



# 5 ways to use generative AI in geospatial: A practical roadmap

CART

Hudson River

Weehawken



**Matthew Forrest**

Field CTO

CART ●

Hoboken

# The Five Steps to Geospatial generative AI

## Code Generation

Writing code to assist with human driven processes

## Feature Labeling

Making geospatial features easy to understand

## Retrieving Data

Getting data from human based text

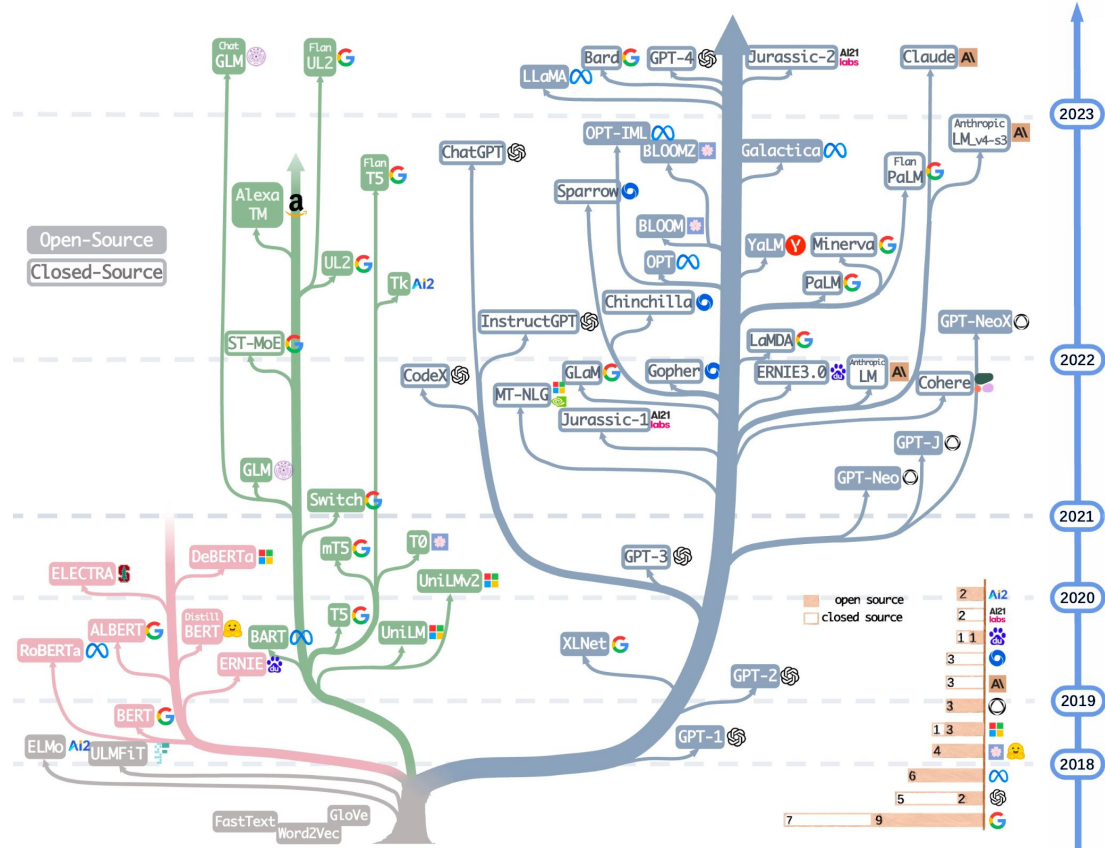
## Interaction

Use text to manipulate your data or application

## Building Custom Models

Building specific models for geospatial

# LLMs have come a long way



# Improving our code and writing it faster



**Let's see  
examples!**

# Making complex data easier to use



How do you  
generate 5,000  
different labels  
for each  
building/score?



**Tree Planting Priority: Very High**

This is a densely populated urban area, with notable youth and senior citizens. It lacks green spaces, has high buildings count, and increased Land Surface Temperature and Nitrogen Dioxide levels. Yet, it has a high Tree Planting Priority due to its exposure score.

