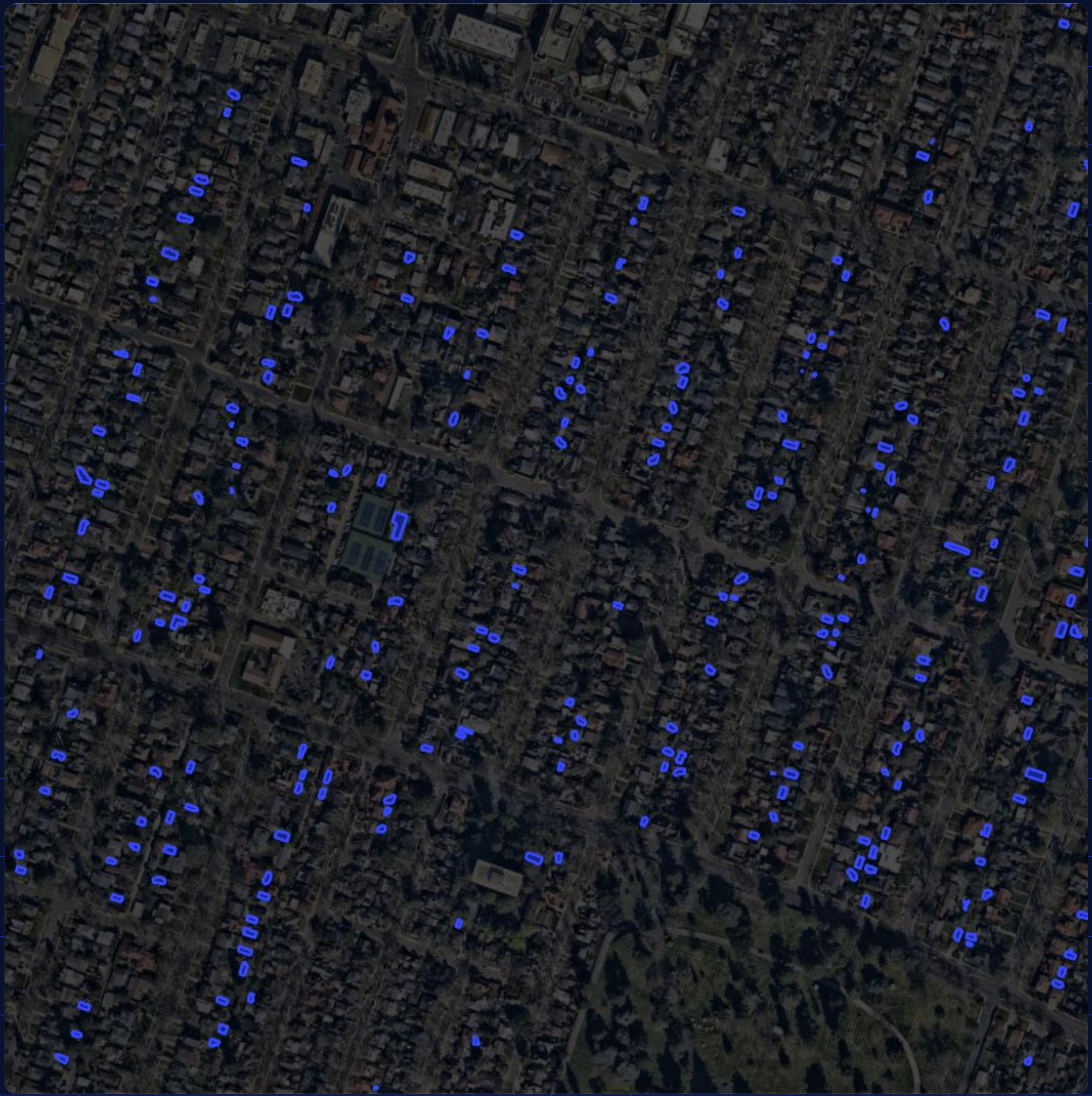
Nearmap

Exec Summit

The Pendry, Park City UT



Nearmap AI in the Age of Gen AI

Michael Bewley

Senior Vice President, AI & Computer Vision, Nearmap

Nearmap

Agenda

- Gen AI Landscape
- Gen AI Showdown (Nearmap vs Claude and Gemini)
- Nearmap Labs & Nearmap Research

Re-Defining AI

Nearmap

The AI Landscape

AI – Artificial Intelligence

- A complete system that performs a task intelligently
- Typically combining ML, software engineering, business logic.

Supervised Machine Learning (“discriminative” ML)

- Training for a specific task – example/answer pairs for training and testing
- Techniques: GLMs, XGBoost, Deep Learning, CNNs, Transformers
- Examples: Swimming pool detection, claim prediction
- Very well understood methodologies for training, evaluation, deployment

[Un/Self]-Supervised Machine Learning (“generative” ML)

- Learn patterns in data from invented, self-supervised tasks (next word prediction, fill holes in an image, …)
- No clear definition of “correct” – compressed patterns in the data
- PATTERNS != MEANING

Gen AI (loose industry term)

- Generative models are of limited use by themselves
- Usually combines Self Supervised “pre-training” & “post-training” with humans providing feedback to be useful (e.g. RLHF, fine tuning, supervised heads)

The AI Landscape (Down the rabbit hole we go…)

AI – Artificial Intelligence

- A complete system that performs a task intelligently
- Typically combining ML, software engineering, business logic.

Supervised Machine Learning ("discriminative" ML)

- Training for a specific task – example/answer pairs for training and testing
- Techniques: GLMs, XGBoost, Deep Learning, CNNs, Transformers
- Examples: Swimming pool detection, claim prediction
- Very well understood methodologies for training, evaluation, deployment

[Un/Self]-Supervised Machine Learning ("generative" ML)

- Learn patterns in data from invented, self-supervised tasks (next word prediction, fill holes in an image, …)
- No clear definition of "correct" – compressed patterns in the data
- PATTERNS != MEANING

Gen AI (loose industry term)

- Generative models are of limited use by themselves
- Usually combines Self Supervised "pre-training" & "post-training" with humans providing feedback to be useful (e.g. RLHF, fine tuning, supervised heads)

Embeddings

- Numerical representations of data produced by pre-training
- Can store embeddings of images for efficient pre-compute
- Open source, or "house-trained". Need lots of data.

Multi-Modal Models

- Jointly pre-train on mixed data sources (text, imagery, other)
- Shared embedding space links modalities (e.g. find images from text)

Zero Shot Learning

- Get an answer with no human labels required (e.g. text search aerial imagery)
- Very flexible semantics, very imprecise
- Very fiddly to improve/validate (prompt engineering?)

Few Shot Learning

- Get an answer with only limited human labels (a few, to hundreds)
- Springboards performance off pre-training / embeddings (zero shot)
- Label more – unlimited performance ceiling

Agentic AI/Systems

- Using Gen AI tools (Claude, ChatGPT, Gemini, etc.) as part of a system.
- Use memory, tools, context and multi-step actions

"We're buying access to Claude / ChatGPT / Gemini.
Can't we just give the aerial images to the AI
and get our answers?"

What are we testing?

Nearmap AI Gen 6

Industrial Grade Supervised Learning

A complete system that performs a task intelligently

- Many (millions) of labelled images
- Strictly controlled ontology of "correct"
- Rigorous training, validation, testing
- Modern deep learning (++)

- ~**1,000 person-years** of expert human labels
- 130+ **rigorously defined** property classes
- **Proprietary** global deep learning model
- **Geometrically precise** outputs
- **Deterministic**: the same image always produces the exact same answer

Gen AI from Frontier Labs

Claude (Anthropic), Gemini (Google)

- **Self supervised**, pre-trained on billions of images
- Post trained for wide variety of tasks.
- Can set up as "**zero shot**" learning, no extra training
- **Stunningly flexible** on task
- Many challenges using for repeatable, clearly defined tasks
- **Non-Deterministic**: Repeating the same request gives different results

The Tasks

1. Classification (multi-modal presence/absence in an image)
2. Counting (objects in an image)
3. Area estimation (from an image, with geometric metadata)
4. Scoring (mimicking roof spotlight index)

These are not different versions of the same approach — they are architecturally different tools.

Nearmap

We ran a rigorous head-to-head study

2,500 US residential properties · 4 tasks · 4 models across 2 providers

Tasks tested

- Swimming pool presence (classification — pool present or not)
- Roof count (object counting — how many rooftops on the property)
- Roof area in square metres (quantitative measurement)
- Roof Spotlight Index — a 0–100 condition score (scoring)

Models tested

- Claude Opus 4.5 and Opus 4.6 (Anthropic)
- Gemini 2.5 Pro and Gemini 3.1 Pro (Google)
- Nearmap AI Gen 6 (purpose-built, pre-computed, included in subscription)

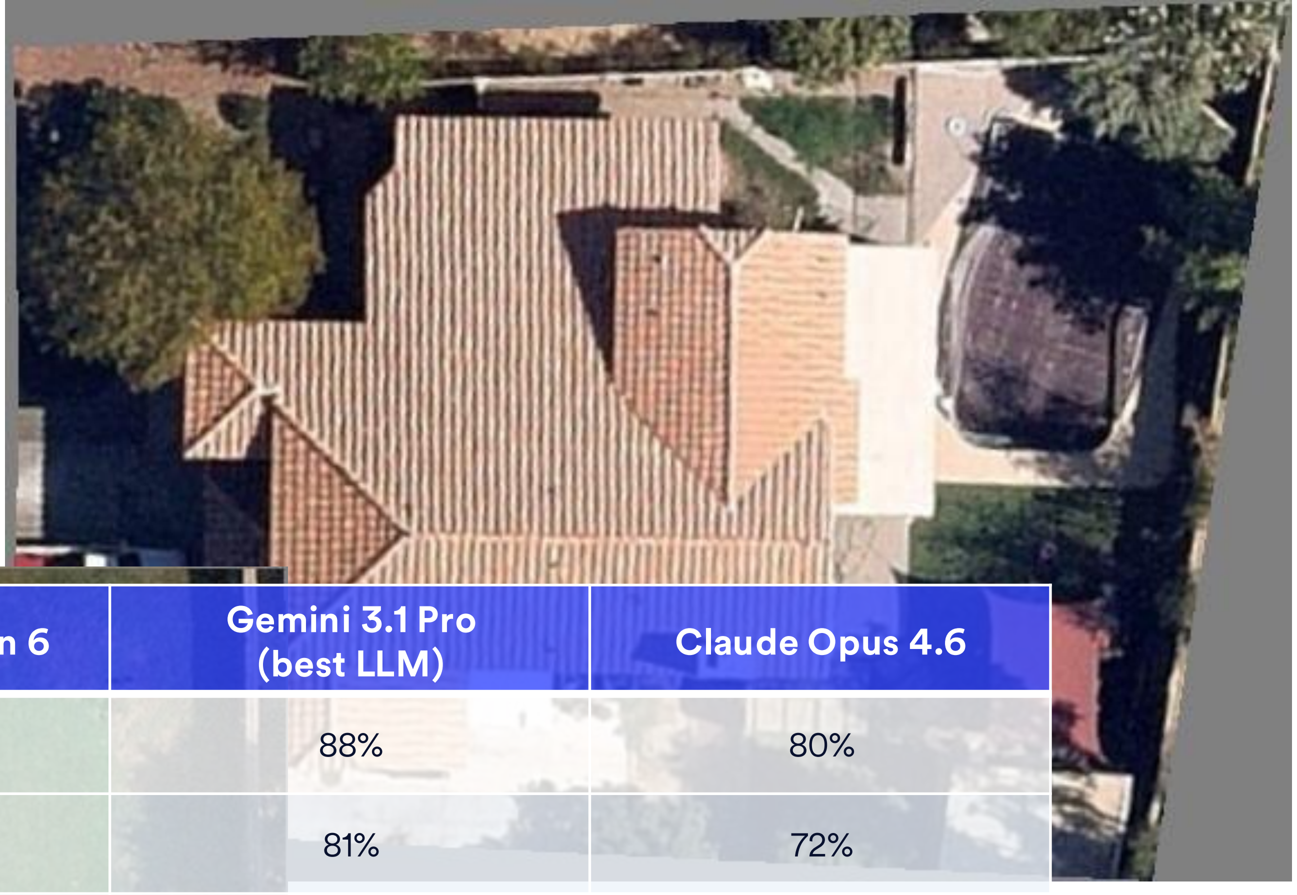
Test conditions

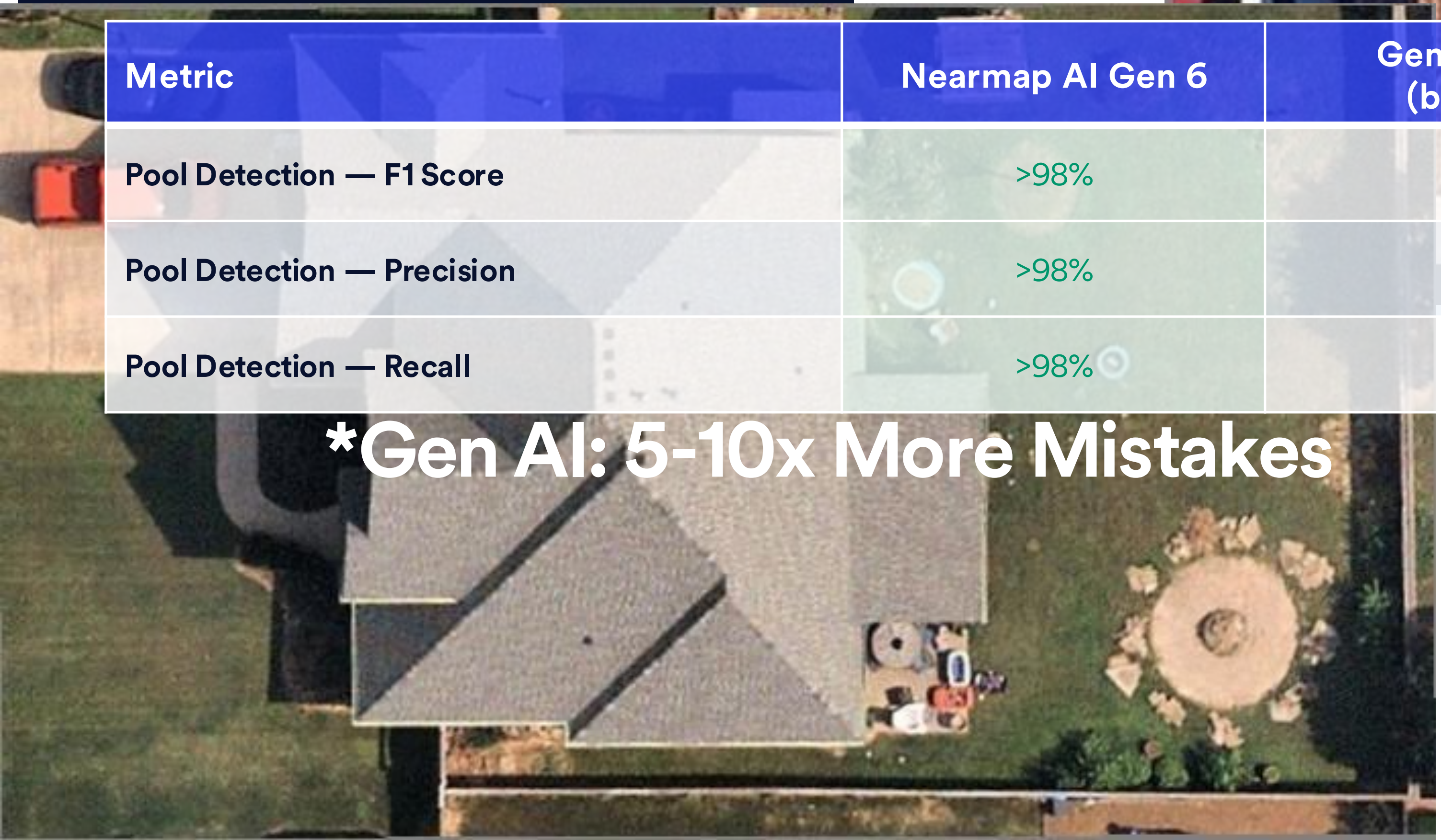
- Native 7.5 cm/pixel resolution · ~500×500 px average property · 3 identical runs per model
- Compared against ground-truth labels from Nearmap's dedicated expert annotation team

Nearmap

Classification with Gen AI

Claude: "Small above-ground pool visible in the backyard to the right of the house. It appears to be a small circular inflatable or above-ground kiddie/splash pool with blue water. There is also what appears to be a smaller blue inflatable pool or water toy nearby. A fire pit or circular landscaping feature is visible in the lower right of the yard but is not a pool."





