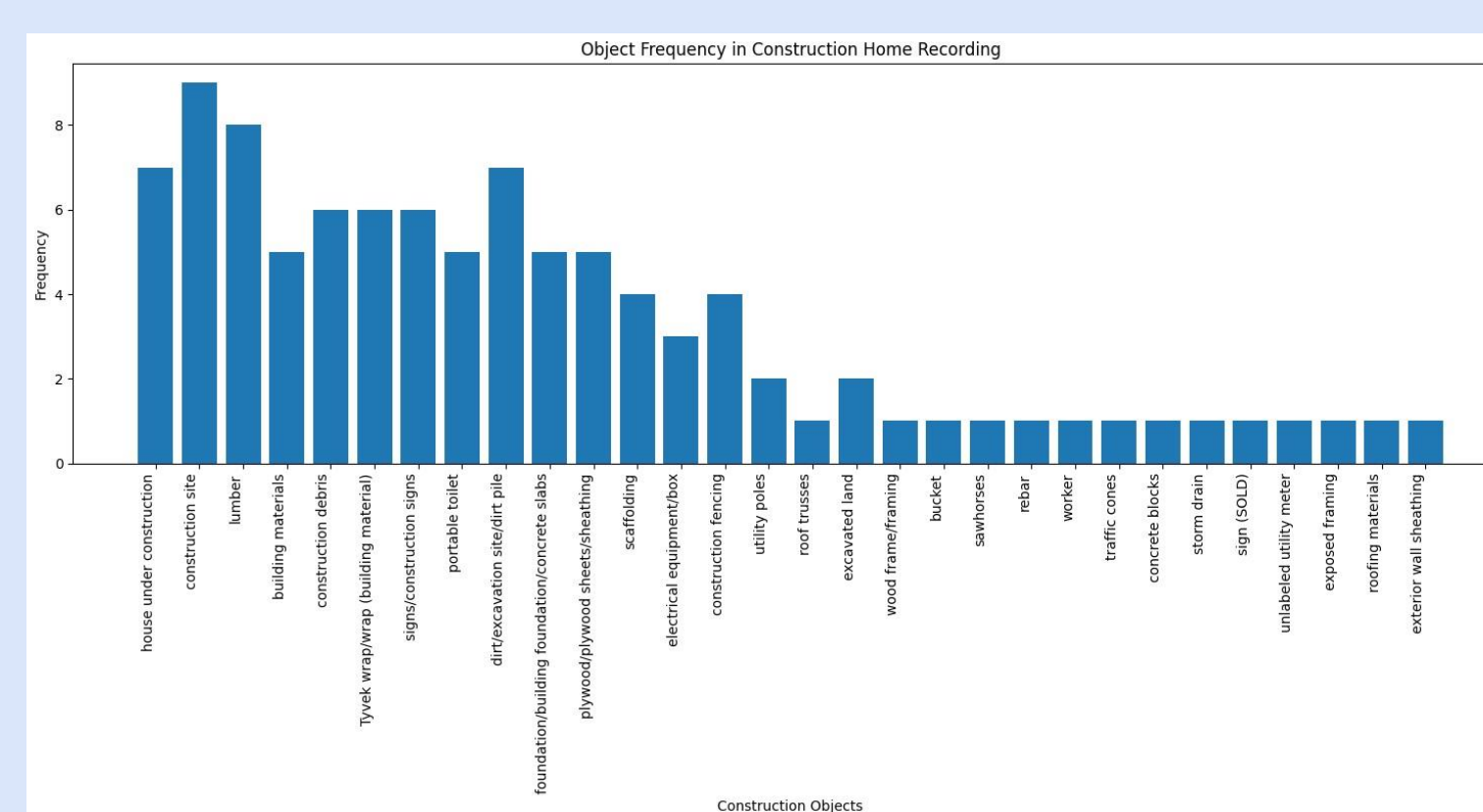
Final Aesthetics & Signage – Generated by Gemini 3

- Structure status evolved from "Construction Site" to "Finished Dwelling"
- New lawn, grass, and freshly planted trees signal the end of the heavy machinery phase
- Mailbox (10 observations) signifying near-occupancy readiness