# Automatic generation of labels for model results

Write your prompt ⓘ

↶ ↷ `{{ }}` Insert test input

Given a set of parameter for a particular location within the city in the format of

**[temperature, no2 concentration, population, vegetation coverage, distance to parks]**

with **values from 1 to 0**, and using the weights |

**W1 = 0.8, W2 = 0.5, W3 = 0.4, W4 = 0.2, and W5 = 0.4,**

you can calculate an index that explains the suitability of the specific place for planting trees. Please calculate the score index and **provide a brief description of the location**. Use following values: `[0.9,0.5,0.4,0.6,0.8]`

∨ Test your prompt ⓘ

INPUT	input	OUTPUT
1	<code>[0.9,0.5,0.4,0.6,0.8]</code>	The score index is 0.62.  The location is a place with a high temperature, a moderate level of NO2 concentration, a low population density, a moderate level of vegetation coverage, and a short distance to parks. It is a suitable place for planting trees.

+ Add test example

✦ Run

⊗ Text Bison ⚙ 0.7

☰ Text preview

# Air Quality, House Price, Crime rate, Amenities and School Quality

This location boasts excellent air quality, high house prices, and a low crime rate, making it suitable for those prioritizing safety and a healthy environment. However, it has limited amenities and poor school quality, which may deter families seeking educational opportunities and a variety of amenities nearby.



# In CARTO Workflows!

The screenshot displays the CARTO Workflows interface for a 'Generative AI' workflow. On the left, a 'Components' panel is open, showing a search bar and two available components: 'Generate Text (OpenAI)' and 'Generate Text (Google)'. The main workspace shows a workflow with two stages: 'Prompts' and 'Predictions'. The 'Prompts' stage contains a 'Custom SQL Select' component, and the 'Predictions' stage contains a 'PaLM2' component. A 'Run' button is visible in the top right. On the right side, a 'Parameters' panel is shown for the 'Generative AI Text' component, with settings for Temperature (0,2), Max output tokens (1024), Top-k (40), and Top-p (0,8). Below the workflow, a text output is visible, containing two paragraphs of generated text.

**CARTO Workflows / Generative AI**

Sources Components

Search component

Generative AI (2)

- Generate Text (OpenAI)
- Generate Text (Google)

Components

Prompts Predictions

Custom SQL Select PaLM2

Run

GENERATIVE AI

GENERATIVE AI TEXT

SUCCESS

Perform text generation with LLMs using the ML.GENERATE\_TEXT from BigQuery ML. You must have a column called 'prompt' that w...

Read more

Generative AI Text documentation

Parameters

Temperature: 0,2

Max output tokens: 1024

Top-k: 40

Top-p: 0,8

What will be the impact of climate change in London?

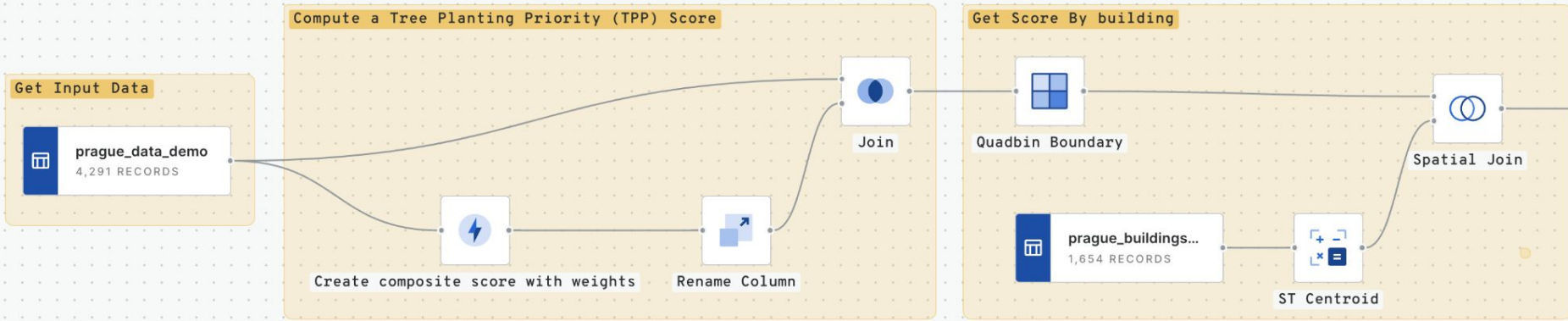
Climate change is expected to have a significant impact on London, with rising temperatures, more frequent extreme weather events, and changes in precipitation patterns. These changes will have a range of impacts on the city, including:

- Increased heat stress: Rising temperatures will make it more difficult for people to stay cool, especially those who are elderly, young, or have underlying health conditions. Heat stress can lead to heatstroke, dehydration, and other health problems.
- More extreme we

Who will win the next champions

"The next Champions League is set to be a thrilling affair, with some of the world's best teams competing for the title. Bayern Munich, Liverpool, Manchester City, and Paris Saint-Germain are all among the favorites to win the tournament, but there are a number of other teams that could also make a run.

One team to watch is Real



**Create Score Index**

**Bring it to buildings**

## Automatic Segmentation of Trees



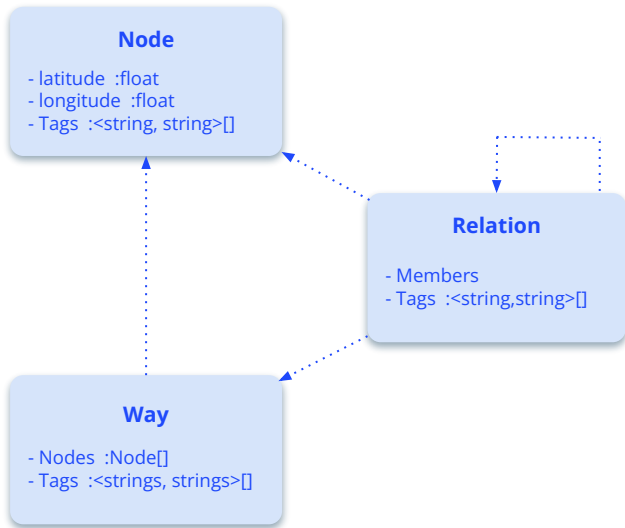
**Not just for vector data!**

**Using text to get  
our data for us**



# Writing SQL for OpenStreetMaps can be painful

All data already available in different platforms



Schema-less



92K tags for 3B objects



# Generation of complex Spatial SQL

Give me all healthcare  
facilities in South of Spain



If you want to find all point features with any value for the 'healthcare' tag in the southern part of Spain, here's the SQL query:

```
sql Copy code
SELECT
  feature_type,
  osm_id,
  osm_timestamp,
  geometry AS geom,
  (SELECT value FROM UNNEST(all_tags) WHERE key = '
FROM `bigquery-public-data.geo_openstreetmap.planet_f
WHERE feature_type = 'points'
AND EXISTS(SELECT 1 FROM UNNEST(all_tags) WHERE key =
AND ST_INTERSECTS(geometry, ST_GEOGFROMTEXT('POLYGON(
LIMIT 1000;
```



Map view



## Layers



A TILES

healthcare in South Spain



Q Search location



+

6

-

Portimão

Faro

Huelva

Sevil

Cádiz

Córdoba

ANDALÚSIA

Granada

GIBRALTAR

Málaga

Almería

Alicante

Murcia

Cartagena

© CARTO, © OpenStreetMap contributors

A SQL Editor - Edited

carto\_dw



```
1 SELECT carto.us.QUERY_OSM('Give me healthcare facilities in South of Spain')
```

SOURCES (1)



Query - BigQuery

SQL Query 2

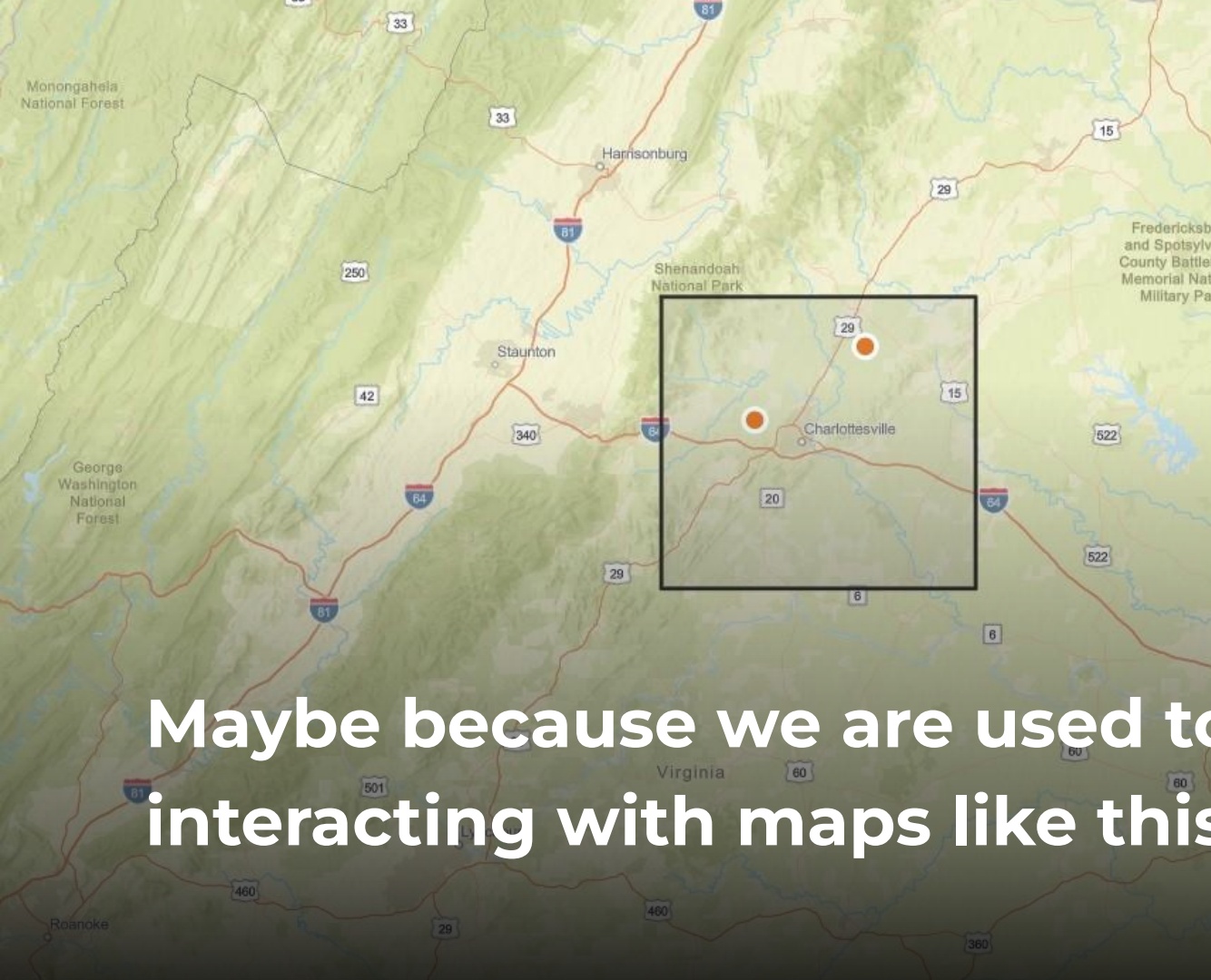


+ Add source from...

Run

00:05

Create table from query



Maybe because we are used to interacting with maps like this

Draw Geometry

Ask Question

Fire Departments

Run Clear

```
{  
  "type": "FeatureCollection",  
  "features": [  
    {  
      "type": "Feature",  
      "geometry": {  
        "type": "Point",  
        "coordinates": [  
          -78.56765747070312,  
          38.066207590396916  
        ]  
      },  
      "properties": {  
        "name": "Harrisonburg Fire Department"  
      }  
    },  
    {  
      "type": "Feature",  
      "geometry": {  
        "type": "Point",  
        "coordinates": [  
          -78.56765747070312,  
          38.066207590396916  
        ]  
      },  
      "properties": {  
        "name": "Harrisonburg Fire Department"  
      }  
    }  
  ]  
}
```

Add to map



# Your personalized expert suggesting datasets



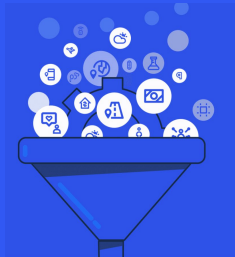
Bard is still in its experimental phase. Chatting with it and rating its responses will help improve the experience.

Enter a prompt here



# Finding data relevant to your analysis

## Fine-tuning the model with CARTO Data Observatory knowledge



12,149 datasets  
120 countries  
40 providers  
8 categories

CARTO

**Housing / Unica360**  
**Cadastral - Spain (Postal code)**

Summary Data Map

Explore

One unique dataset on parcels and buildings, based on official data from all of the 5 different Cadaster databases in Spain, at the 100x100 metre grid aggregation level. It provides data on the total of buildings, properties, properties by use (premises, dwellings, offices), by year of construction, surface area, garden, heights, etc. It allows managing the enormous volume of data of the 5 Spanish Cadastre sources, in an agile and effective way.

**Key variables** (show all)

- Number of portals
- Number of buildings
- Number of properties
- Parcel surface
- Property surface by use
- Property surface
- Number of premises by construction quality
- Number of single family properties

**License**  
Premium data

**Country**  
Spain

**Source**  
Unica360

**Platform**  
Postal Code

**Temporal aggregation**  
Yearly

**Update frequency**  
-

**Associated Geography**  
Postal Code - Spain

**ChatGPT**

I am trying to find how the residence and the daily mobility patterns of my customers affect the sales of my restaurants. Do you have any variables to do so in Barcelona?

**CARTO (GPT4 + Data Observatory)**

I am trying to find how the residence and the daily mobility patterns of my customers affect the sales of my restaurants. Do you have any variables to do so in Barcelona?

Based on your query, I suggest the following top 4 datasets that could be useful in enriching your data with real estate data for Madrid:<sup>1</sup>. The paper analyzes the effect of closing bars and restaurants on mobility patterns in Cataluña. It might be useful for you to understand how mobility patterns affect sales of restaurants.

- **Unica360 Housing Cadaster for Spain - Postal Code Level:** This dataset provides information on parcels and buildings at the postal code level. It includes data on the number of residential properties, floor area of residential and commercial properties, and other property types, as well as other variables such as year of construction, surface area, garden, and heights.
- **Unica360 Housing Cadaster for Spain - Parcel Level:** This dataset provides socio-demographic and socio-economic data at the cadastral parcel level in Spain. It includes information on various aspects such as profits from real estate capital, floor area of residential and commercial properties, number of properties of different sizes, and data on resident population by gender/age, income, education level, and expenses, among others.
- **Unica360 Housing Cadaster for Spain - Grid 100m Level:** This dataset provides data on parcels and buildings at the 100x100 metre grid aggregation level in Spain. It includes data on the number of residential properties, floor area of residential and commercial properties, and other property types, as well as other variables such as floor area, number of properties under 50 square meters, and number of properties of different sizes.
- **The Data Appeal Company Places Sentiment Main Listing:** This dataset could be useful in enriching your data with information on points of interest (POIs) in Madrid. It provides location, contact, and category information for a wide range of POIs, as well as additional KPIs such as customer feedback, sentiment, and footfall. The dataset includes data from various online platforms such as Booking.com, Airbnb, TripAdvisor, Google, Facebook, etc.

Send to ChatGPT

Send to CARTO

Send to all ...

Send

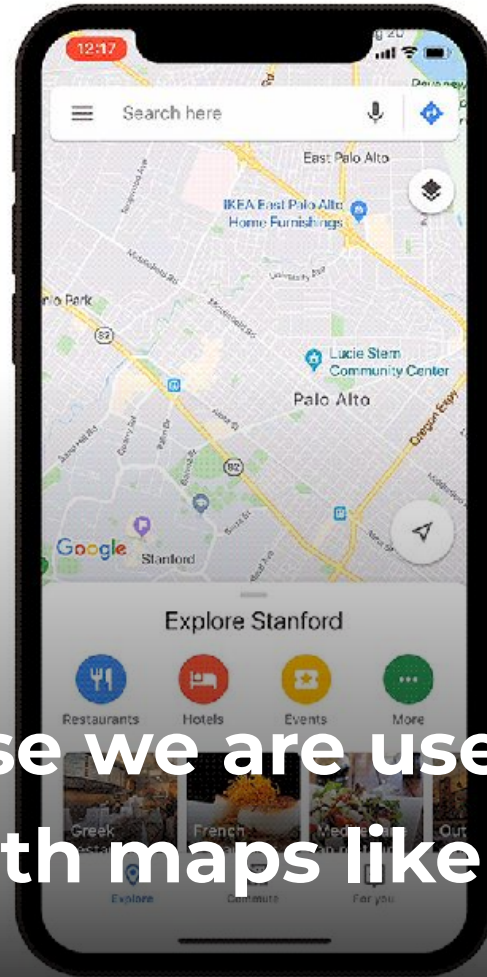
# Interacting with our maps



**AI is the  
new UI**

**And SQL remains the  
lingua franca for  
Analytics, now on a  
new AI world of LLMs**





Maybe because we are used to interacting with maps like this

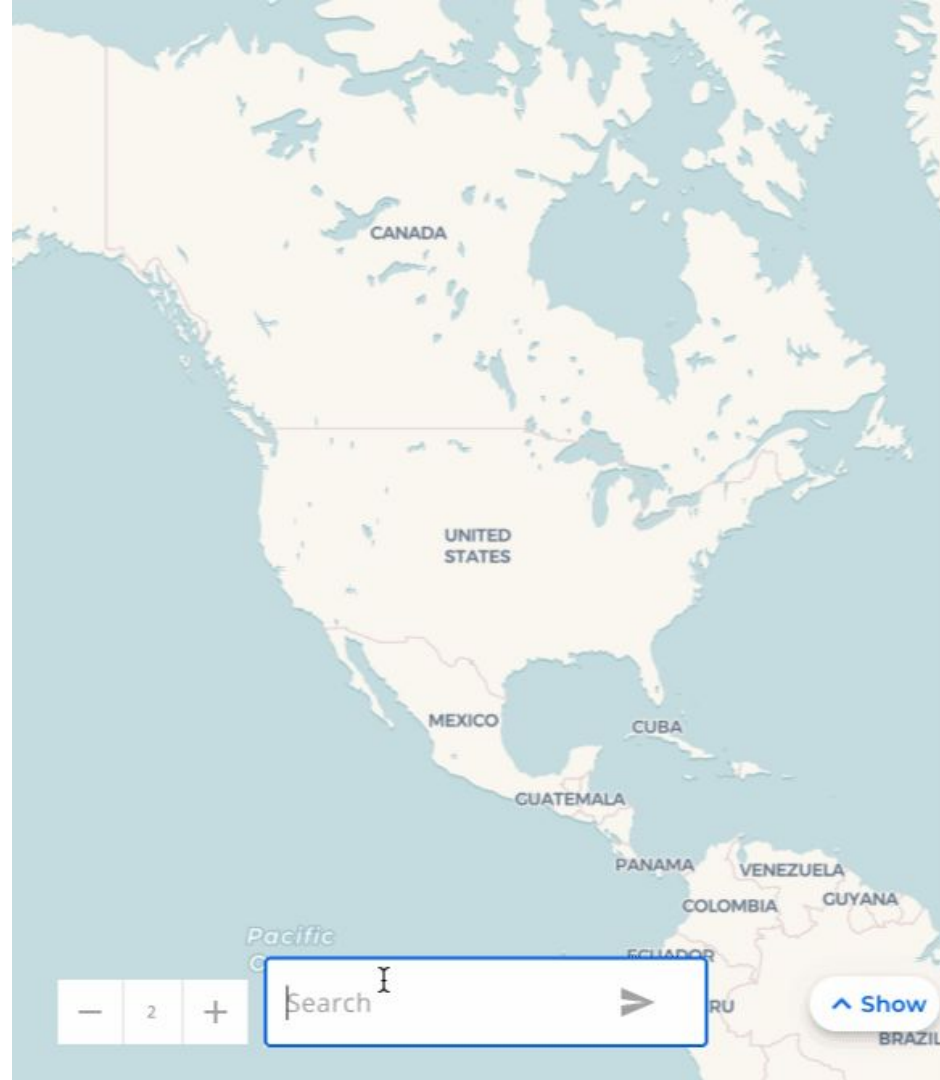
 Prompt Engineering

 Embedding CARTO  
DO

 Spatial **SQL** in LLMs

 CARTO Maps API

 **MapGPT!**





[Home](#) [About](#) [Services](#) [Blogs](#)

[Contact](#)

[BOOK A  
DEMO](#)

[LOGIN /  
SIGNUP](#)

# Revolutionising Carbon Management With AI

[JOIN WAITLIST](#)

[LEARN MORE](#)

AI-driven, cloud-based carbon footprint management platform designed for organizations

## Connecting text based datasets to geospatial data

Carbon GPT

Login to Your Account

Enter your username & password to login

Your Email



So where is  
this going?



# Custom models for geospatial

**First, what are  
these models  
good for?**

Salesforce/xgen-7b-8k-base like 297

Text Generation Transformers PyTorch llama text-generation-inference License: apache-2.0

Model card Files and versions Community 28 Train Deploy Use in Transformers

Edit model card

Downloads last month  
7,435



Hosted inference API

Text Generation

Examples

My name is Teven and I am



Compute

\*+Enter

0.0

This model can be loaded on the Inference API on-demand.

ON Output

Maximize

Spaces using Salesforce/xgen-7b-8k-base 9

## XGen-7B-8K-Base

Official research release for the family of XGen models (7B) by Salesforce AI Research:

Title: [Long Sequence Modeling with XGen: A 7B LLM Trained on 8K Input Sequence Length](#)

Authors: [Erik Nijkamp\\*](#), [Tian Xie\\*](#), [Hiroaki Hayashi\\*](#), [Bo Pang\\*](#), [Congying Xia\\*](#), [Chen Xing](#), [Jesse Vig](#), [Semih Yavuz](#), [Philippe Laban](#), [Ben Krause](#), [Senthil Purushwalkam](#), [Tong Niu](#), [Wojciech Kryscinski](#), [Lidiya Murakhovs'ka](#), [Prafulla Kumar Choubey](#), [Alex Fabbri](#), [Ye Li](#), [Rui Meng](#), [Lifu Tu](#), [Meghana Bhat](#), [Chien-Sheng Wu](#), [Silvio Savarese](#), [Yingbo Gao](#), [Han Hyohan](#), [Ziwei Tang](#), [Oleg](#)

(\* indicates equal contribution)

Correspondence to: [Shafiq Rayhan Joty](#), [Caiming Xiong](#)

Models

First, we need to understand text generation and model size

# Improving Machine Learning from Human Feedback

Erin Mikail Staples + Nikolai Lubimov  
PyData DE 2023







## IBM NASA Geospatial

Community

### Research interests

Geospatial foundation models using HLS2 data

### Team members 38

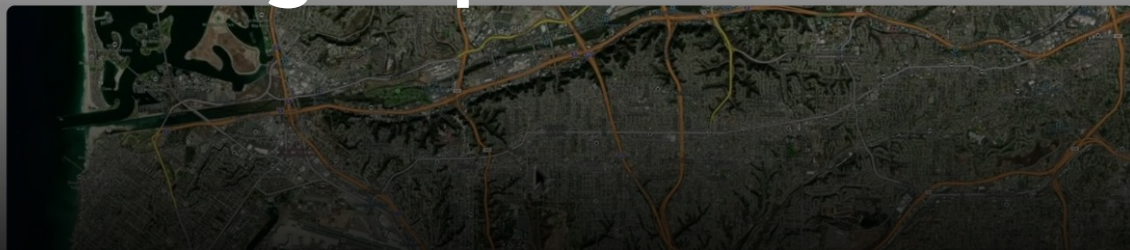


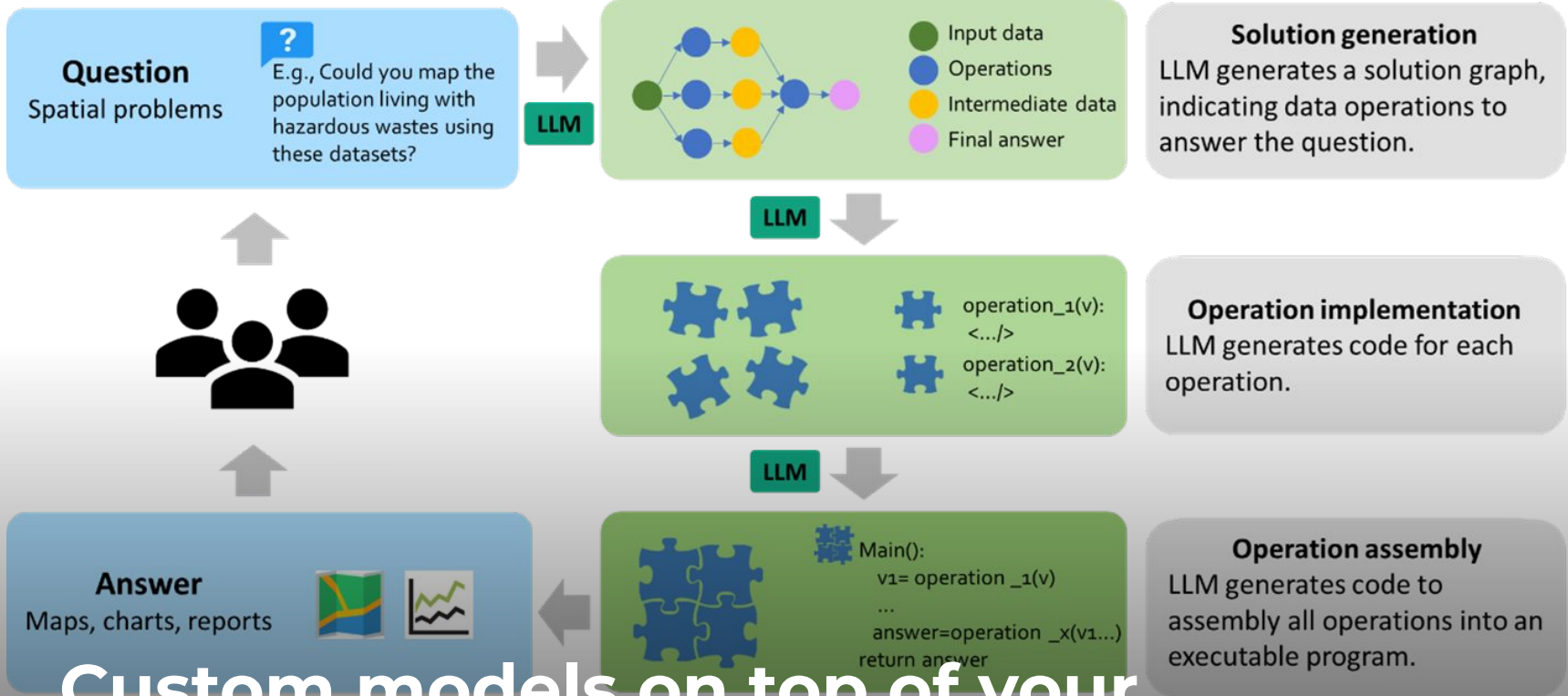
### Organization Card

About org cards

NASA and IBM have teamed up to create an AI Foundation Model for Earth Observations, using large-scale satellite and remote sensing data, including the Harmonized Landsat and Sentinel-2 ([HLS](#)) data. By embracing the principles of open AI and open science, both organizations are actively contributing to the global mission of promoting knowledge sharing and accelerating innovations in addressing critical environmental challenges. With Hugging Face's platform, they simplify geospatial model training and deployment, making it accessible for open science users, startups, and enterprises on multi-cloud AI platforms like [watsonx](#). Additionally, Hugging Face enables easy sharing of the pipelines of the model family, which our team calls [Prithvi](#), within the community, fostering global collaboration and engagement.

# Models for specific geospatial use cases





**Custom models on top of your data**