Metric	Nearmap AI Gen 6	Gemini 3.1 Pro (best LLM)	Claude Opus 4.6
Pool Detection — F1 Score	>98%	88%	80%
Pool Detection — Precision	>98%	81%	72%
Pool Detection — Recall	>98%	98%	90%

*Gen AI: 5-10x More Mistakes

Claude: "The property shows a large residential roof with tile roofing, a driveway, some landscaping, and what appears to be a red structure or covering in the right side of the backyard. No swimming pool is visible on this property. The red area appears to be a patio cover or awning, not a pool."

Nearmap

Measurement with Gen AI

Counting (e.g. "How many rooftops?")

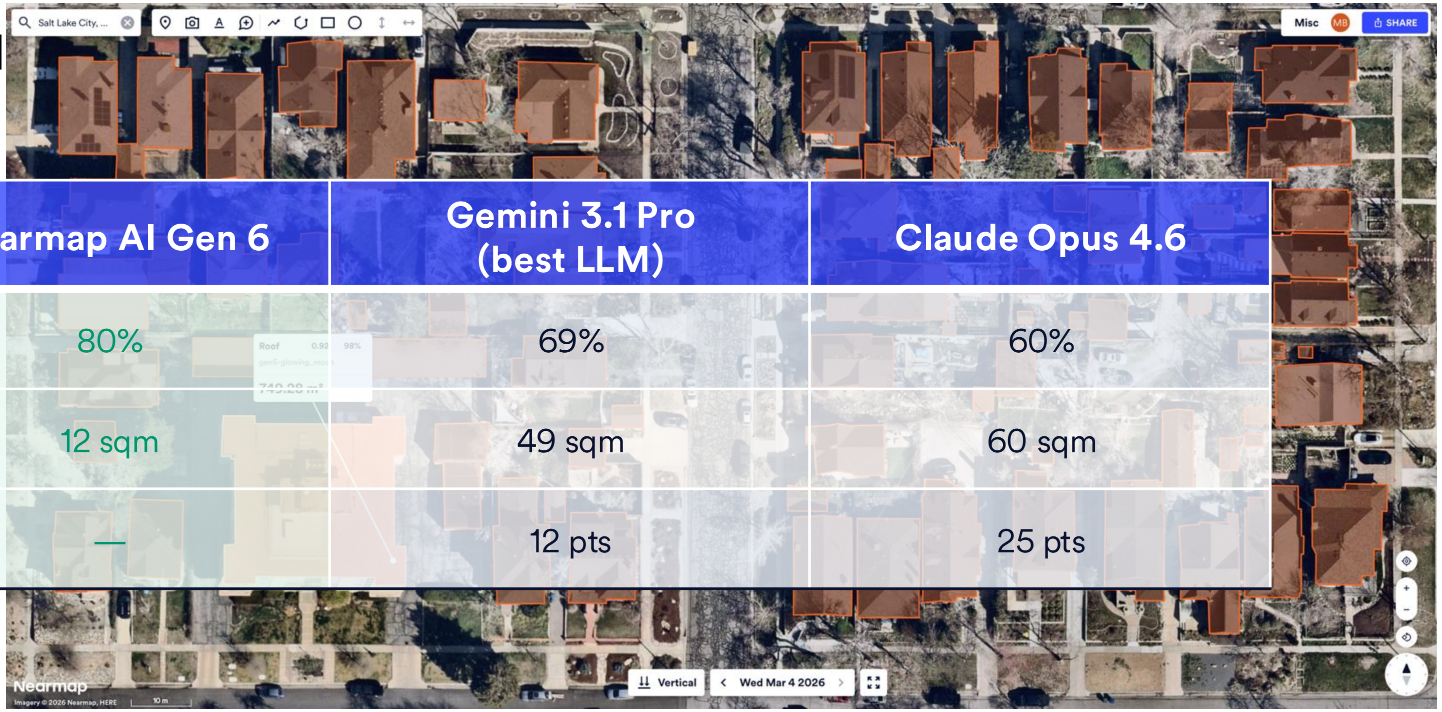
- Easy task – random guesses are good!
- For Residential Properties ~1-4 roofs
- 19% of LLM answers change between (identical) runs.

Measuring (e.g. "What is the roof area?")

- Provided clues (imagery dimensions)
- Guesses round numbers (increments of 5sqm)

Updated Claude Estimates WITH orange outlines

While Claude revised the guess from 63 down to 48 with more info… the correct answer was 68



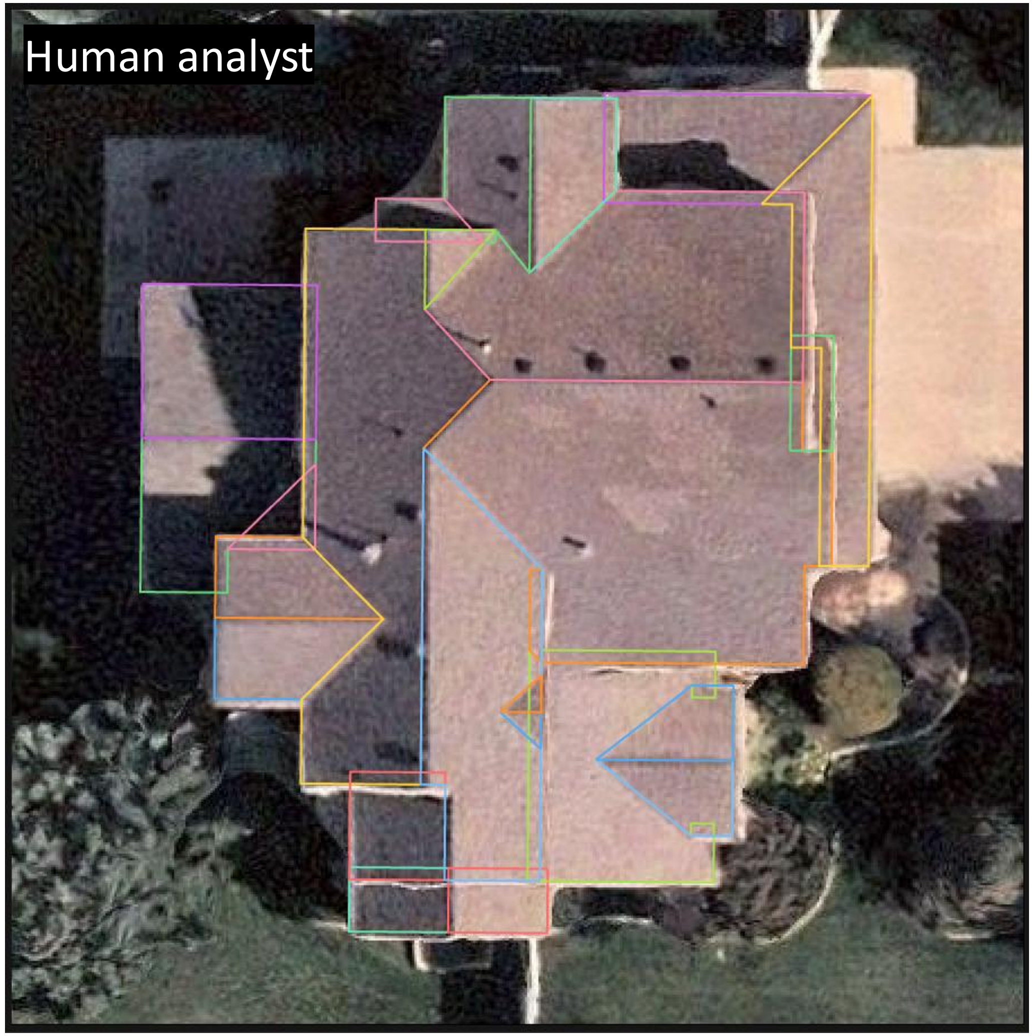
Metric	Nearmap AI Gen 6	Gemini 3.1 Pro (best LLM)	Claude Opus 4.6
Roof Count — Exact Match	80%	69%	60%
Roof Area — Mean Abs. Error	12 sqm	49 sqm	60 sqm
Roof Condition Score — MAE	—	12 pts	25 pts

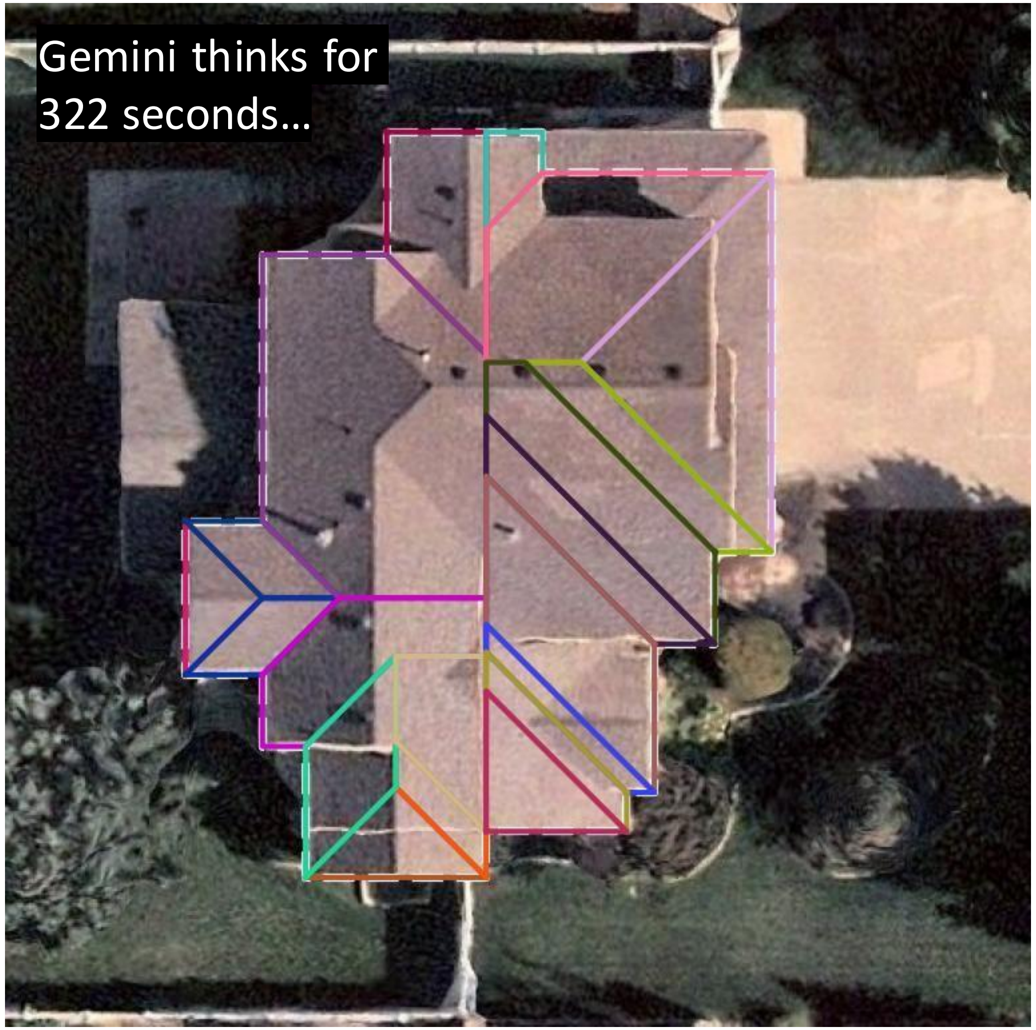
Gen AI: Costlier, Slower Responses, Lower Throughput (50-1000x)

Metric	Nearmap AI Gen 6	Gemini 3.1 Pro (best LLM)	Claude Opus 4.6
Pool Detection — F1 Score	>98%	88%	80%
Pool Detection — Precision	>98%	81%	72%
Pool Detection — Recall	>98%	98%	90%
Roof Count — Exact Match	80%	69%	60%
Roof Area — Mean Abs. Error	12 sqm	49 sqm	60 sqm
Roof Condition Score — MAE	—	12 pts	25 pts
Response Time (4 tasks)	**< 1 sec**	**55–112 sec**	**20–70 sec**
Cost per Property (4 tasks)	**Included**	**~$0.02**	**~$0.06**
Throughput	**20–1,000+/sec**	**< 10/sec**	**< 10/sec**

Nearmap

What happens if we get Gen AI to draw roof panels?