OBJECT FREQUENCY

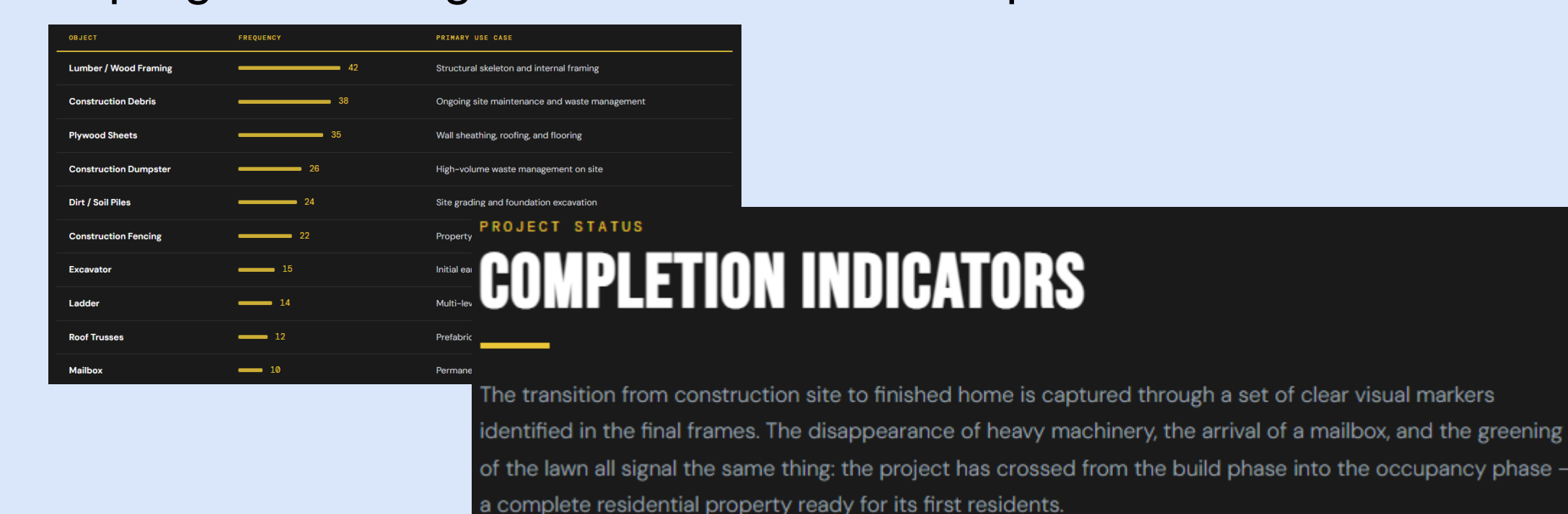
- Histograms were created from JSON outputs generated per frame using Gemini 3, showing object frequency from Phase I and Phase II
- Object correlation of the top vocabulary word that has appeared in each frame
- Objects that were most critical to the project's progression



REPORT GENERATION

When creating the automated report for the progression of the construction of the home, it created a conclusion based on:

- Gemini 3 identifies what objects were present across time, how frequently they appeared, and how their counts shifted (lumber frequency drops as framing completes; roofing materials appear in later images).
- SAM 3 visually confirmed where those objects were in each image, providing a spatial and visual layer of evidence that the VLM's language-based conclusions were grounded in actual image content.
- Gemini 3 then synthesized both the frequency trends and the segmentation-validated findings into a cohesive automated report — essentially a data-driven narrative of construction progress from groundwork to near-completion.



DISCUSSION

- The use of an LLM (Gemini 3) to automatically classify objects from video frames demonstrated the feasibility of AI-driven site monitoring without requiring manual annotation, suggesting strong potential for scalable, low-cost construction documentation.
- Frame-by-frame object detection successfully captured the sequential nature of residential construction, with distinct object categories (heavy machinery, structural materials, site logistics) aligning closely with recognized industry phases.
- The sequential appearance of mailbox installation, freshly laid sod, and newly planted trees in the final frames served as reliable visual markers of near-occupancy readiness, suggesting these objects could function as automated "project completion" indicators in future AI monitoring systems.
- Future research could incorporate temporal weighting of observations to distinguish between objects that appeared once versus those maintained continuously across many frames, improving the precision of phase-duration estimates.

REFERENCES & ACKNOWLEDGEMENTS

Special thanks to the Lawrence Technological University Department of Math and Computer Science for their hardware and software support.



Scan the QR Code to See the System Output: