

From Images to Energy Models: Generative AI for Urban Energy Data

Jackson Eshbaugh ¹, Chetan Tiwari ^{2 3}, Jorge Silveyra ¹

¹ Department of Computer Science, Lafayette College

² Departments of Geosciences & Computer Science, Georgia State University

³ Center for Disaster Informatics & Computational Epidemiology, Georgia State University

Modeling and Simulation

- What is modeling?
- Why do we model?
- Energy models

ENERGY MODELERS



How to simulate the energy “consumption” of one home?

- EnergyPlus is an energy analysis and thermal load simulation program.
- Based on a user’s description of a building’s geometry, construction materials, usage, and systems
 - a. EnergyPlus will calculate the heating and cooling loads necessary to maintain thermal control setpoints
 - b. conditions throughout secondary HVAC system and coil loads, and
 - c. The energy consumption of primary plant equipment as well as many other simulation details that are necessary to verify that the simulation is performing as the actual building would.



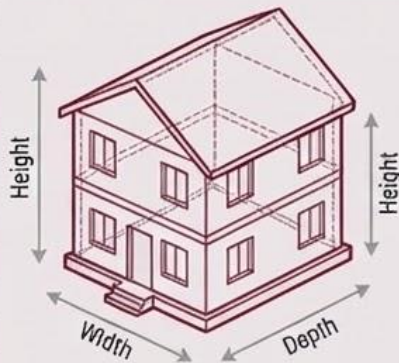
Data Inputs for Home Energy Simulation

Weather Data



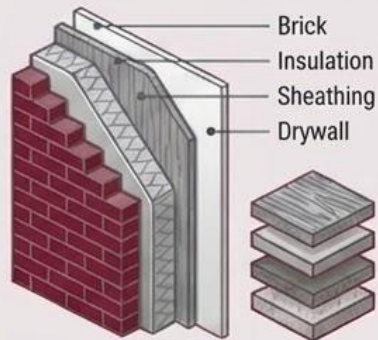
Historical weather files, solar radiation, temperature.

Building Geometry



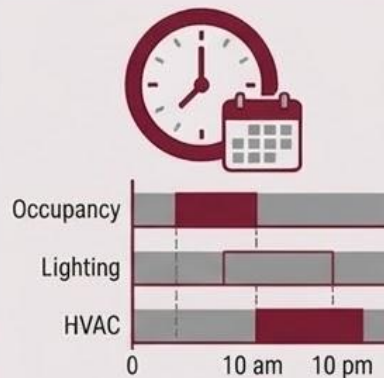
Floor plan, elevation, roof shape, window locations.

Construction/ Material Properties



Thermal resistance (R-values), heat capacity, material types.

Operational Schedules



Daily occupancy patterns, thermostat settings, settings, appliance usage.

Where do we get the data from?

Weather Data (EPW)



Weather Data (EPW)

Historical weather files, solar radiation, temperature.

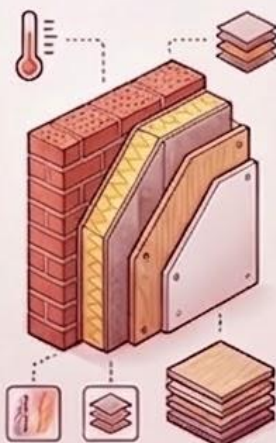
Building Geometry (OpenStreetMap/GIS)



Building Geometry (OpenStreetMap/GIS)

Floor plan, elevation, roof shape, window locations.

Construction/Materials (ASHRAE/DOE)



Construction/Materials (ASHRAE/DOE)

Thermal resistance (R-values), heat capacity, material types.

Operational Schedules (EIA/ASHRAE)



Operational Schedules (EIA/ASHRAE)

Daily occupancy patterns, thermostat settings, appliance usage.

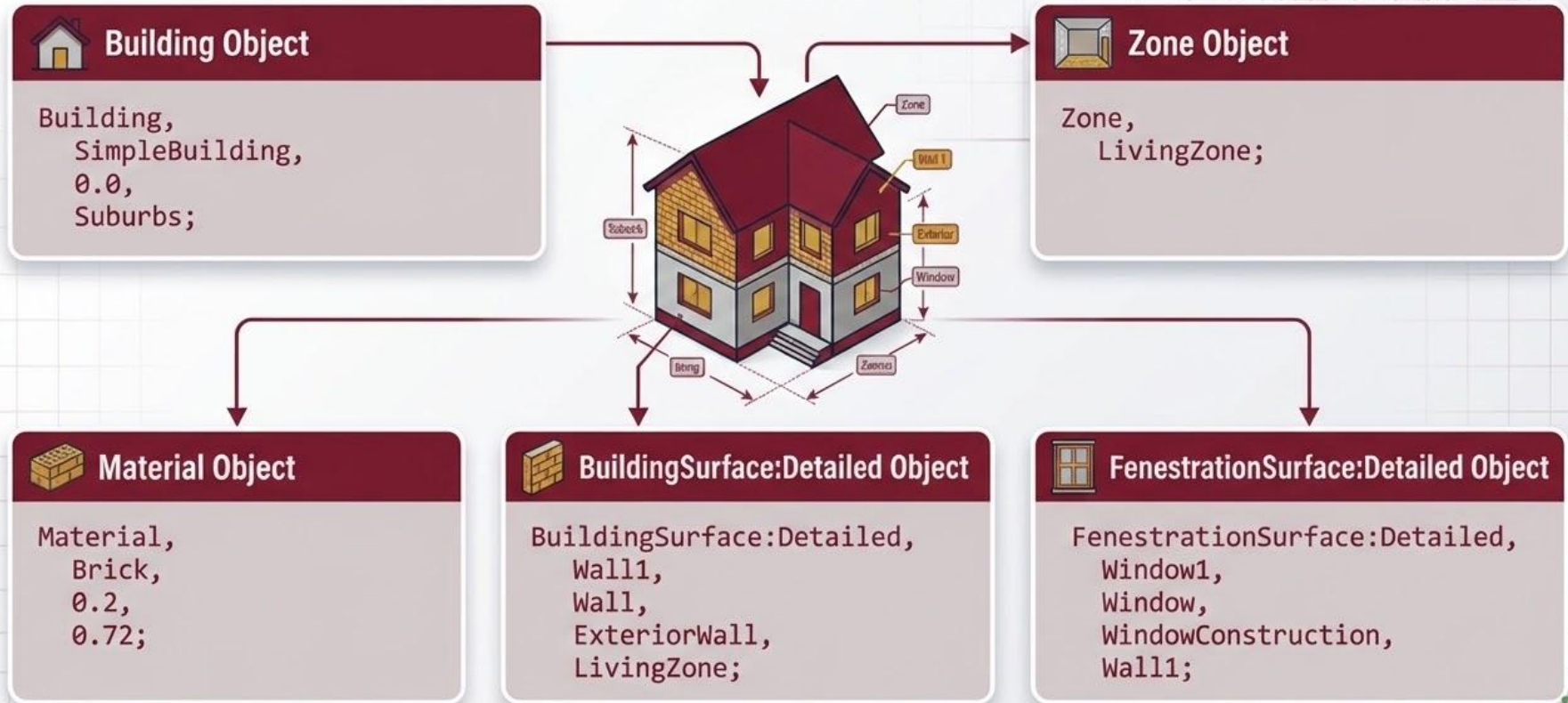
HVAC Assumptions



HVAC Assumptions

System types, efficiency, performance curves.

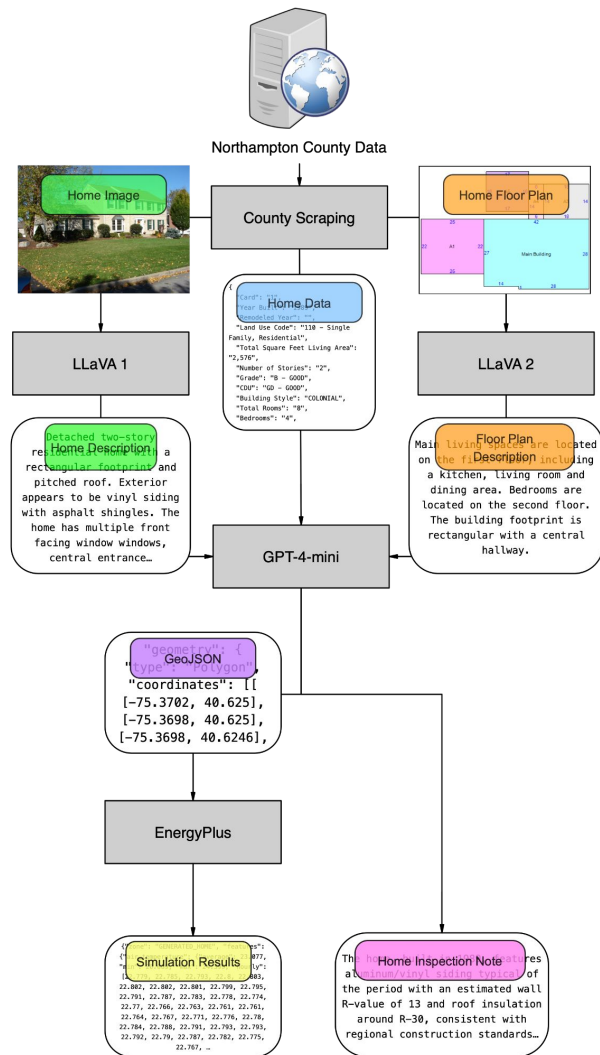
What is the output? IDF files





Solution

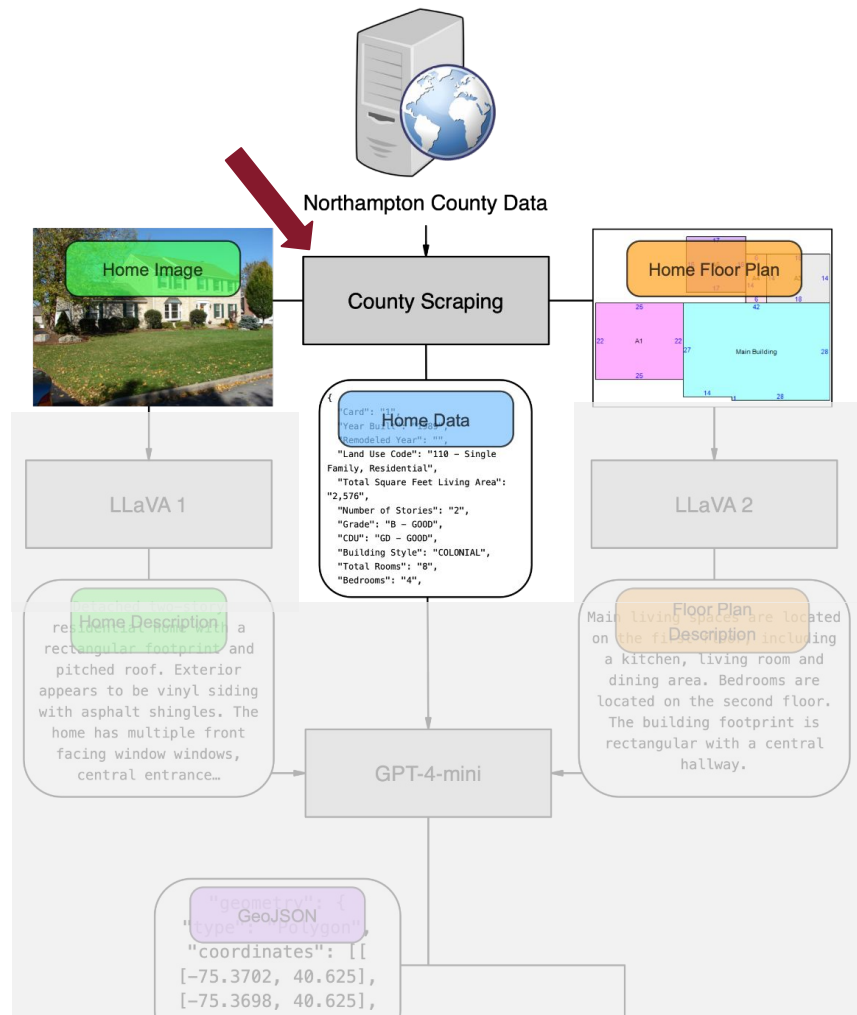
A multimodal pipeline powered by generative AI which can generate synthetic versions of this sought after data at a dramatically lower price.



Stage 1: Web Scraping

Input: Northampton County data

Outputs: home data, floor plan, and image




Web Scraping

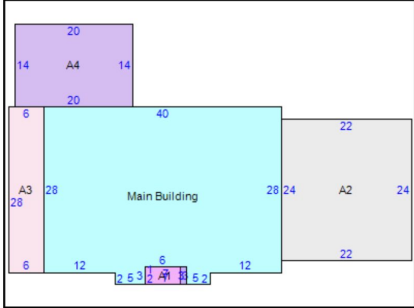
- With the aim of making our generated data **realistic** existing data as a starting point

Results			Residential	
All	Parcel ID▲	Owner	Card	1
<input type="checkbox"/>	M7NW3 1 2 0205	DONCHEZ REBECCA A	Year Built	1988
<input type="checkbox"/>	M7NW3 12 7 0205	GRASING DAVID &	Remodeled Year	
<input type="checkbox"/>	M7NW3 13 1 0205	MANTZ STEPHEN P & RENEE	Land Use Code	110 - Single Family, Residential
<input type="checkbox"/>	M7NW3 13 2 0205	STRONG THOMAS D & LINDA	Total Square Feet Living Area	2,457
<input type="checkbox"/>	M7NW3 13 3 0205	SAHL ANDREW J &	Number of Stories	2
<input type="checkbox"/>	M7NW3 13 4 0205	VERMA MAHESHWER B & PUN	Grade	B+ - GOOD +
<input type="checkbox"/>	M7NW3 13 5 0205	PARVEZ RAJA M & SHAHIDA M	CDU	GD - GOOD
<input type="checkbox"/>	M7NW3 13 6 0205	SKEE DAVID C & JANET L	Building Style	COLONIAL
<input type="checkbox"/>	M7NW3 13 7 0205	IRR HANS J &	Total Rooms	8
<input type="checkbox"/>	M7NW3 13 8 0205	RAY GREGORY T & JEANETTE	Bedrooms	4
<input type="checkbox"/>	M7NW3 13 9 0205	LEE DAVID S Y & MARYSE S	Full Baths	2
<input type="checkbox"/>	M7NW3 14 4 0205	ASHCRAFT MICHAEL FRANCIS	Half Baths	1
<input type="checkbox"/>	M7NW3 14 5 0205	YOUNG PETER W & CHERIE C	Additional Fixtures	0
<input type="checkbox"/>	M7NW3 14 6 0205	DRAPER ANDREW J & CONNIE	Total Fixtures	10
<input type="checkbox"/>	M7NW3 14 7 0205	EDWARDS THOMAS E & MARC	Heat/Air Cond	AIR COND
			Heating Fuel Type	ELECTRIC
			Heating System Type	WARM AIR
			Attic Code	1 - NONE
			Unfinished Area	
			Rec Room Area	560
			Finished Basement Area	0
			Fireplace Openings	1
			Fireplace Stacks	1
			Prefab Fireplaces	0
			Basement Garage (Number of Cars)	
			Condo Level	
			Condo/Townhouse Type	-
			Basement	FULL
			Exterior Wall Material	ALUMINUM/VINYL SIDING
			Physical Condition	AV

Printable Version

Results Page: [1] 2 3 Next >>





Options ▾

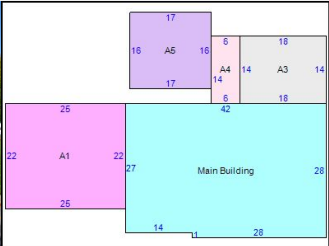
Type	Line #	Item
Dwelling	0	Main Building
Dwelling	1	A1 - 11/10:OPF OPEN FRAME PORCH/1S FR ONE STORY FRAME
Dwelling	2	A2 - 13:FR GR FRAME GARAGE
Dwelling	3	A3 - 50/10:UNFIN BSMT/1S FR ONE STORY FRAME
Dwelling	4	A4 - 31:WDDCK WOOD DECKS

Click on an item to display it independently.

Web Scraping

- With the aim of making our generated data **realistic**, we opted to use existing data as a starting point.
 - Northampton County, PA has a surprising amount of data openly available.
- We scraped this data using **SeLenuim**, a Python package that can find the data on the website by running a program-controlled instance of a browser

```
{
  "Card": 1,
  "Year Built": 1989,
  "Total Square Feet Living Area": "2,576"
  ...
}
```

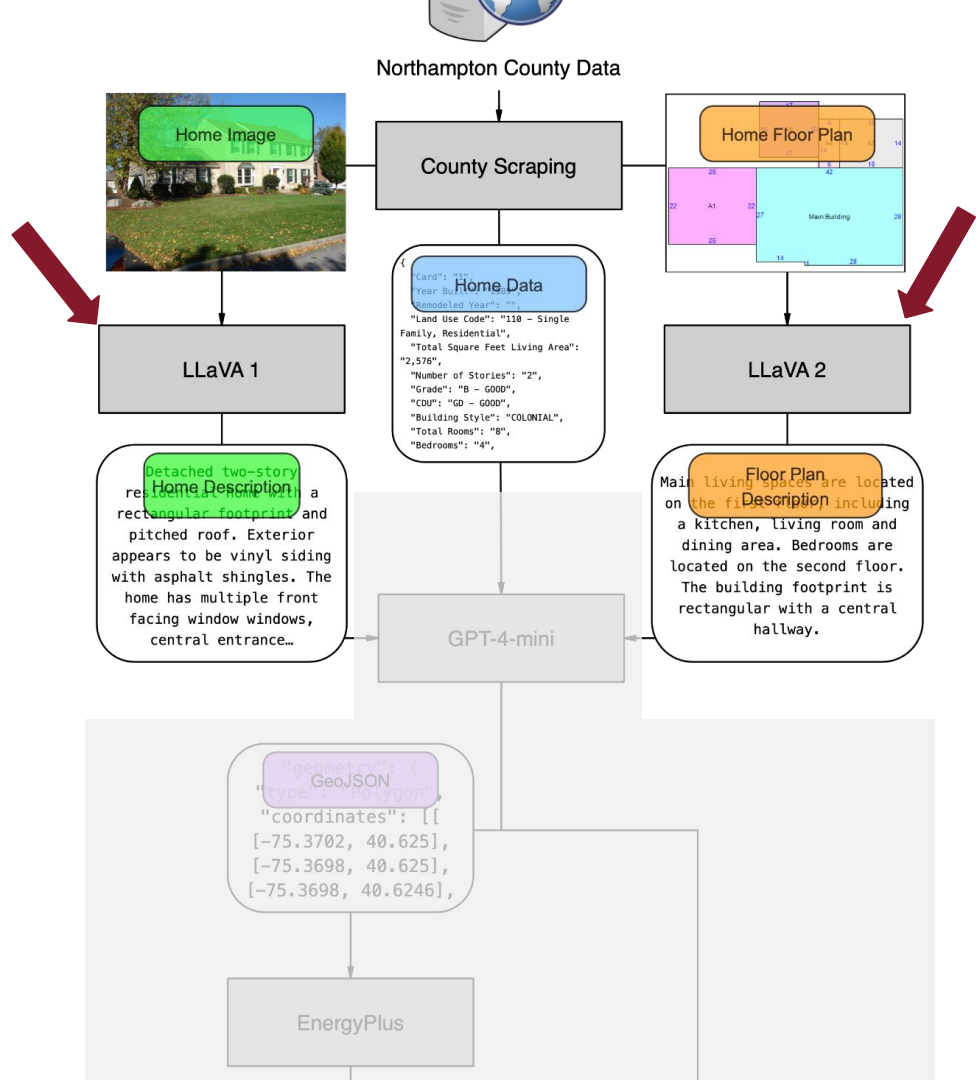


```
<tbody><tr><td width="40%" valign="top" align="left"
class="DataletSideHeading">Card</td><td width="*" valign="top" align="left"
class="DataletData">1</td></tr><tr><td width="40%" valign="top" align="left"
class="DataletSideHeading">&nbsp;</td><td width="*" valign="top" align="left"
class="DataletData">&nbsp;</td></tr><tr><td width="40%" valign="top"
align="left" class="DataletSideHeading">Year Built</td><td width="*"
valign="top" align="left" class="DataletData">1989</td></tr><tr><td width="40%"
valign="top" align="left" class="DataletSideHeading">Remodeled Year</td><td
width="*" valign="top" align="left" class="DataletData">&nbsp;</td></tr><tr><td
width="40%" valign="top" align="left" class="DataletSideHeading">Land Use Code</
td><td width="*" valign="top" align="left" class="DataletData">110 - Single
Family, Residential</td></tr><tr><td width="40%" valign="top">
```

Stage 2: Image Processing

Inputs: home images & floor plans

Outputs: textual descriptions of each input



```
def describe_exterior(image_path: str) → str:
    prompt = (
        "Describe the visible exterior features of this residential building in detail. "
        "Mention materials (brick, siding, etc.), number of visible stories, roof type, windows, "
        "doors, and any other notable architectural features. "
        "Do not speculate beyond what is clearly visible."
    )
    return run_llava(image_path, prompt)
```

- Using LLaVA for visual question answering (VQA) involves translating instructions into prompts that the model can understand.
- Philosophy: One task per model—this is much more difficult for a model to perform multiple steps.
- Especially important for models that would need to consider different data modalities.*
- Since LLaVA is a **large language model (LLM)**, we use a prompt to direct its actions.

```
def describe_floorplan(sketch_path: str) → str:
    prompt = (
        "Extract and list all labeled rooms and their approximate dimensions from this floorplan "
        "sketch. "
        "Reproduce the room names and measurements exactly as written. "
        "Do not summarize or invent details beyond what is explicitly labeled."
    )
    return run_llava(sketch_path, prompt)
```

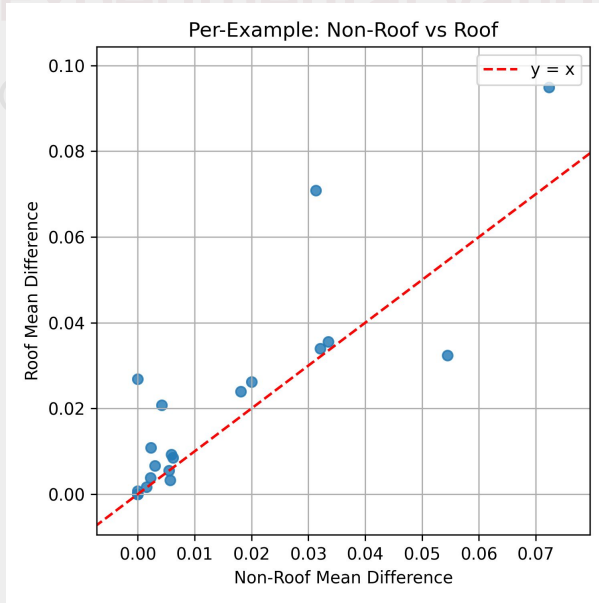
Why specifically use LLaVA?

Image Processing: *Why LLaVA?*

- 1. Structure:** LLaVA's internal structure is *known*, and it does not feature many *cross-attention layers* like other modern models have.
- 2. Experimental Validation:** We found that LLaVA outperformed GPT in our *focus* (occlusion) test.
 - 20 home images (10 with good-condition roofs, 10 with roofs in need of work)
 - Each image divided into 100 cells
 - Each cell run through the models and compared to a baseline output
 - Prompted each model to evaluate the roof
 - We would therefore expect cells covering the roof to be more important to determine the output—we expect the model to *focus on the roof*.

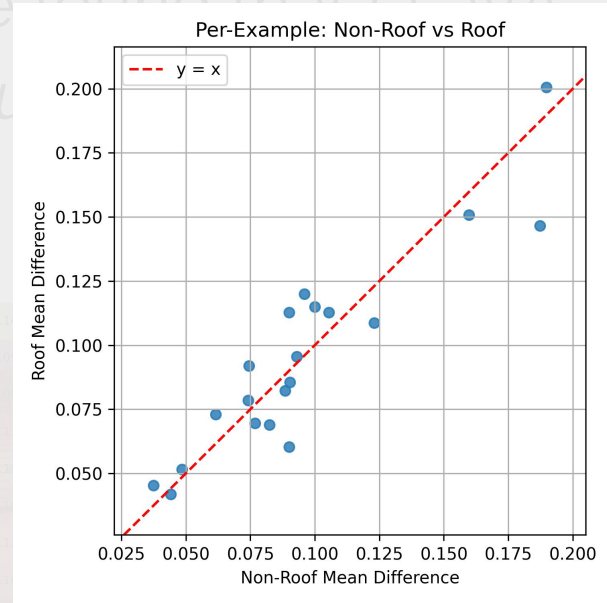
Image Processing Overall Test Results

2. Experimental Validation: We found that LLaVA is on our focus



LLaVA

Roof Mean Difference: 0.0208
Non-Roof Mean Difference: 0.0149



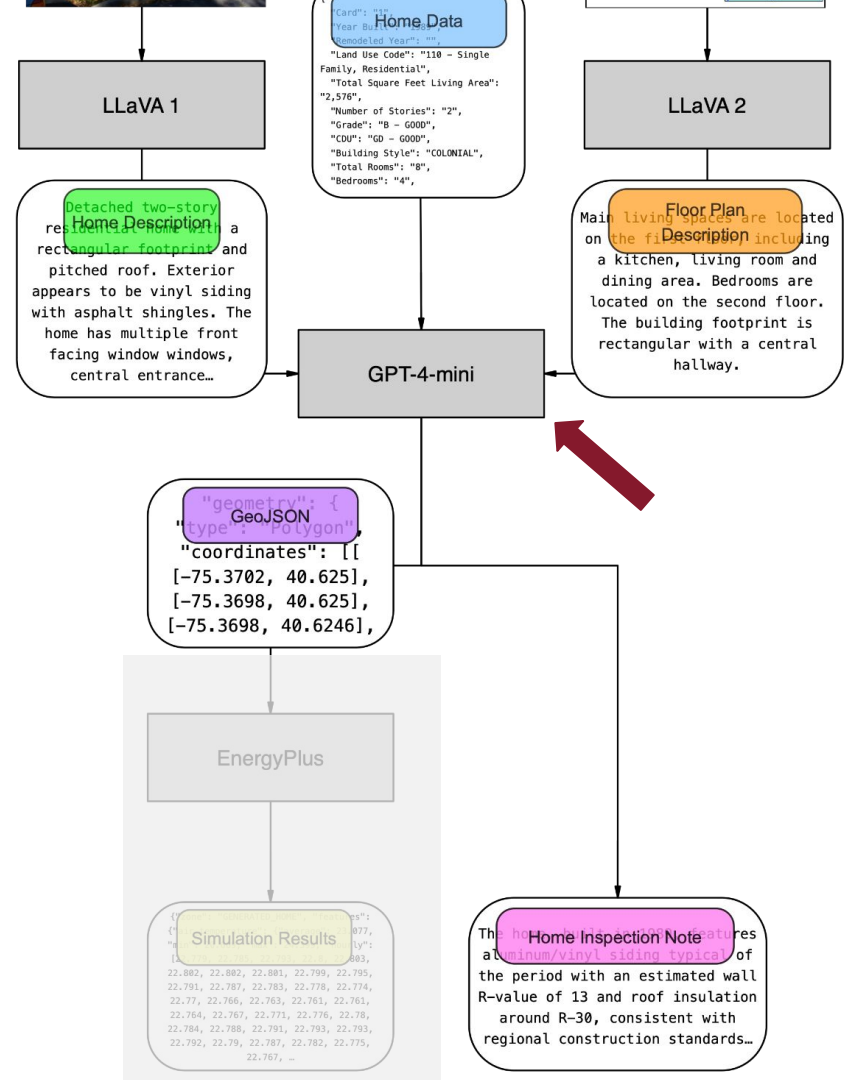
GPT

Roof Mean Difference: 0.0956
Non-Roof Mean Difference: 0.0956

Stage 3: Generating GeoJSON

Inputs: home image
description, floor plan
description, home data

Outputs: GeoJSON, home
inspection note



```
./GSU_data/  
comp_sci_coords.txt
```

```
[33.75487, -84.38967]  
[33.75487, -84.38937]  
[33.75462, -84.38937]  
[33.75462, -84.38967]  
[33.75487, -84.38967]
```

```
./GSU_data/comp_sci.json
```

```
{  
  "department": "Computer Science",  
  "university": "Georgia State University",  
  "contact": {  
    "phone": "404-413-5700",  
    "fax": "404-413-5717",  
    "chair": "Dr. Armin Mikler"  
  },  
  "research_focus_areas": [  
    "Artificial Intelligence",  
    "Machine Learning",  
    "Bioinformatics",  
    "Cybersecurity",  
    "Databases",  
    "Graphics and Visualization"  
  ],  
  "programs_offered": [  
    "B.S. in Computer Science",  
    "B.S. in Data Science",  
    "M.S. in Computer Science",  
    "M.S. in Data Science and Analytics",  
    "Ph.D. in Computer Science"  
  ],  
  "description": "The Department of Computer Science at Georgia State University offers undergraduate and graduate programs with a strong emphasis on interdisciplinary research and preparing students for careers in industry, government, and academia."  
}
```

./GSU_data/comp_sci.geojson

```
{
  "department": "Computer Science",
  "university": "Georgia State University",
  "contact": {
    "phone": "404-413-5700",
    "fax": "404-413-5717",
    "chair": "Dr. Armin Mikler"
  },
  "building_perimeter": [
    ["33.75487", "-84.38967"],
    ["33.75487", "-84.38937"],
    ["33.75462", "-84.38937"],
    ["33.75462", "-84.38967"],
    ["33.75487", "-84.38967"]
  ],
  "programs_offered": [
    "B.S. in Computer Science",
    "B.S. in Data Science",
    "M.S. in Computer Science",
    "M.S. in Data Science and Analytics",
    "Ph.D. in Computer Science"
  ],
  "research_focus_areas": [
    "Artificial Intelligence",
    "Machine Learning",
    "Bioinformatics",
    "Cybersecurity",
    "Databases",
    "Graphics and Visualization"
  ],
  "description": "The Department of Computer Science at Georgia State University offers undergraduate and graduate programs with a strong emphasis on interdisciplinary research and preparing students for careers in industry, government, and academia."
}
```

```

{
  "type": "FeatureCollection",
  "mappers": [],
  "project": {
    "id": "502caa70-b30a-4cb2-8622-17a1229f25d5",
    "name": "Generated Project",
    "begin_date": "2024-08-14T00:00:00Z",
    "end_date": "2025-08-14T00:00:00Z",
    "cec_climate_zone": null,
    "climate_zone": "4A",
    "default_template": "90.1-2013",
    "import_surrounding_buildings_as_shading": null,
    "surface_elevation": null,
    "tariff_filename": null,
    "timesteps_per_hour": 1,
    "weather_filename": "weather.epw"
  },
  "scenarios": [
    {
      "feature_mappings": [],
      "id": "7efb3e6c-be76-47c7-9210-6ea6377ef48b",
      "name": "Base Scenario"
    }
  ],
  "features": [
    {
      "type": "Feature",
      "properties": {
        "id": "2b358096-bf52-4cd5-a15f-75d8f2640fd6",
        "name": "Generated Home",
        "type": "Building",
        "building_type": "Single family",
        "floor_area": 2576,
        "number_of_stories": 2,
        "inspection_note": "The home, built in 1989, features aluminum/vinyl siding typical of the period with an estimated wall R-value of 13 and roof insulation around R-30, consistent with regional construction standards. The electric warm air heating system paired with central air conditioning appears to be in good working order, with a heating COP of approximately 0.85 and cooling COP near 3.2, indicating moderately efficient HVAC equipment. The full basement and two-story colonial design suggest moderate air leakage; the estimated air change rate of 0.35 reflects average tightness. The 12 windows, including a bay window, are consistent with the home's style but may benefit from energy upgrades. No attic insulation was noted, which is typical for this house type. Overall, the home shows evidence of some energy considerations but could benefit from insulation and window improvements.",
        "Year Built": 1989,
        "Total Square Feet Living Area": 2576,
        "Building Style": "COLONIAL",
        "Exterior Wall Material": "ALUMINUM/VINYL SIDING",
        "Heating Fuel Type": "ELECTRIC",
        "Heating System Type": "WARM AIR",
        "Heat/Air Cond": "AIR COND",
        "Bedrooms": 4,
        "Full Baths": 2,
        "Half Baths": 1,
        "Basement": "FULL",
        "Number of Stories": 2,
        "Grade": "B - GOOD",
        "hvac_heating_cop": 0.85,
        "hvac_cooling_cop": 3.2,
        "wall_r_value": 13,
        "roof_r_value": 30,
        "air_change_rate": 0.35
      }
    }
  ],
  "geometry": {
    "type": "Polygon",
    "coordinates": [
      [
        [-75.3702, 40.625],
        [-75.3698, 40.625],
        [-75.3698, 40.6246],
        [-75.3702, 40.6246],
        [-75.3702, 40.625]
      ]
    ]
  }
}

```

Generating GeoJSON

Inputs

Red property data

Interior description

Interior description

Footprint

Energy parameters

Note

Simulating with EnergyPlus

- Simulation allows us to convert our synthetic home data to **hourly energy usage data**.
 - We use **EnergyPlus**¹ (Crawley *et al.*, 2001) to run

GeoJSON Data

IDF File

EnergyPlus

Hour-by-Hour
Simulation Results

necessary data from the GeoJSON we generated in Stage 3. This data is inserted into a template **IDF** (input description file) file.

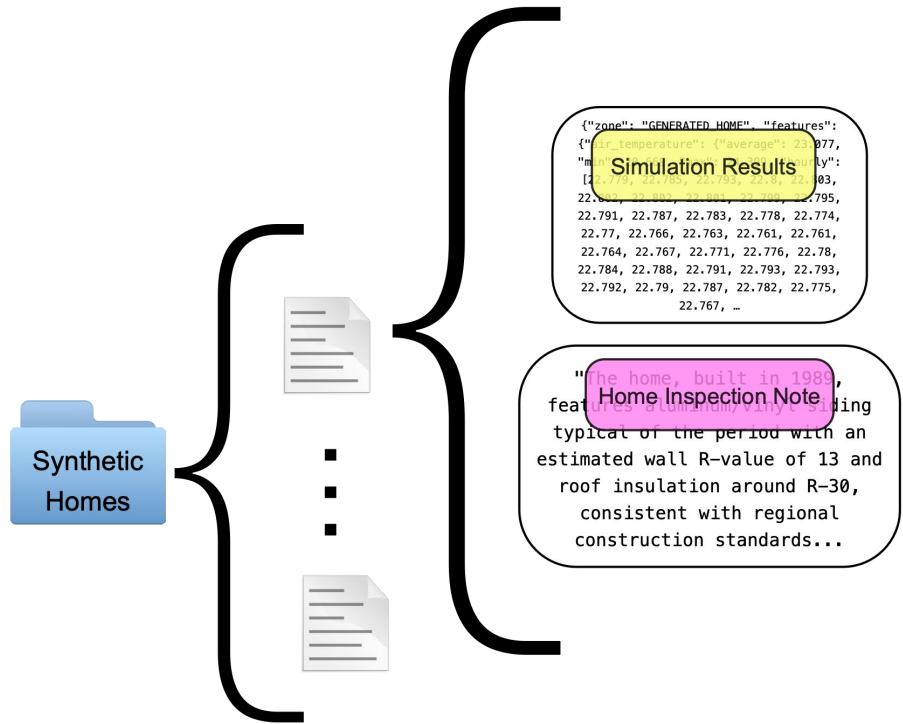
¹ EnergyPlus™ is a trademark of the United States Department of Energy.

MATERIAL,				OUTPUT:VARIABLEDICTIONARY,	
Wall Material,	!- Name	BUILDINGSURFACE:DETAILED,		IDF;	!- Key Field
Rough,	!- Roughness	Generated_Home_Roof,	!- Name		
0.2,	!- Thickness	Roof,	!- Surface	OUTPUT:VARIABLE,	
8.736295e-02,	!- Conductivity	Type		*	!- Key Value
1400,	!- Density	Roof Construction,	!-	Zone Air Temperature,	!- Variable Name
1000,	!- Specific Heat	Construction Name		Hourly;	!- Reporting
0.9,	!- Thermal Absorptance	Generated_Home,	!- Zone	Frequency	
0.7,	!- Solar Absorptance	Name			
0.7;	!- Visible Absorptance	,	!- Space	OUTPUT:VARIABLE,	
		Name		*	!- Key Value
MATERIAL,		Outdoors,	!- Outside	Heating Coil Gas Energy,	!- Variable
Roof Material,	!- Name	Boundary Condition		Name	
Rough,	!- Roughness	,	!- Outside	Hourly;	!- Reporting
0.15,	!- Thickness	Boundary Condition Object		Frequency	
2.839296e-02,	!- Conductivity	SunExposed,	!- Sun		
800,	!- Density	Exposure		OUTPUT:VARIABLE,	
1200,	!- Specific Heat	WindExposed,	!- Wind	*	!- Key Value
0.9,	!- Thermal Absorptance	Exposure		Cooling Coil Electric Energy,	!-
0.7,	!- Solar Absorptance	autocalculate,	!- View	Variable Name	
0.7;	!- Visible Absorptance	Factor to Ground		Hourly;	!- Reporting
		4,	!- Number	Frequency	
MATERIAL,		of Vertices			
Floor Material,	!- Name	0,	!- Vertex 1	OUTPUT:VARIABLE,	
MediumSmooth,	!- Roughness	Xcoordinate		*	!- Key Value
0.15,	!- Thickness	0,	!- Vertex 1	Fan Electric Energy,	!- Variable Name
1.4,	!- Conductivity	Ycoordinate		Hourly;	!- Reporting
2200,	!- Density	3.048,	!- Vertex 1	Frequency	
1000,	!- Specific Heat	Zcoordinate			
0.9,	!- Thermal Absorptance	15.469910406980384,	!- Vertex 2	OUTPUT:VARIABLE,	
0.7,	!- Solar Absorptance	Xcoordinate		*	!- Key Value
0.7;	!- Visible Absorptance	0,	!- Vertex 2	HVAC Electric Energy,	!- Variable Name
		Ycoordinate		Hourly;	!- Reporting
		3.048,	!- Vertex	Frequency	

{}zone": "GENERATED_HOME", "features": {"air_temperature": {"average": 23.077, "min": 20.669, "max": 24.399, "hourly": [22.779, 22.785, 22.793, 22.8, 22.803, 22.802, 22.802, 22.801, 22.799, 22.795, 22.791, 22.787, 22.783, 22.778, 22.774, 22.77, 22.766, 22.763, 22.761, 22.761, 22.764, 22.767, 22.771, 22.776, 22.78, 22.784, 22.788, 22.791, 22.793, 22.793, 22.792, 22.79, 22.787, 22.782, 22.775, 22.767, 22.758, 22.748, 22.736, 22.723, 22.71, 22.696, 22.681, 22.665, 22.648, 22.632, 22.615, 22.597, 22.579, 22.561, 22.543, 22.525, 22.507, 22.489, 22.471, 22.454, 22.437, 22.42, 22.405, 22.391, 22.377, 22.363, 22.348, 22.334, 22.32, 22.308, 22.297, 22.287, 22.278, 22.272, 22.266, 22.261, 22.256, 22.253, 22.249, 22.246, 22.242, 22.239, 22.235, 22.232, 22.229, 22.225, 22.219, 22.212, 22.205, 22.195, 22.185, 22.175, 22.165, 22.156, 22.15, 22.148, 22.149, 22.154, 22.161, 22.17, 22.179, 22.187, 22.194, 22.198, 22.201, 22.202, 22.202, 22.199, 22.195, 22.188, 22.18, 22.171, 22.159, 22.144, 22.127, 22.11, 22.094, 22.079, 22.068, 22.061, 22.058, 22.059, 22.063, 22.069, 22.077, 22.086, 22.094, 22.103, 22.112, 22.12, 22.128, 22.135, 22.141, 22.147, 22.151, 22.154, 22.156, 22.157, 22.158, 22.158, 22.157, 22.156, 22.155, 22.154, 22.153, 22.153, 22.153, 22.153, 22.154, 22.155, 22.157, 22.159, 22.161, 22.162, 22.163, 22.165, 22.167, 22.169, 22.17, 22.17, 22.169, 22.168, 22.166, 22.165, 22.164, 22.164, 22.166, 22.168, 22.171, 22.175, 22.179, 22.184, 22.189, 22.194, 22.198, 22.2, 22.202, 22.203, 22.202, 22.201, 22.198, 22.195, 22.191, 22.186, 22.18, 22.173, 22.165, 22.157, 22.148, 22.14, 22.132, 22.124, 22.114, 22.104, 22.093, 22.08, 22.065, 22.049, 22.031, 22.011, 21.988, 21.964, 21.938, 21.911, 21.882, 21.853, 21.823, 21.792, 21.761, 21.728, 21.696, 21.664, 21.634, 21.605, 21.579, 21.558, 21.54, 21.526, 21.515, 21.507, 21.501, 21.496, 21.491, 21.486, 21.481, 21.474, 21.467, 21.461, 21.455, 21.45, 21.443, 21.435, 21.428, 21.419, 21.41, 21.403, 21.397, 21.394, 21.394, 21.398, 21.406, 21.417, 21.432, 21.449, 21.467, 21.486, 21.504, 21.523, 21.541, 21.558, 21.575, 21.591, 21.606, 21.619, 21.631, 21.641, 21.648, 21.653, 21.655, 21.657, 21.657, 21.657, 21.656, 21.656, 21.657, 21.658, 21.66, 21.662, 21.663, 21.663, 21.662, 21.66, 21.656, 21.65, 21.643, 21.633, 21.622, 21.61, 21.595, 21.58, 21.564, 21.547, 21.527, 21.506, 21.485, 21.464, 21.443, 21.424, 21.409, 21.397, 21.388, 21.383, 21.38, 21.377, 21.375, 21.372, 21.368, 21.362, 21.354, 21.345, 21.335, 21.323, 21.31, 21.296, 21.282, 21.266, 21.249, 21.23, 21.212, 21.195, 21.179, 21.166, 21.155, 21.148, 21.143, 21.14, 21.138, 21.137, 21.138, 21.139, 21.142, 21.145, 21.149, 21.154, 21.158, 21.163, 21.168, 21.174, 21.18, 21.188, 21.195, 21.204, 21.213, 21.223, 21.235, 21.249, 21.266, 21.285, 21.307, 21.331, 21.357, 21.382, 21.407, 21.43, 21.452, 21.473, 21.493, 21.51, 21.525, 21.538, 21.549, 21.556, 21.561, 21.565, 21.566, 21.566, 21.567, 21.567, 21.569, 21.574, 21.581, 21.591, 21.604, 21.618, 21.632, 21.647, 21.661, 21.675, 21.686, 21.696, 21.704, 21.709, 21.713, 21.715, 21.714, 21.711, 21.706, 21.697, 21.686, 21.671, 21.656, 21.639, 21.621, 21.605, 21.591, 21.579, 21.568, 21.561, 21.554, 21.549, 21.543, 21.538, 21.531, 21.525, 21.518, 21.51, 21.503, 21.495, 21.488, 21.48, 21.472, 21.465, 21.457, 21.449, 21.442, 21.437, 21.434, 21.433, 21.435, 21.44, 21.448, 21.458, 21.47, 21.483, 21.496, 21.51, 21.523, 21.536, 21.548, 21.559, 21.569, 21.578, 21.585, 21.591, 21.596, 21.6, 21.603, 21.605, 21.606, 21.606, 21.607, 21.607, 21.608, 21.611, 21.614, 21.618, 21.622, 21.627, 21.632, 21.636, 21.639, 21.642, 21.645, 21.646, 21.647, 21.647, 21.646, 21.645, 21.643, 21.64, 21.637, 21.633, 21.629, 21.625, 21.622, 21.62, 21.619, 21.62, 21.621, 21.624, 21.627, 21.63, 21.634, 21.637, 21.64, 21.643, 21.644, 21.645, 21.646, 21.646, 21.646, 21.646, 21.645, 21.643, 21.641, 21.637, 21.633, 21.629, 21.625, 21.622, 21.621, 21.624, 21.631, 21.64, 21.653, 21.667, 21.683, 21.698, 21.714, 21.727, 21.739, 21.748, 21.755, 21.76, 21.763, 21.764, 21.763, 21.762, 21.759, 21.754, 21.749, 21.744, 21.739, 21.735, 21.733, 21.734, 21.739, 21.747, 21.758, 21.769, 21.781, 21.794, 21.806, 21.817, 21.826, 21.833, 21.838, 21.841, 21.843, 21.843, 21.838, 21.83, 21.82, 21.807, 21.792, 21.776, 21.76, 21.743, 21.728, 21.714, 21.703, 21.693, 21.686, 21.682, 21.679, 21.677, 21.676, 21.676, 21.676, 21.677, 21.678, 21.679, 21.681, 21.683, 21.685, 21.686, 21.687, 21.689, 21.691, 21.693, 21.695, 21.697, 21.699, 21.702, 21.704, 21.706, 21.707, 21.709, 21.71, 21.712, 21.713, 21.715, 21.715, 21.715, 21.714, 21.714, 21.714, 21.714, 21.713, 21.712, 21.711, 21.71, 21.709, 21.709, 21.709, 21.71, 21.713, 21.718, 21.725, 21.733, 21.741, 21.75, 21.758, 21.765, 21.77, 21.772, 21.772, 21.77, 21.765, 21.758, 21.749, 21.738, 21.725, 21.709, 21.69, 21.669, 21.646, 21.622, 21.6, 21.581, 21.565, 21.553, 21.545, 21.541, 21.54, 21.541, 21.544, 21.546, 21.548, 21.549, 21.548, 21.546, 21.541, 21.534, 21.525, 21.513, 21.501, 21.487, 21.472, 21.455, 21.437, 21.418, 21.398, 21.379, 21.359, 21.341, 21.323, 21.306, 21.29, 21.275, 21.261, 21.247, 21.234, 21.221, 21.207, 21.193, 21.18, 21.167, 21.156, 21.145, 21.134, 21.124, 21.115, 21.106, 21.098, 21.092, 21.088, 21.085, 21.086, 21.089, 21.096, 21.104, 21.115, 21.127, 21.14, 21.153, 21.166, 21.178, 21.187, 21.194, 21.198, 21.201, 21.201, 21.199, 21.194, 21.185, 21.174, 21.161, 21.147, 21.131, 21.113, 21.096, 21.079, 21.064, 21.05, 21.038, 21.027, 21.017, 21.007, 20.996, 20.985, 20.972, 20.958, 20.942, 20.925,

Overall: Generated Homes & Price

- We generated **258 synthetic homes**.
- Worst case cost: **\$0.0039 per synthetic home**.
 - For 258 homes, that is **\$1.0062—about \$1.01**.
 - This data can cost **thousands of dollars** to license.
- This pipeline is **open source** and **designed with modularity in mind**.



How can we demonstrate that our data
is **realistic**?

What does it even mean for our data to
be **realistic**?

Validation: A word on *realism*

- What does it mean for data to be *realistic*?
 - the overlap of synthetic values with the 10%–90% range of a trusted reference on the home characteristics being generated.
- Our trusted reference is **ResStock** (Present *et al.* 2024), a dataset engineered to match the stock of residential homes in the United States.
- We generate four variables for each synthetic home:
 - Wall R-Value
 - Roof R-Value
 - Cooling COP (Coefficient of Performance)
 - Heating COP

Validation: Results

Variable	Overlap (%)	ResStock IQR (Q1–Q3)	Synthetic Mean	Synthetic Median
Wall R-Value	93.41	0.704 (1.94–2.64)	2.79	2.29
Roof R-Value	100.00	3.35 (3.35–6.69)	5.04	5.28
Cooling COP	66.28	0.70 (3.1–3.8)	2.99	3.00
Heating COP	100.00	0.20 (0.8–1.0)	0.84	0.85

Conclusion: Our data is realistic according to our definition.

Urban level

Labeling of homes to train ML models

Potential retrofits and areas of improvement

Budget management



From Images to Energy Models: Generative AI for Urban Energy Data

Jackson Eshbaugh ¹, Chetan Tiwari ^{2 3}, Jorge Silveyra ¹

¹ Department of Computer Science, Lafayette College

² Departments of Geosciences & Computer Science, Georgia State University

³ Center for Disaster Informatics & Computational Epidemiology, Georgia State University