Nearmap

Factors in Choosing an AI Model

Key factors in determining the proper AI Model

Economics

Speed

Consistency

Confidence

Regulatory Defensibility

Nearmap

The economics break down at national scale

LLM API cost per property (4 visual tasks: pool, roof count, area, condition)

- Gemini 3.1 Pro: ~$0.02 per property
- Claude Opus 4.6: ~$0.06 per property

At 150 million US properties, a single national refresh costs:

- Gemini: ~$3 million USD — every time you refresh, change the model, tweak the prompts, or add an attribute
- Claude: ~$9 million USD
- Nearmap AI: 130+ attributes in a single inference pass — included in the subscription

Resolution also matters for accuracy:

- LLMs see a ~500×500 px crop of a typical residential property at full resolution
- Nearmap AI processes areas ~10× larger in its proprietary model
- Scaling to higher resolution: costs 4× worse per zoom level — quickly unworkable

Nearmap

Speed & throughput: a 100× gap

Response time per property (4 sequential visual tasks)

- Claude Opus 4.6: 20–70 seconds
- Gemini 2.5 Pro: 55 seconds
- Gemini 3.1 Pro: 55–112 seconds
- Nearmap AI (API): < 1 second

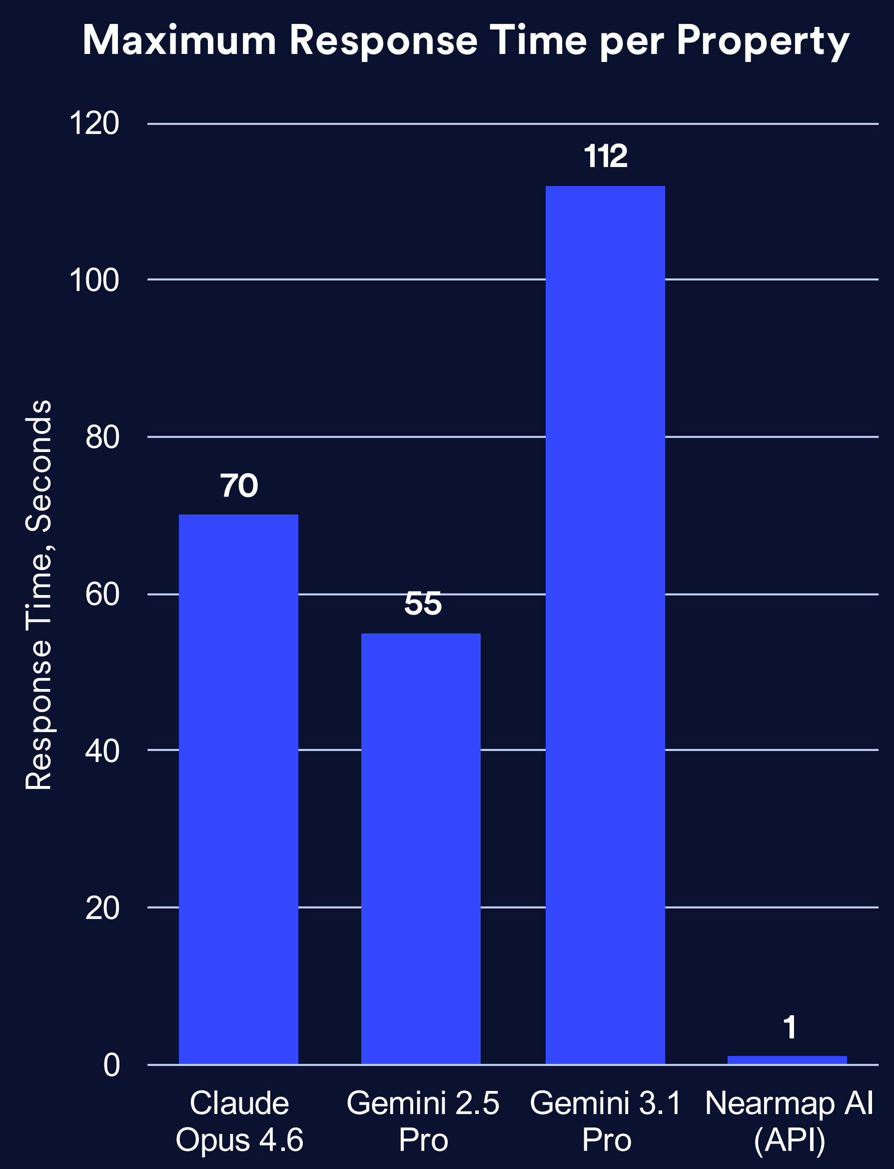
Throughput — properties processed per second

- Any LLM tested: < 10 properties / second (at maximum public rate limits)
- Nearmap AI real-time: up to 20 properties / second
- Nearmap AI bulk: 20 to 1,000+ properties / second

What this means at national scale

- LLMs at max rate: hundreds of thousands of properties per day — months to cover the USA
- Nearmap AI bulk: 150 million properties processable in hours, not months

Gen 6 launch 2024: 3 million km² processed in 10 days on a 10,000-node GPU cluster.



LLM answers change — even when nothing changes

3 identical runs, same images, same prompts. How often did the answer differ?

Claude Opus 4.6 (most consistent LLM)

- Pool presence: 0.2% of properties got a different answer → 300,000 unstable at US scale
- Roof count: 2.2% of properties got a different answer
- Roof area: mean spread of 5.4 sqm between runs
- Roof condition: mean spread of 0.5 points between runs

Gemini 3.1 Pro (most accurate LLM)

- Pool presence: 2.6% of properties changed → 3.9 million unstable at US scale
- Roof count: 18.6% of properties changed → nearly 1-in-5 gives a different answer each time
- Roof area: mean spread of 52.7 sqm — nearly as large as the error itself
- Roof condition: mean spread of 9.7 points between runs

Nearmap AI Gen 6: identical answer every time — deterministic by design.

Model/Prompt updates change your results

You don't train an LLM like Gemini or Claude – you steer it awkwardly, from a distance with prompts, and have to rerun large validation sets to explore for a better solution, THEN rerun at scale once you find one.

Claude Opus 4.5 → Opus 4.6 (a minor version increment)

- Pool presence: 2.3% of properties changed answer
- Roof count: 20% of properties changed answer
- Roof area: mean difference of 22.7 sqm
- Roof condition: Opus 4.6 scored 7.8 points lower on average — systematically different risk view

Gemini 2.5 Pro → Gemini 3.1 Pro

- Pool presence: 11.8% of properties changed answer
- Roof count: 30.2% of properties changed answer
- Roof area: mean difference of 52.5 sqm
- Roof condition: shifted 9.2 points in the opposite direction to Claude

Nearmap

Regulatory compliance: Rife with Challenges

State insurance rate filings challenges:

- Reproducible model outputs — same input must always produce same output
- Documented training data and methodology
- Model versioning with full audit trails
- Consistent behaviour over time
- Calibrated confidence scores ("80% confidence" must mean ~80% accuracy)

Nearmap AI delivers all of these

- Gen 6 ships with a 59-page System Release Report documenting every layer
- Per-class precision, recall, F1, and failure mode analysis
- Platt and Temperature scaling for calibrated confidence on every prediction
- Full version lineage at system, model, and individual feature level

LLM APIs deliver none of these

- Non-deterministic outputs that cannot be reproduced
- No training data lineage, no model versioning, no audit trail
- No confidence calibration — hallucinations are a feature, not a bug

only

21%

of companies report currently having a mature model for AI governance

-Deloitte 2026

Nearmap

So where does Gen AI fit? (Hint: we use it extensively at Nearmap)

Nearmap

Gen AI is transformative — for the right jobs

We are not "anti" Gen AI. We use it extensively — the question is the right tool for the right job.

Underwriting Document Intelligence — ideal for Gen AI

- Parse hundreds of pages of underwriting guidelines automatically
- Extract individual rules (e.g. "roof must be in good condition for eligibility")
- Map rules to Nearmap AI layers, scores, and 3rd-party data sources
- Already running on hundreds of US insurer guideline documents

Accelerating New AI Layer Development — ideal for Gen AI

- Text-to-image search across billions of aerial images to find rare object classes
- Few-shot bootstrapping to accelerate labelling of new property features
- LLM-assisted ontology definition to precisely specify new layer definitions
- Result: new AI layers reach production faster with better edge-case coverage

Accelerating Software Development — ideal for Gen AI

- Claude Code is transformative. We're finding new ways to deliver an onslaught of value.
- Take this study… this presentation. The code, testing, analysis, write up – massively faster.

The most powerful approach to business innovation combines both Gen AI and Purpose-built models

---

Purpose-built AI for production property intelligence
+
Gen AI for the workflows around it

The right tool for the job

Need	Nearmap AI	Gen AI
Classify: pool, solar, roof type	✓ >98% F1, deterministic	~ Approximate, inconsistent
Count objects (rooftops, trees)	✓ Geometric precision	~ Approximate only
Measure areas & geometries	✓ Calibrated measurements	✗ Cannot measure from imagery
130+ attributes in one pass	✓ Single inference pass	✗ One API call per question
National scale (150M properties)	✓ 1,000+ properties/sec	✗ <10 properties/sec
Regulatory audit trail	✓ 59-page system report	✗ No docs, no version control
Reproducible outputs	✓ Deterministic	✗ Stochastic by design
Calibrated confidence scores	✓ Per-class calibration	✗ Not available
Parse underwriting documents	✗ Not the right tool	✓ Exactly what Gen AI does best
Explore and search for novel things / R&D prototyping	~ Possible but costly	✓ Fast, flexible

Nearmap

"The question was never Gen AI or Nearmap AI.
It's: what's the right tool
for each part of the job?"

— Nearmap TM033 Study Conclusion | March 2026

Nearmap

Nearmap Labs

Nearmap Labs

A window into the future of property intelligence

Showcase of foundational technologies

Early access to our thinking and building

Behind-the-scenes look at the data and models shaping Nearmap's future

Nearmap

Nearmap Labs

A window into the future of property intelligence

- Agentic Rules Mapper + SERFF Explorer
- Zero-Shot Learning
- Few-Shot Learning
- Foundation Models
- Polygon Deep Learning
- Temporal Reasoning Engine
- Agentic Change Detection

Nearmap Research

Technical whitepapers and analyses

Insurance Lift Analysis with Image Recency

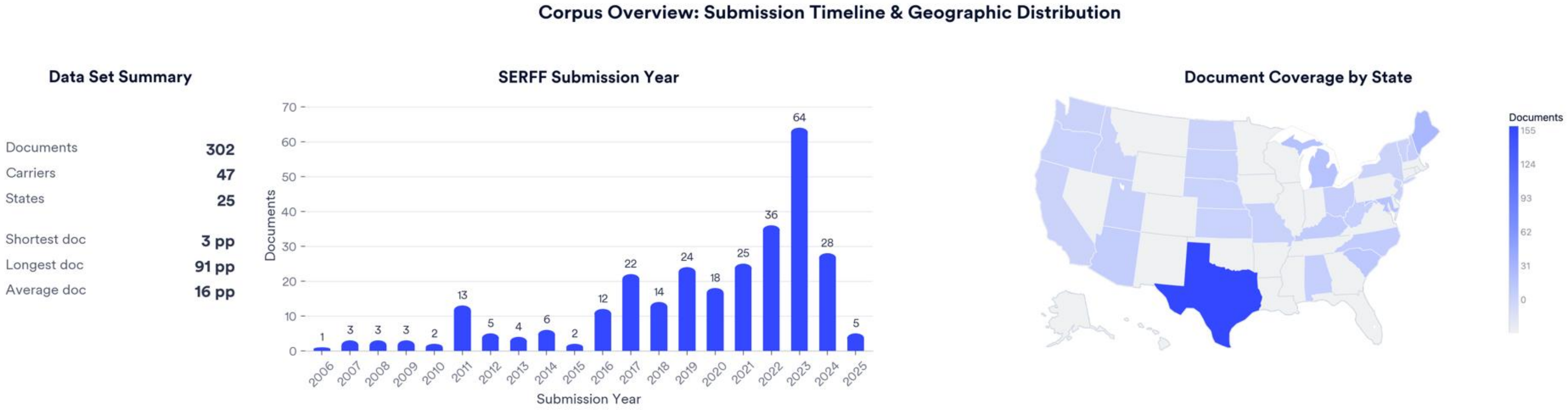
National Roof Age Study

Nearmap

Nearmap Labs #1b: SERFF Exploration

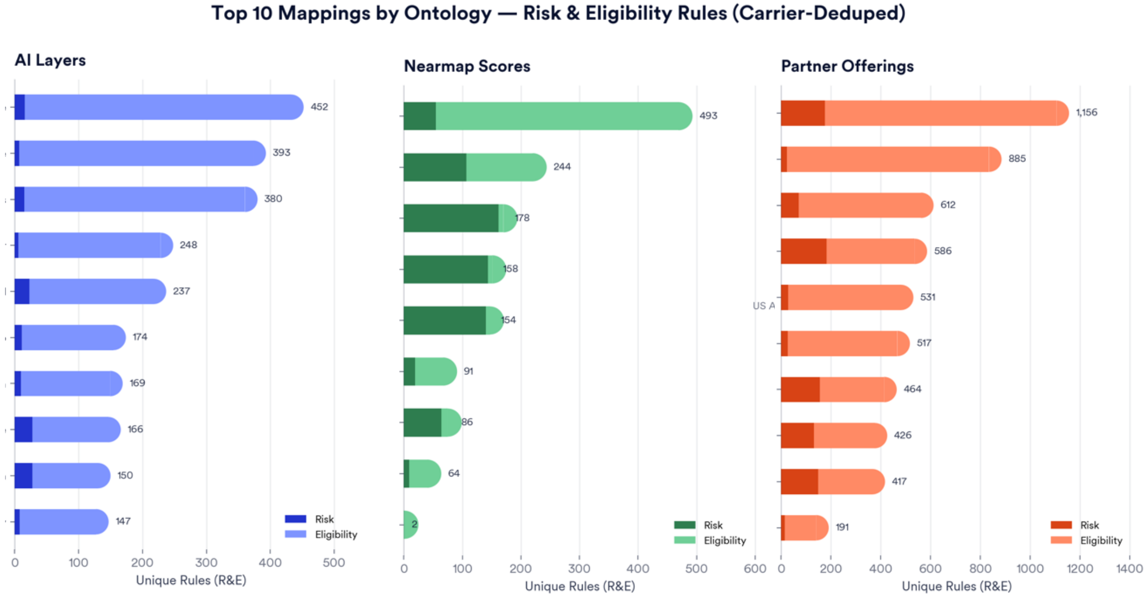
What is it?

- The obvious next step: **>300 underwriting guidelines docs** from the open SERFF platform covering **47 carriers** in **25 states**
- **Analysis and Exploration** app, and a **Whitepaper**



So What?

- Unique **strategic understanding** of the whole underwriting industry, and our place in it
- Driving **use case focused innovation & insights** as we build scores, layers and experiences



Nearmap

Documents | Data Sources | Property Features | Search

By Carrier | By State

INSURANCE COMPANY	5
Access Home Insurance Company	1
Acuity, A Mutual Insurance Company	2
Aegis Security Insurance Company	26
Agricultural Workers Mutual Auto Insurance Company	2
Alabama Insurance Underwriting Association	2
Allied Trust Insurance Company	8
Allstate Indemnity Company	14
Allstate Vehicle and Property Insurance Company	9
American Bankers Insurance Company of Florida	19
American Family Connect Property and Casualty Insurance Company	4
American Hallmark Insurance Company of Texas	3
American Modern Home Insurance Company	4
American Modern Property and Casualty Insurance Company	37
American National Lloyds Insurance Company	27
American National Property And Casualty Company	1
American Reliable Insurance Company	1
American Risk Insurance Company, Inc.	1

Select a document to view.

Select a document on the left first

ELIGIBILITY RISK ROUTING PROCEDURE APPLICABILITY

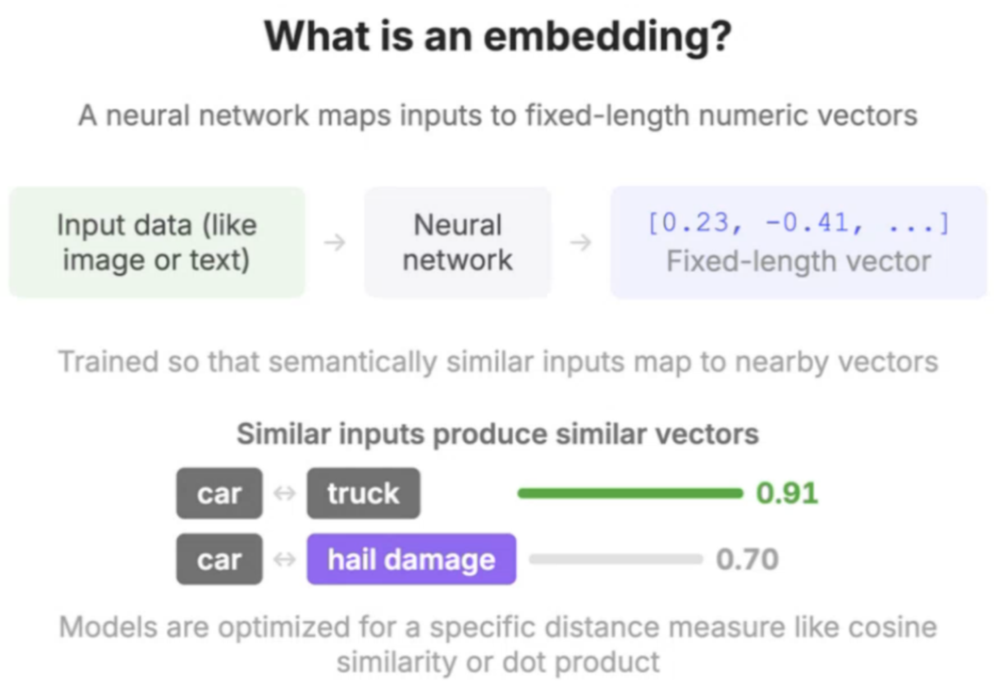
Nearmap Labs #2: Zero Shot Learning (Text Search Embeddings)

What is it?

- Embeddings executed on vertical and multi-angle imagery of properties
- Current demo at 130k properties / **1M images** (Billion image, **national scale data** in next couple of months)
- Search for arbitrary keywords, including obliques

So What?

- Developed initially for internal purposes – efficiently building new AI Layers
- A Caution: Like all **embedding** approaches - **approximate search only**
- **Not exhaustive, variable accuracy**
- **Flexible search**, but **performance un-quantifiable**



Nearmap

Property search

Dataset

2026-05-12 21:30 · 132,875 properties · 132,875 vert · 1,053,609 oblique · siglip2_so400m_parcel_vert_naflex / siglip2_so400m_parcel_obliques_naflex

Refresh

Query text	Mode	Rerank	Top K	
e.g. swimming pool, red roof, dirt driveway	Both	Max	10	Search

Why not wrap it in a MCP Connector?

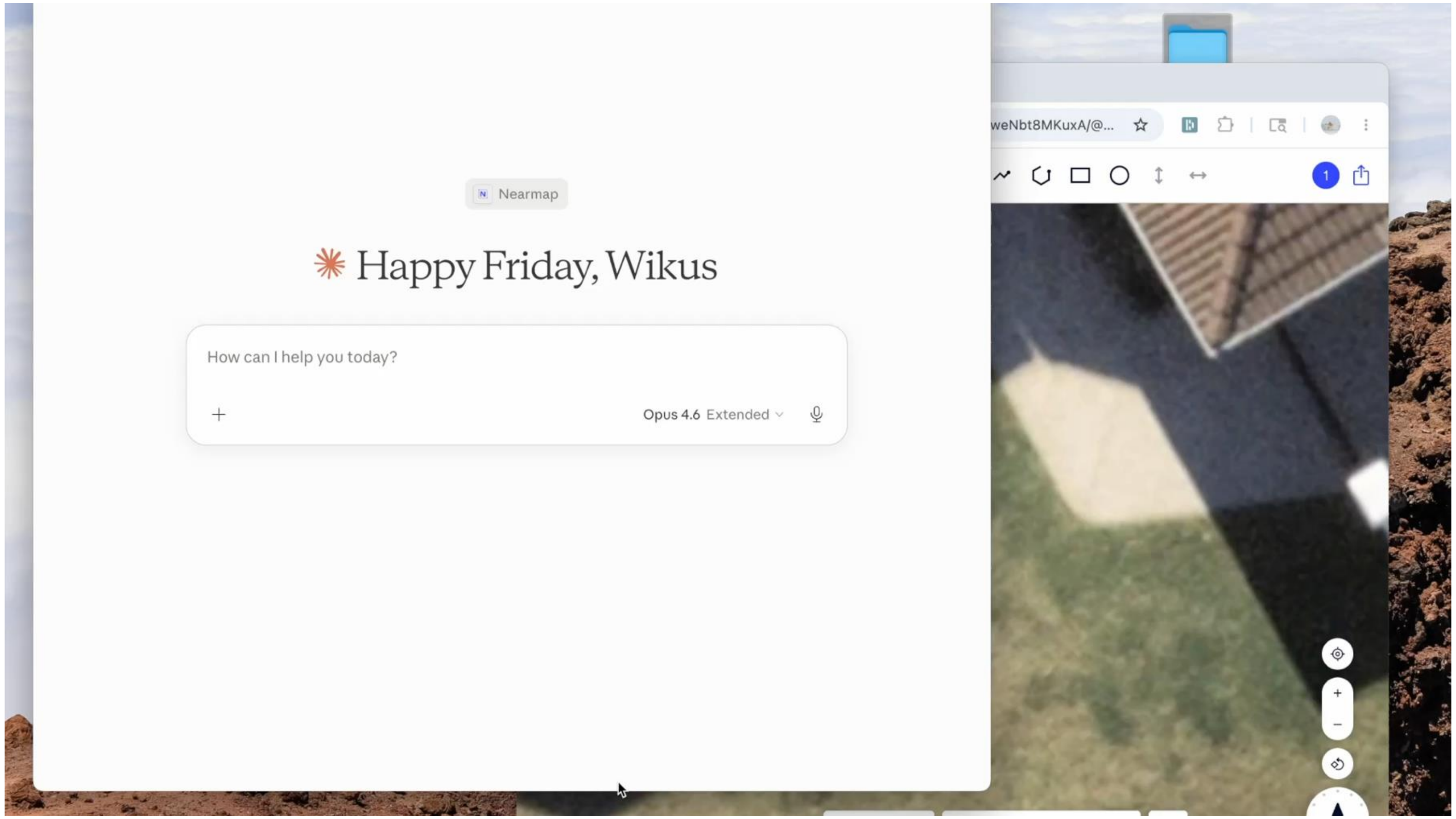
An MCP is not the hard thing – it's like an API.

It's the data and technology capabilities that it connects an LLM to that matter.

This one is a connector to the Zero Shot Embedding Search demo

… and some basics like image viewing/reviewing

Embeddings are bad at combinations – so the MCP does basic "this AND NOT that" style expresxsions



Nearmap

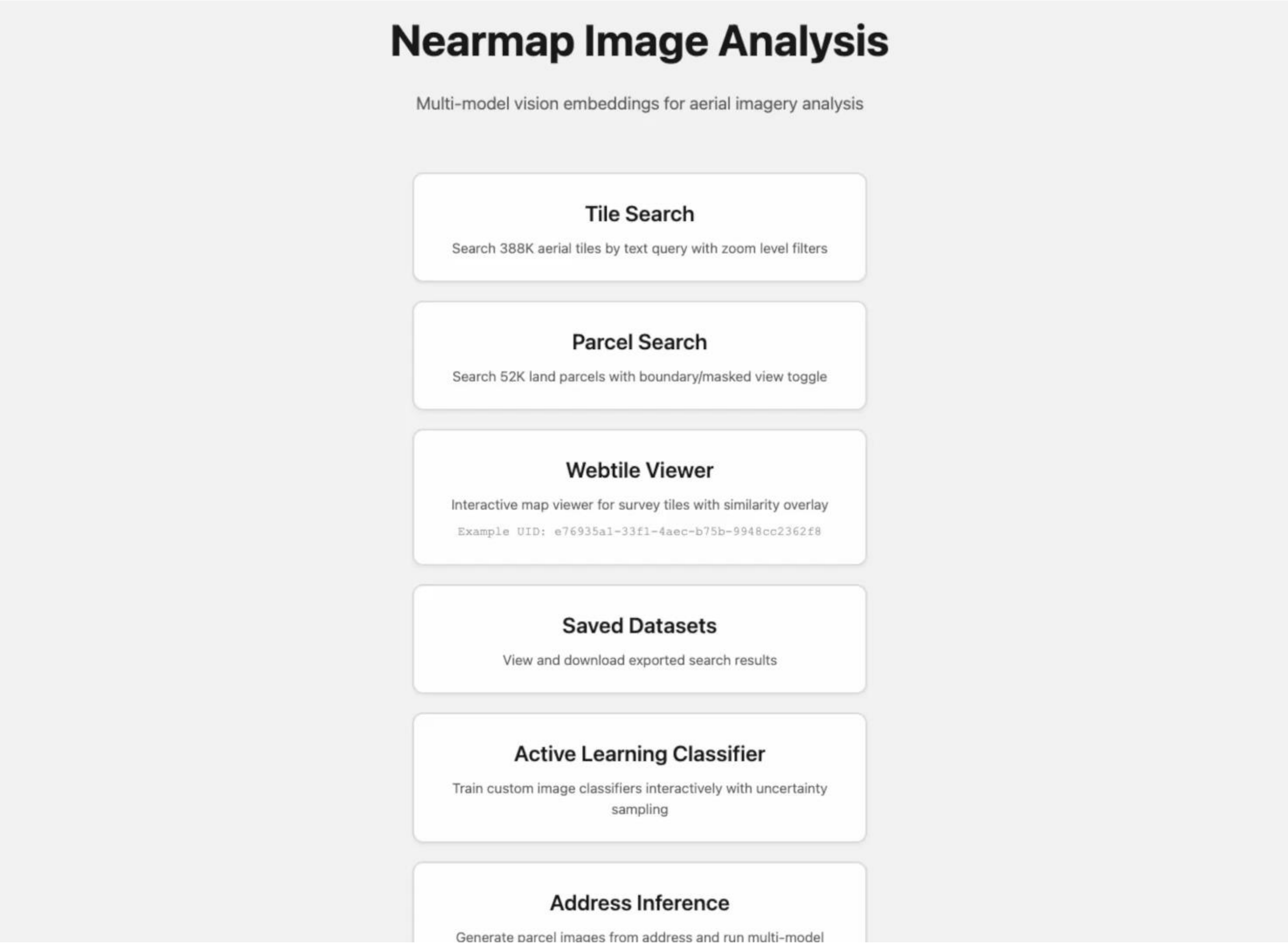
Nearmap Labs #3: Few Shot Learning

What is it?

- **Start** with zero shot **approximate search**
- **Click some examples** to correct the answers
- **Train a supervised embedding model head**
- Refine until **desired performance** reached

So What?

- **Smooth scaling** beyond text search
- **Verifiable performance** from human supervised answers
- Used internally for **efficient** AI Layer creation
- **Customer specific** AI Layers on our embeddings?
- **Custom embeddings?**



Nearmap

Foundation Model Workstreams

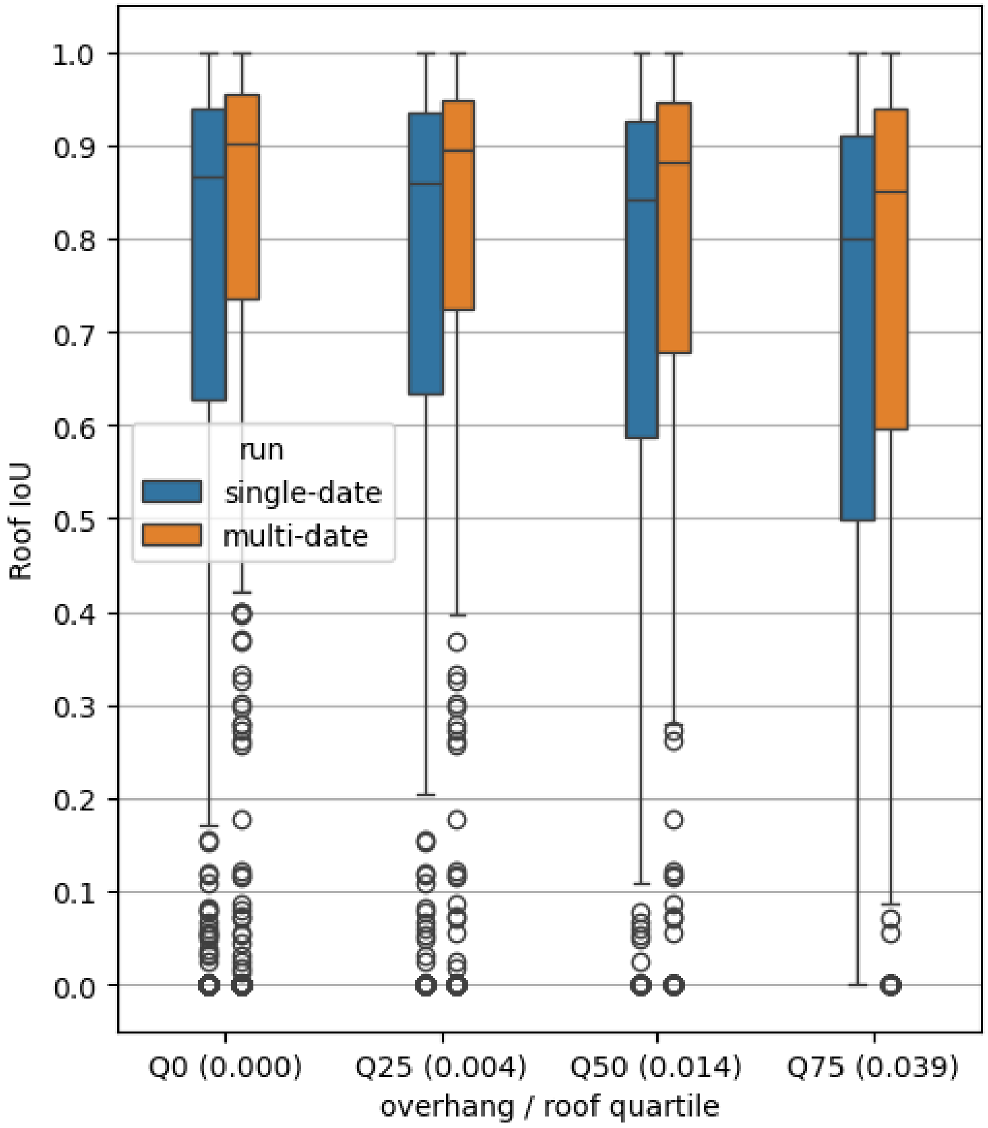
Model training experiments with hundreds of millions to **billions of images in training data**

Open source embedding models like DINOv3 are great for single image – they have learned about the structure of the world from internet scale imagery.

Nearmap trained embedding models… are adapted to the nuances of our data set:

- Learn implicit 3D reconstruction from multi-view
- Learn temporal patterns from multi-date
- Learn multi-modal info from near infrared, DSM, DEM

Distribution of roof IoU by run, filtered by overhang / roof quartile

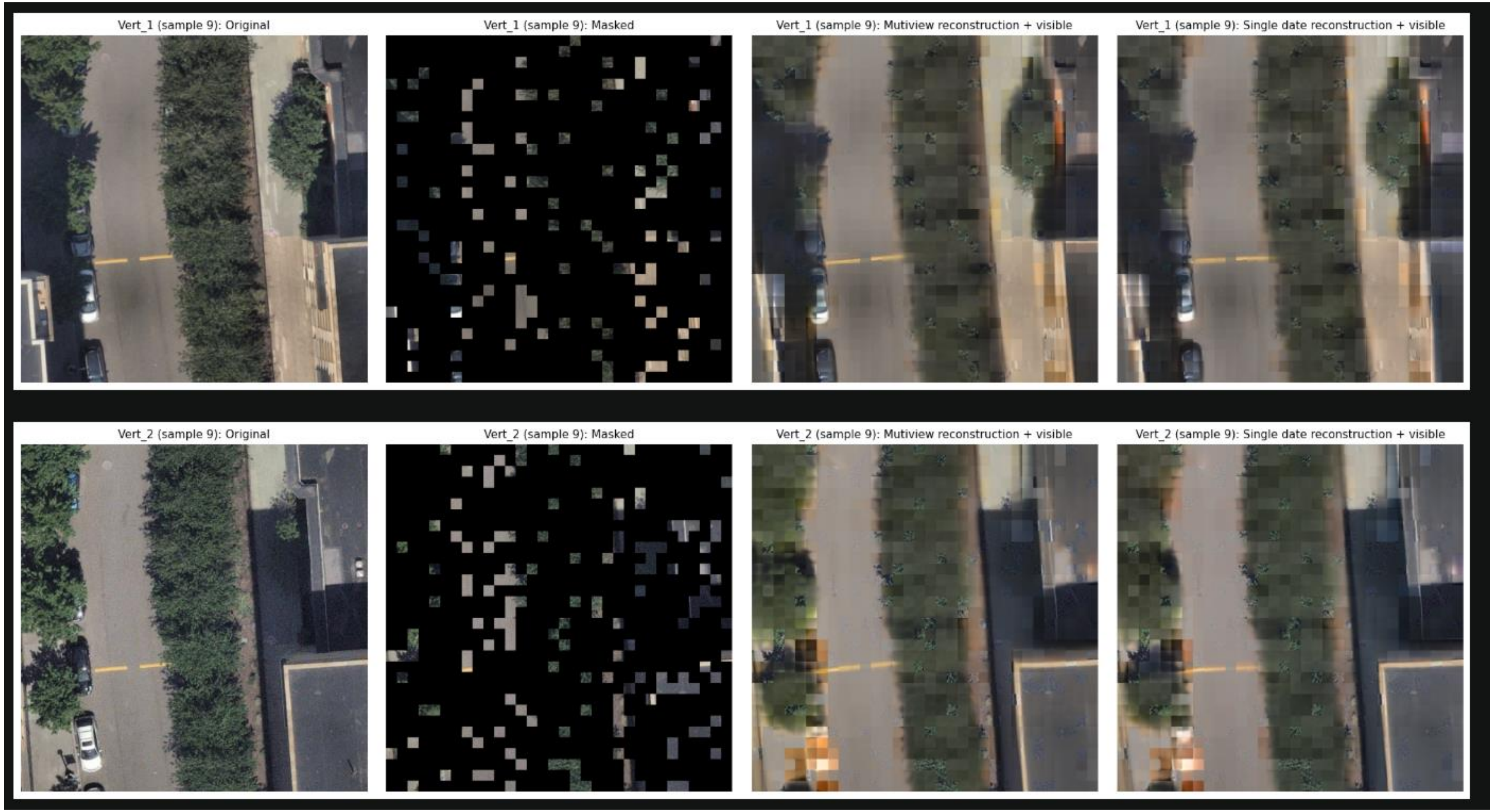


Nearmap

Foundation Model Workstreams

Columns (Left->Right)

- **Original** Image
- **Masked** Image
- Reconstruction from masked using **multi-view**
- Reconstruction from masked using **multi-date**

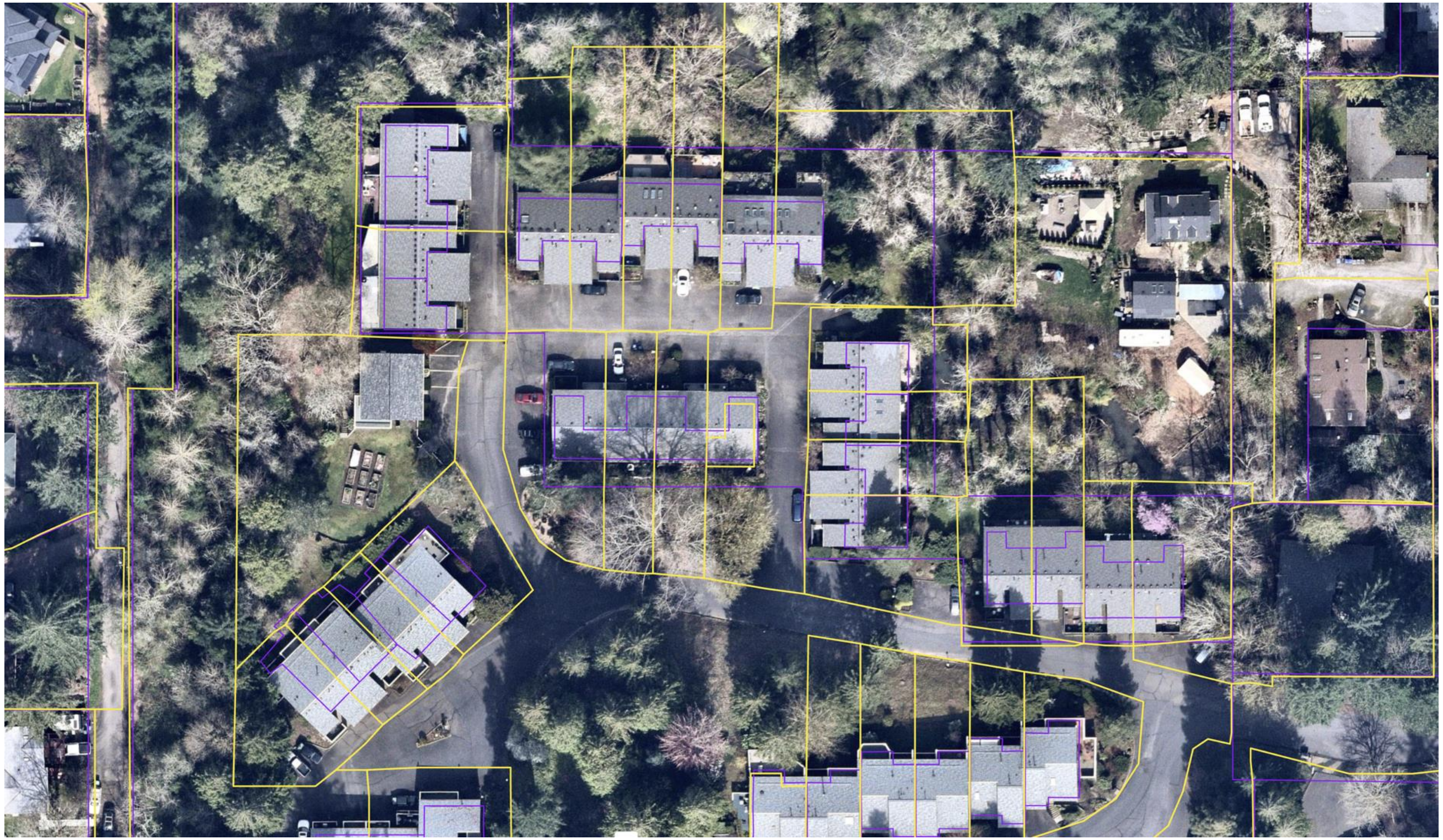


Nearmap

Nearmap Labs #5: Polygon Deep Learning (Visual Parcels)

What is it?

Proprietary "direct to polygon" transformer deep learning model Identify property boundaries purely from imagery, or enhancing existing parcels.

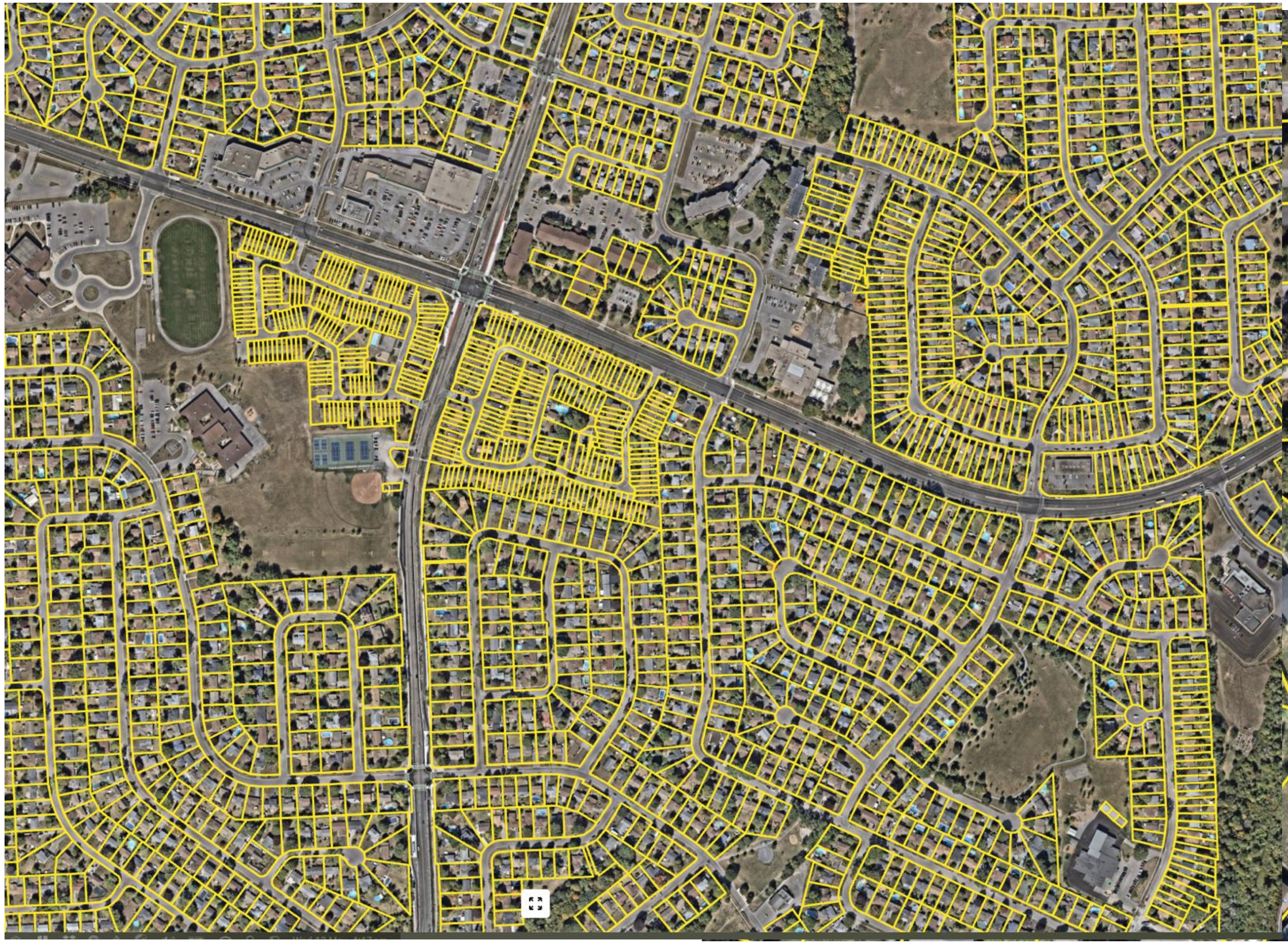


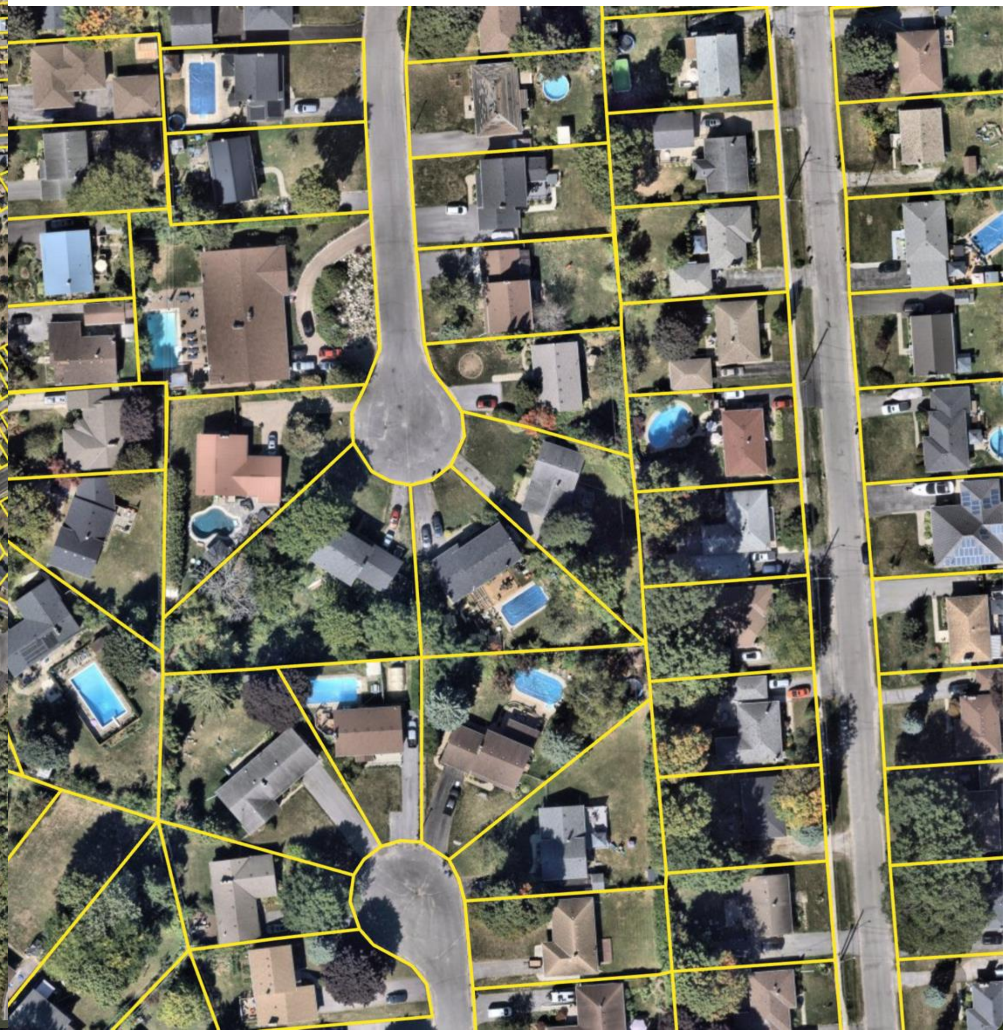
So What?

1. **Bad Parcels** are common (misalignment, errors) and break data
2. **New Developments** lag with parcel boundaries
3. **Whole Jurisdictions** (e.g. Canada) lack parcels
4. **Other Polygons...?**

Nearmap

Visual Parcels (pure imagery) – Kingston, Canada



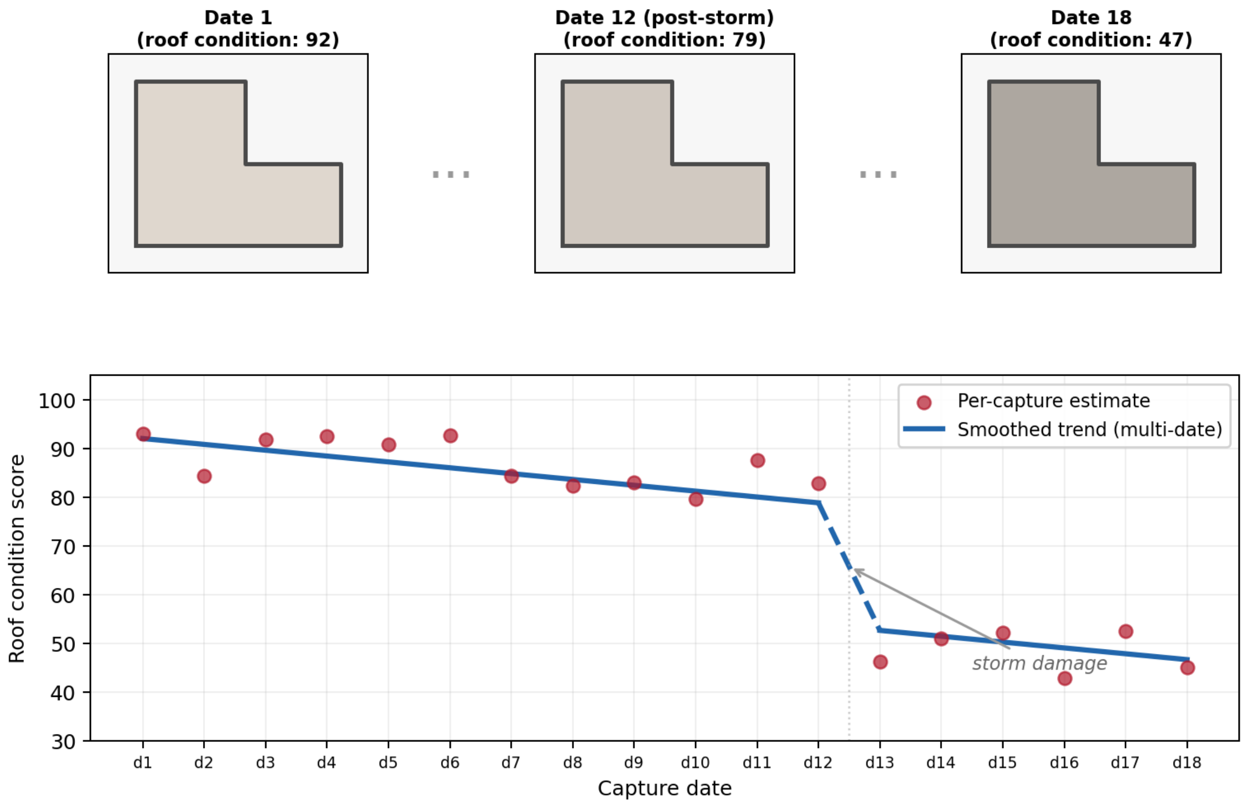


Nearmap

Nearmap Labs #6: Temporal Reasoning Engine

What is it?

- New multi-date, multi-source first reasoning system
- First used for roof age – every building, every date, multiple data sources
- Every date on every building now **has full Gen 6 AI**
- Extra **temporal embedding** models to enrich
- 3rd party data ingest (assessor & permit)
- Meta algorithms reasoning about full history
- Simplified data set



So What?

- New Damage Conflation product
- Pre-cat footprints, reasoning about cloud and occlusion
- "Best and latest" view – **unified data set updated each survey** in an event
- What's next? Applying this to **everything about a roof**

Nearmap

Event Name

Hurricane Milton

Event Jira

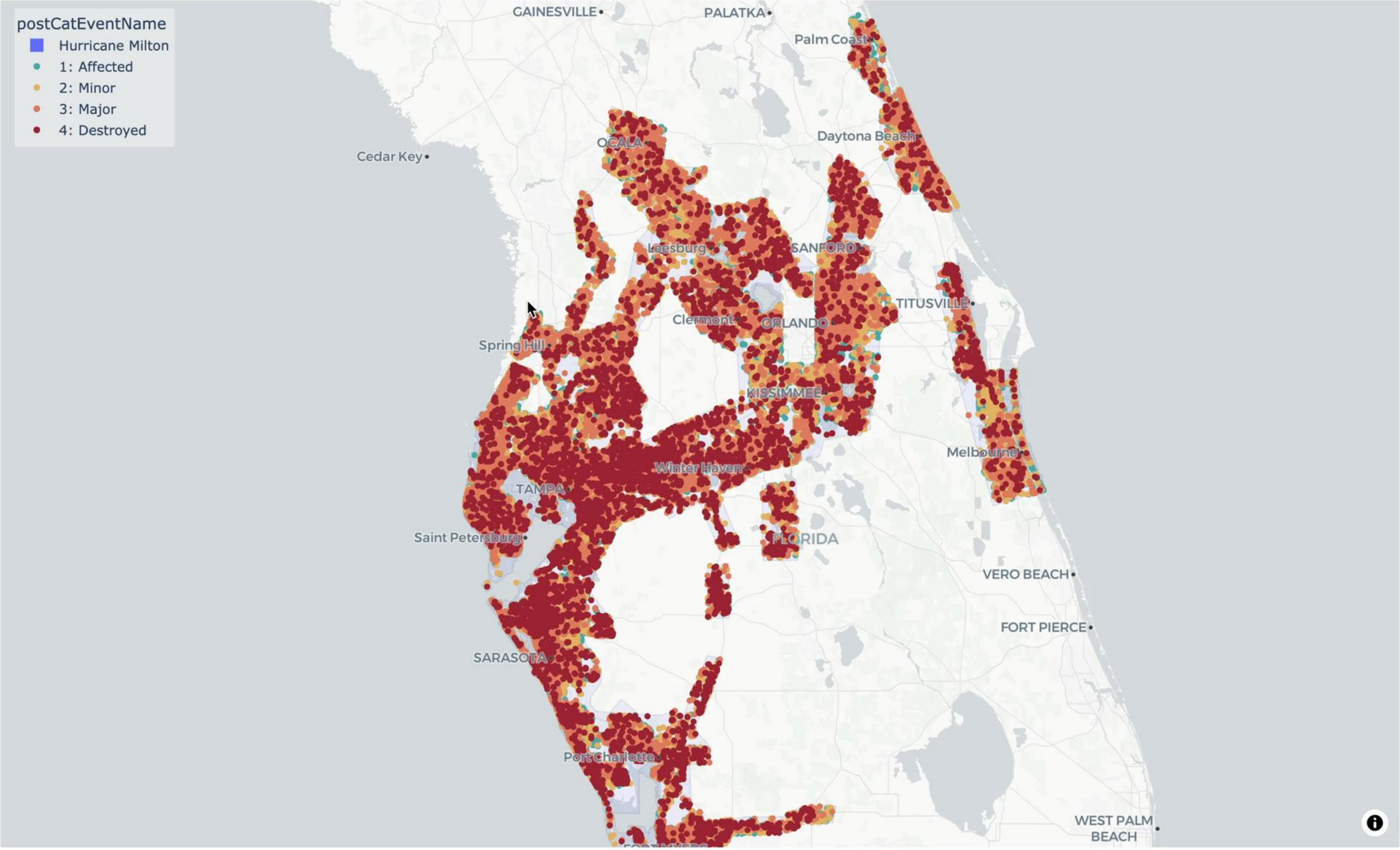
PCAT-299

Latest Imagery Online Time

2024-10-21 13:12:08

Task ID: 429943c3-c841-5848-af1b-20eb5dd9ebb5

Number of Buildings	Impacted Buildings	Major or Destroyed Buildings	Buildings with Pre-Event Data
4,297,711	244,476	42,474	4,094,064



Number of Filtered Buildings from Event

1,860,582

Filter Thresholds:

Minimum Area (sqm) 30.0

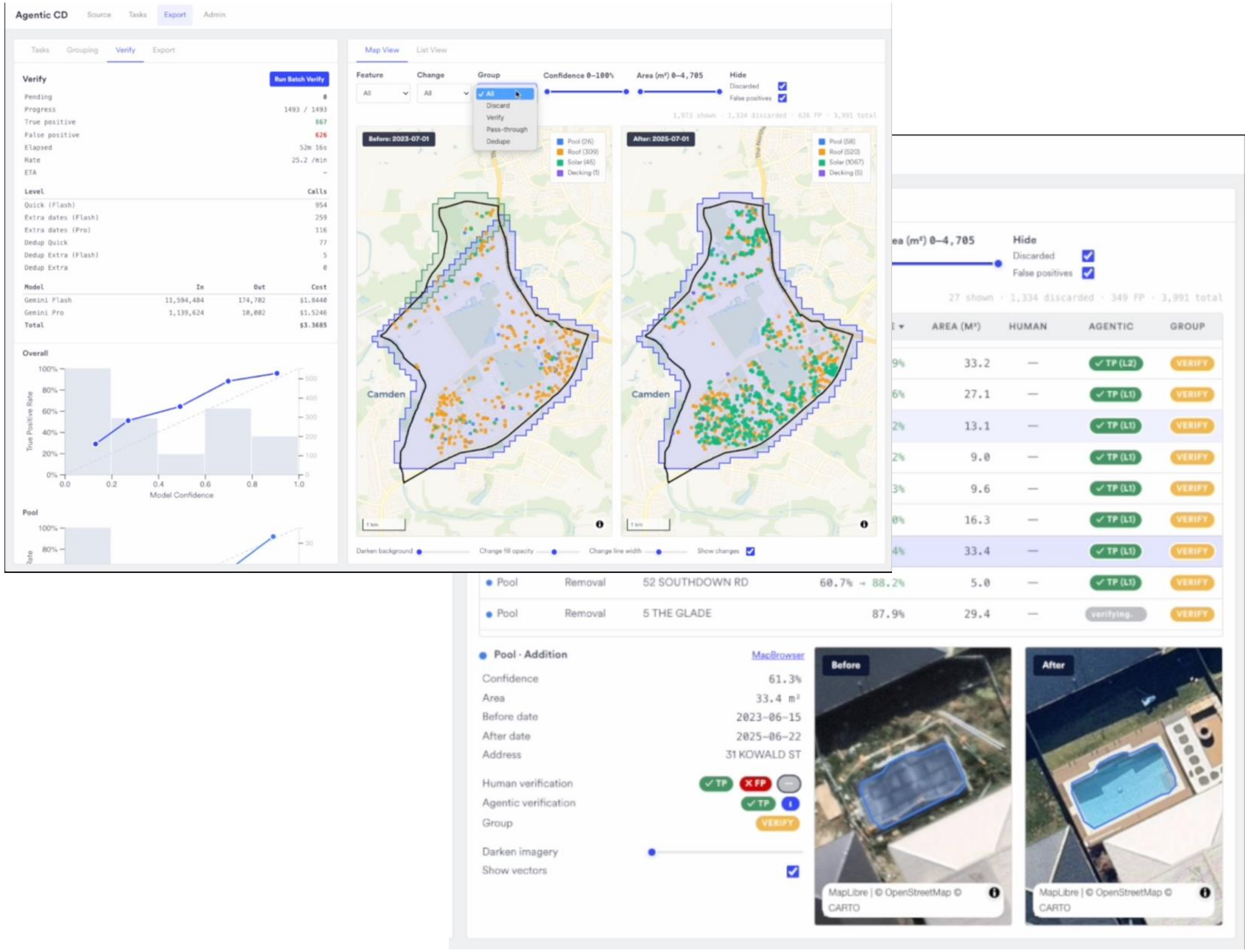
Nearmap Labs #7: Agentic Change Detection

What is it?

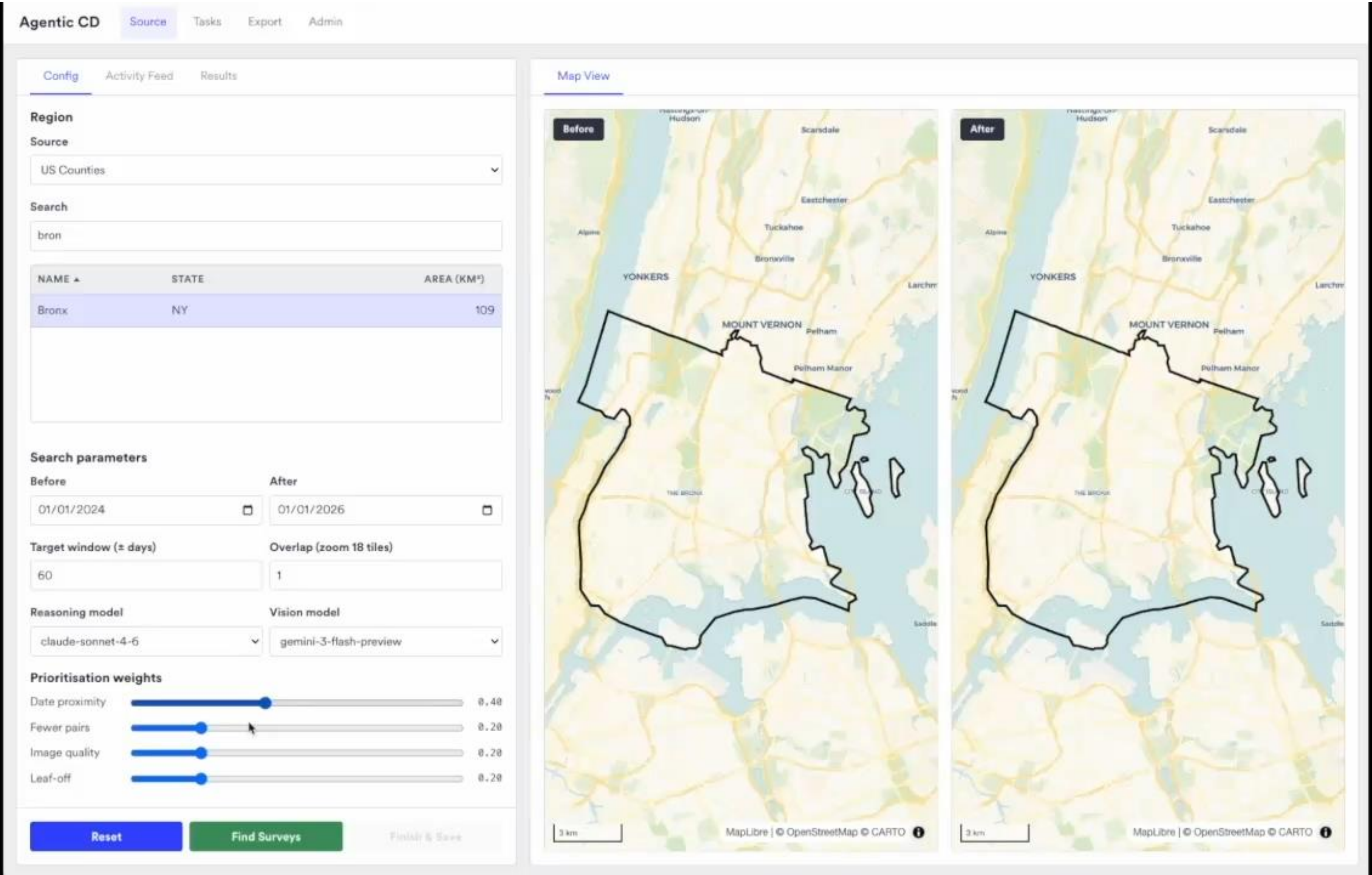
- Advanced Heuristic Change Detection (Gen 6 "intelligent" two date comparison)
- Temporal reasoning with a twist - Agentic Refinement

So What?

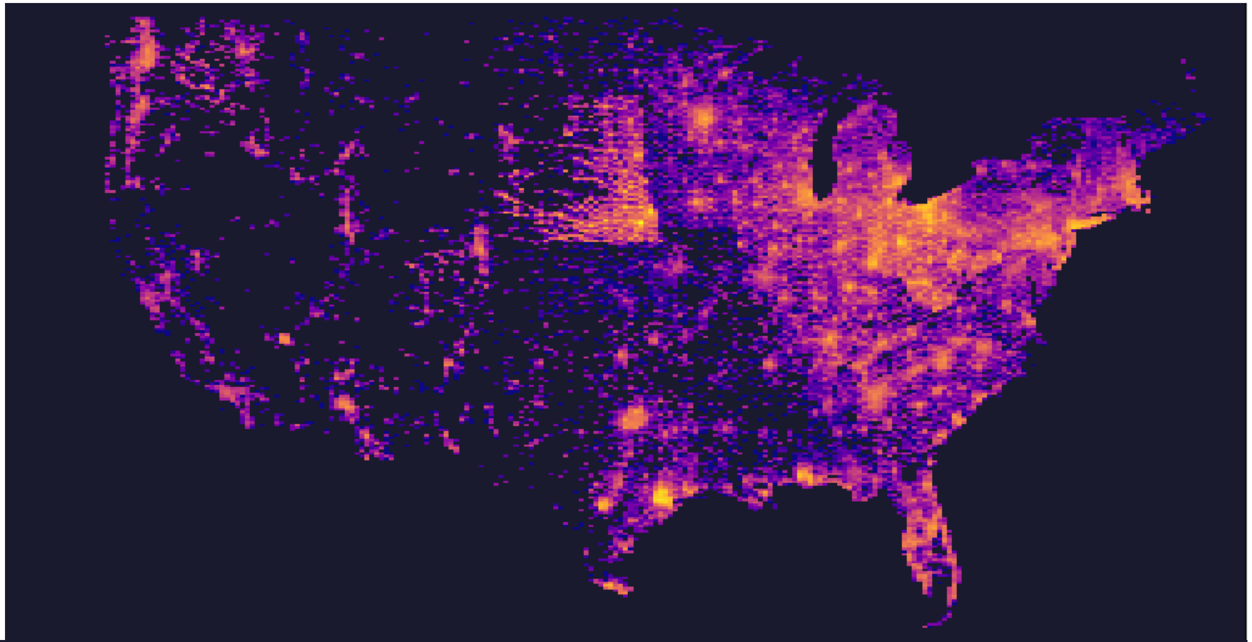
- Change Detection is Hard (bad signal-to-noise ratio, in a messy physical world)
- Scalable Success on a range of change tasks requires **extremely high grade imagery and supervised ML**
- Even that needs some **curation by an agent**… human is traditional, machine scales



Nearmap


Agentic CD
Source
Tasks
Export
Admin
Config
Activity Feed
Results
Region
Source
US Counties
Search
bron
NAME
STATE
AREA (KM²)
Bronx
NY
109
Search parameters
Before
01/01/2024
After
01/01/2026
Target window (± days)
60
Overlap (zoom 18 tiles)
1
Reasoning model
claude-sonnet-4-6
Vision model
gemini-3-flash-preview
Prioritisation weights
Date proximity
0.40
Fewer pairs
0.20
Image quality
0.20
Leaf-off
0.20
Reset
Find Surveys
Map View
Before
After
Scarsdale
Eastchester
Tuckahoe
Bronxville
YONKERS
MOUNT VERNON
Pelham
Pelham Manor
MapLibre | © OpenStreetMap © CARTO

Nearmap Research #1: Gen 6 Lift Analysis

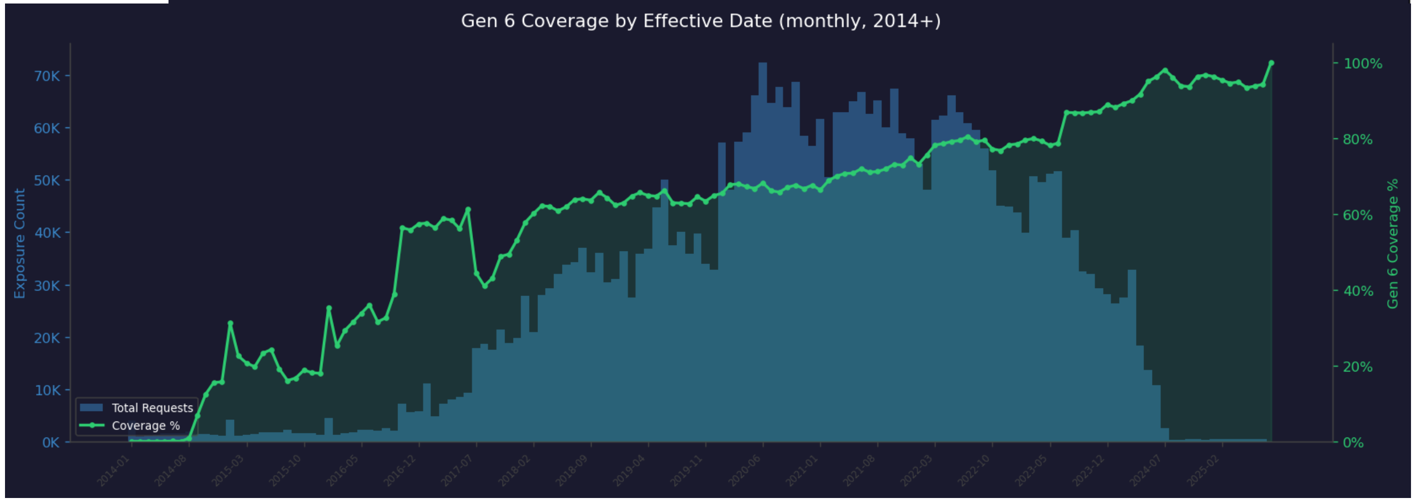


What is it?

- Analysis against an aggregated data set of claims data equivalent in size to a **top ten US insurer** (with claim types and loss value)
- **Comprehensive lift analysis** on existing Gen 6 Attributes and Scores

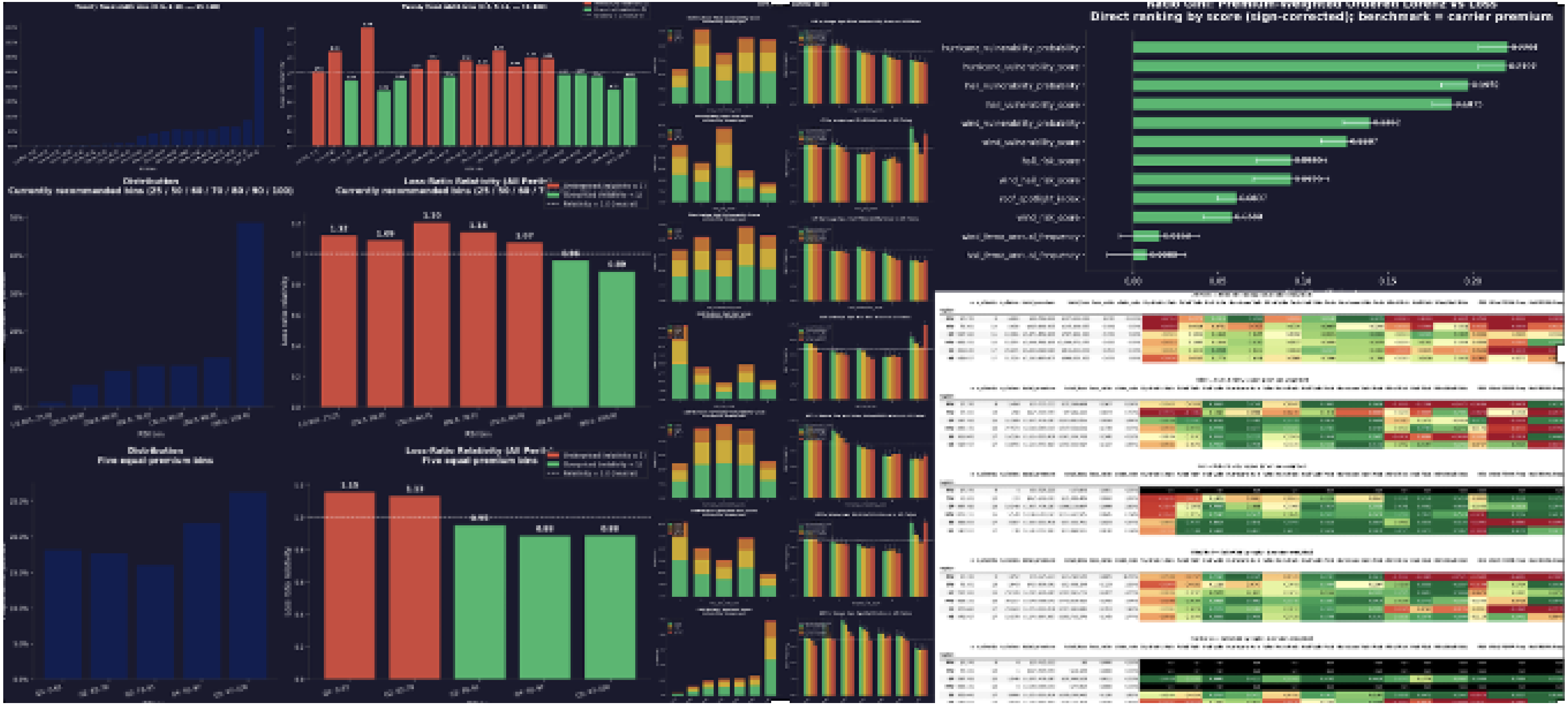
So What?

- Demonstrate value of existing attributes
- Guide complex data tests more effectively, and engineer better features
- Building a **next generation** of features and scores using Gen 6
- Already found **large lift opportunities**



Nearmap

“Top Ten” Equivalent Nearmap Lift Analysis



Nearmap

Nearmap Research #2: National Roof Age Study

What is it?

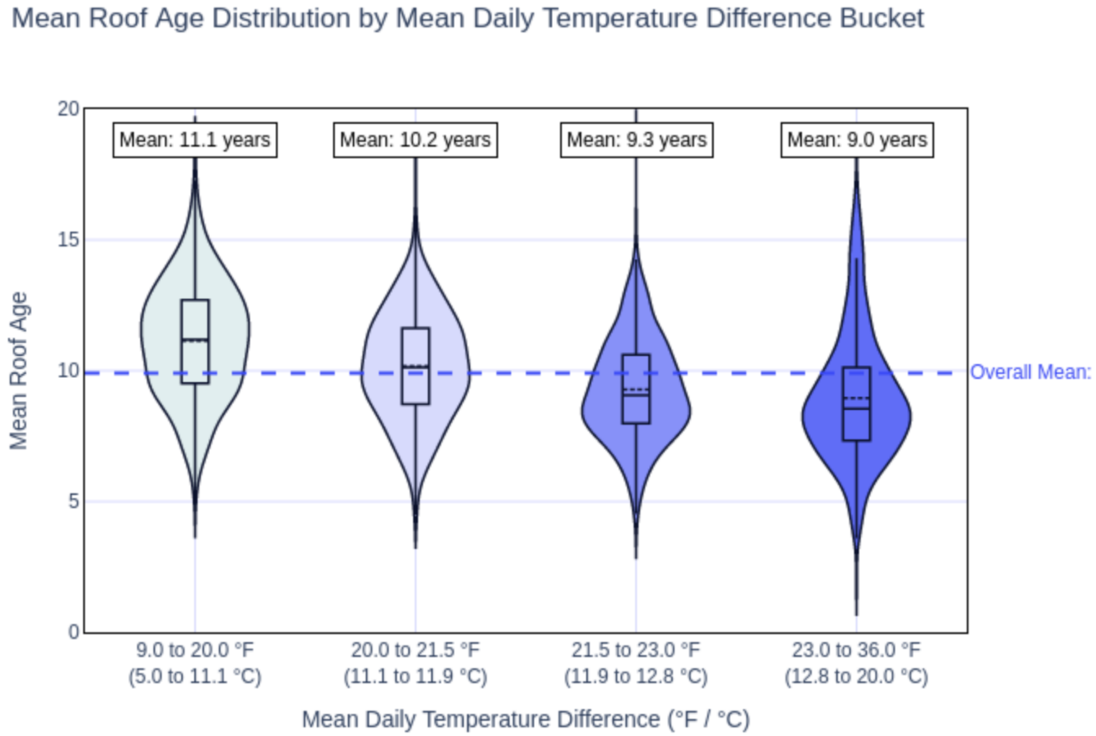
- Based on full, national US roof age analysis
- Aggregate statistics and trends
- Combined with climate data (temperature, humidity, precipitation)

So What?

- Only possible with consistent, **national pre-processing**
- **Benchmarking** and **Analysis**: See your properties vs the nation

Thermal Stress and Roof Age

One of the clearest trends in the data is the relationship between large daily temperature swings and younger roofs. Counties experiencing mean daily temperature fluctuations above 23°F (12.8°C) had an average roof age of 9 years, compared to 11.1 years in counties where fluctuations stayed below 20°F (11.1°C). This seemingly small difference in weather behaviour resulted in a roughly 23% increase in mean roof age.



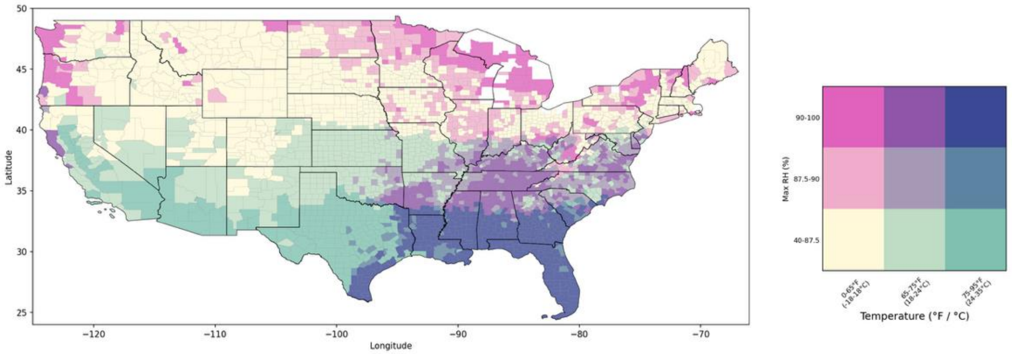
Nearmap

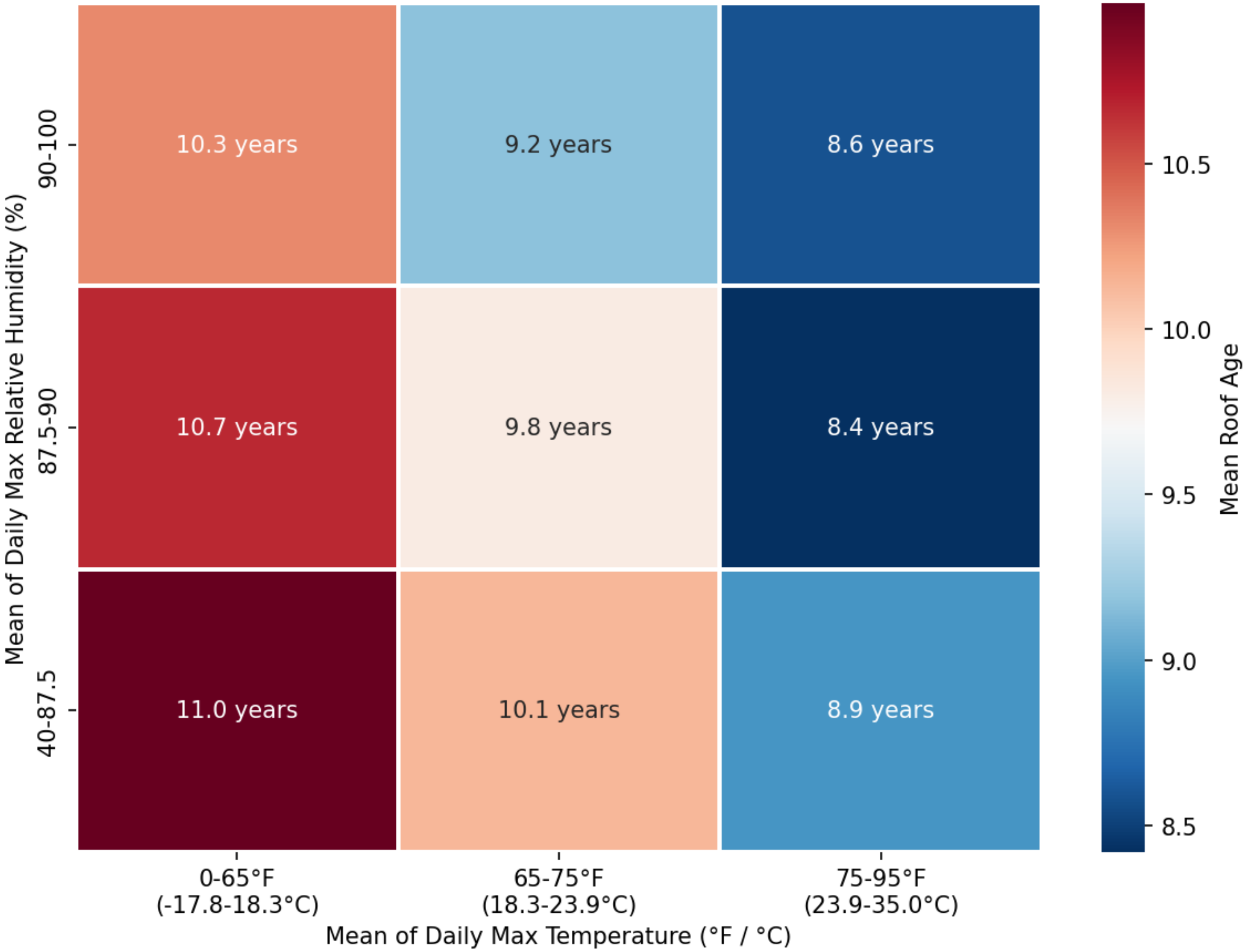
Weathering the Elements: How Climate Impacts Roof Lifespan in the US

Divide USA based on climate:

- Average **daily max temperature**
- Average **daily max humidity**

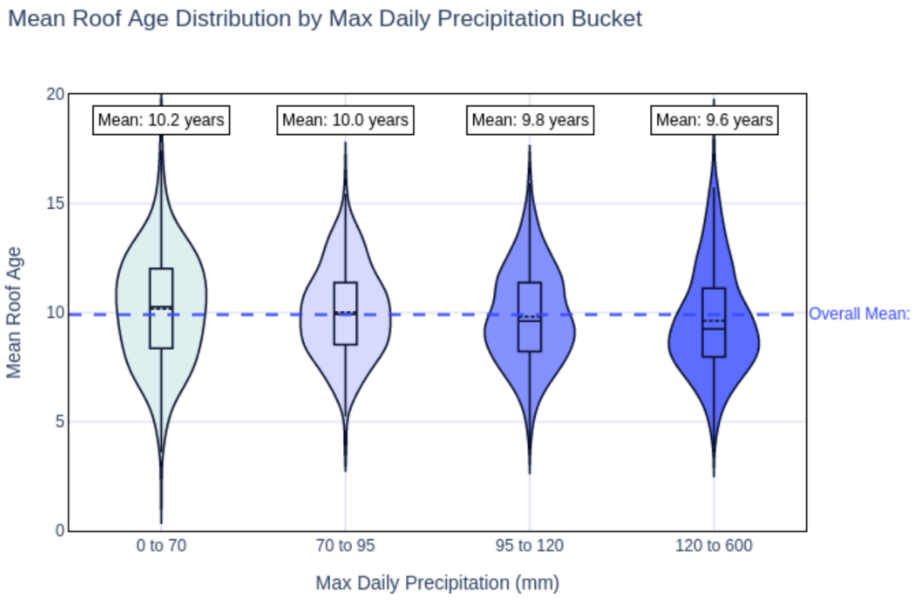
For each of the 9 groups:
calculate **average roof age**

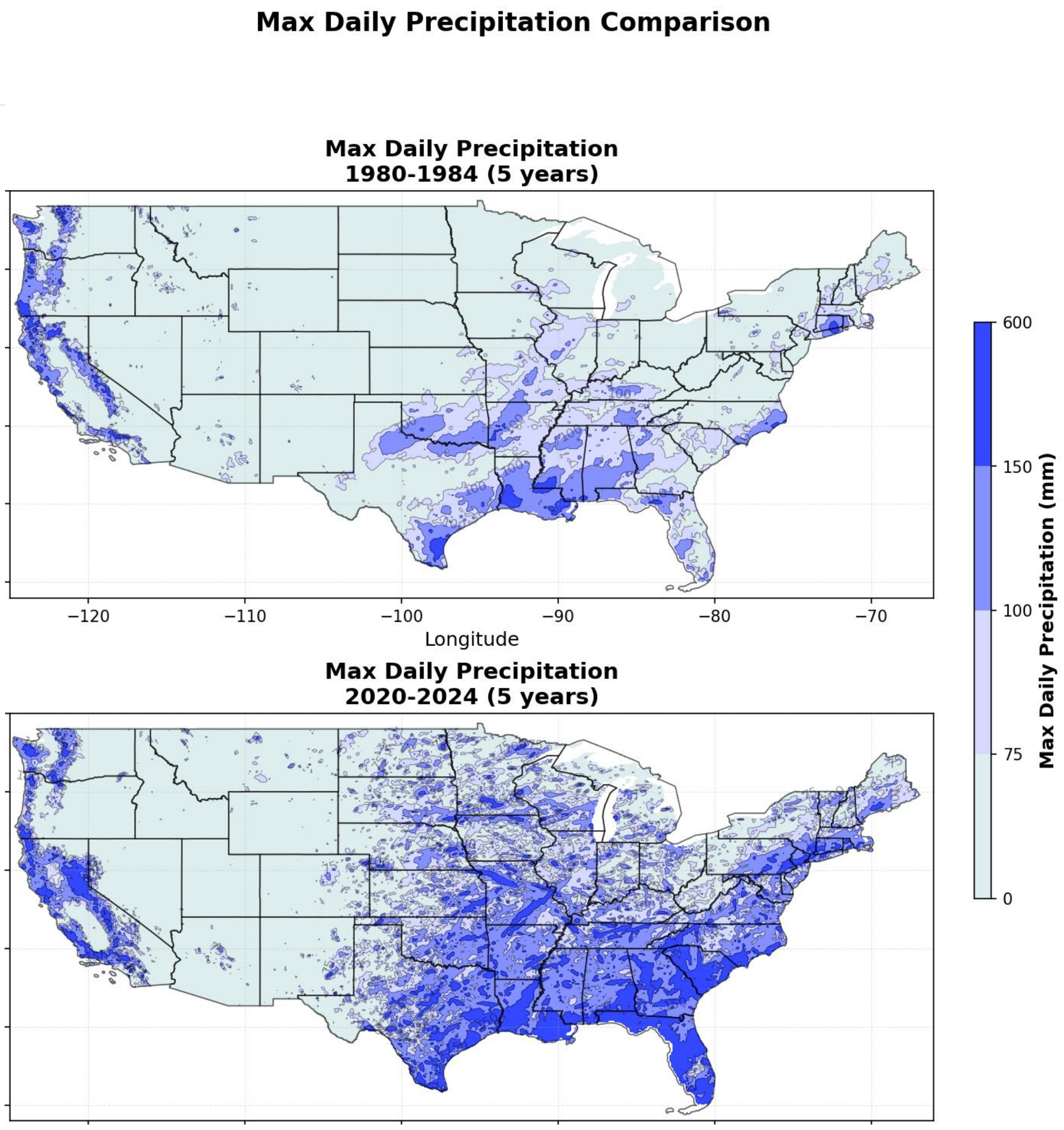




The Rising Toll of Heavy Rain

Extreme rainfall events, often associated with Severe Convective Storms (SCSs), are becoming more prevalent [5] and have significant effects on roof damage [6]. The data shows a clear relationship between maximum daily rainfall and roof age: counties experiencing the highest rainfall intensity tend to have younger roofs, indicating more frequent replacements. In regions where maximum daily rainfall exceeded 3 inches, mean roof ages were considerably lower than in areas with more moderate precipitation patterns.





Nearmap

Nearmap

Let's talk

nearmap.com/insurance

name@nearmap.com
+1(000) 000-0000
+1(000) 000-0000

10897 S River Front
Parkway, Suite 150
South Jordan, UT 84095