<epam>

Applications of Generative AI for GIS data

October 2023

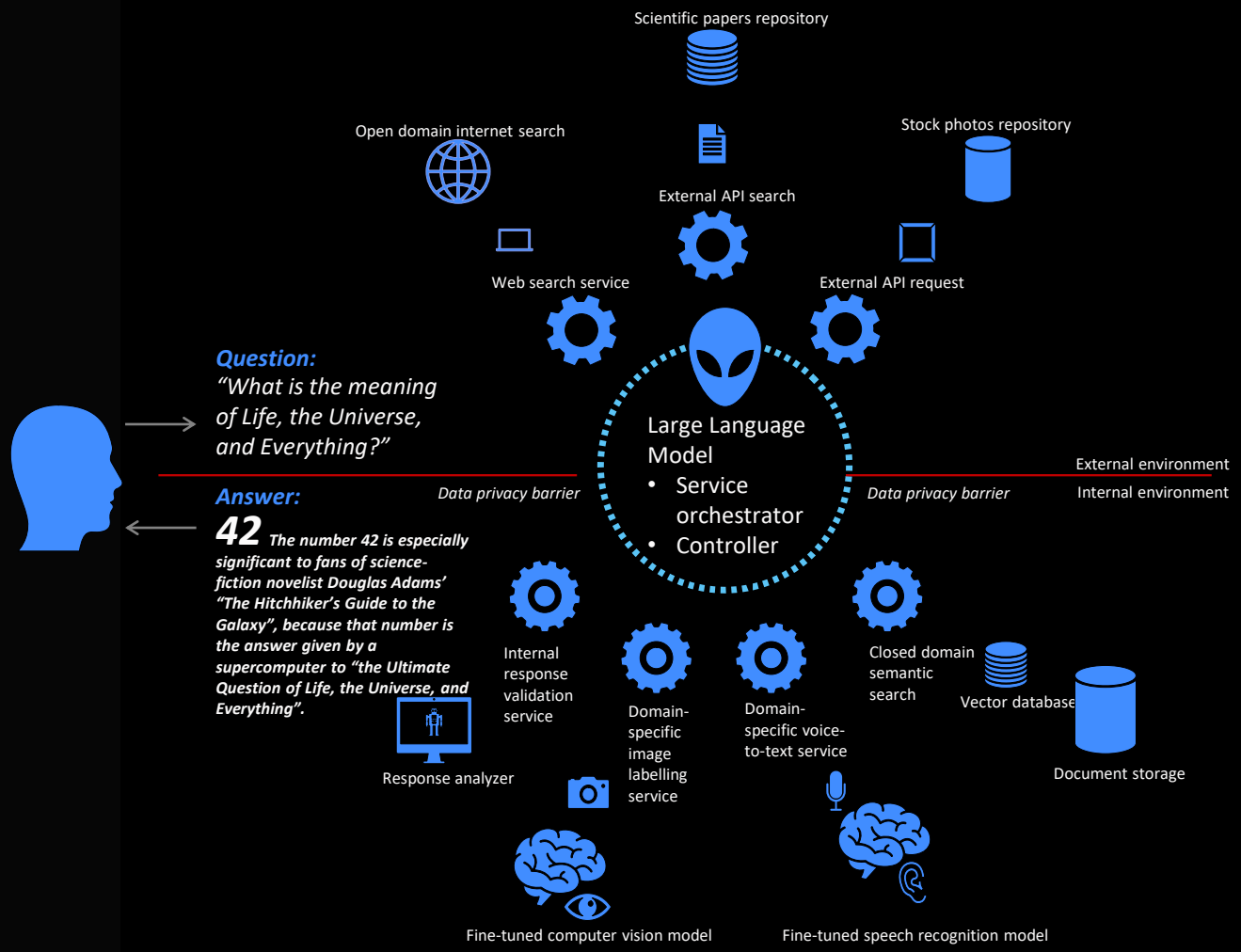
GenAI assisted analytics for GIS data

| Hub and Spokes | Hybrid Architecture

- GenAI enables new approach to Geographic Information System
- Capability to define an arbitrary area is not difficult, but analysing it, is content is not trivial
- Generative AI gives that capability, and is capable of acting as a digital assistant and can perform numerous role of a service orchestrators or controller
- Our current solution is capable of simultaneously connecting to:
  - → Data stored internally in databases & repositories (collections)
  - → Generic open domain search (Ask PaLM 2 for text)
  - → Specific external API request (Google Maps)

AI Solution Strategies, Key to Success

- Central GenAI role can be expanded in the future by increasing number of connection functionalities, for instance to perform graphic search, or voice-to-text conversion
- Subsequent relevant external API searches can be implemented
- Response quality can be ensured by enhancing the system with an internal response validation system
- There is a clear data privacy boundary dividing the internal and external systems introducing the security by design



Built upon the experience from an existing PoC

- Built on VertexAI (for quick access of the selected dataset) and PaLM 2 for text (for all other queries)
- Integrated with CARTO for React for the front-end
- Allows GIS data search results visualization

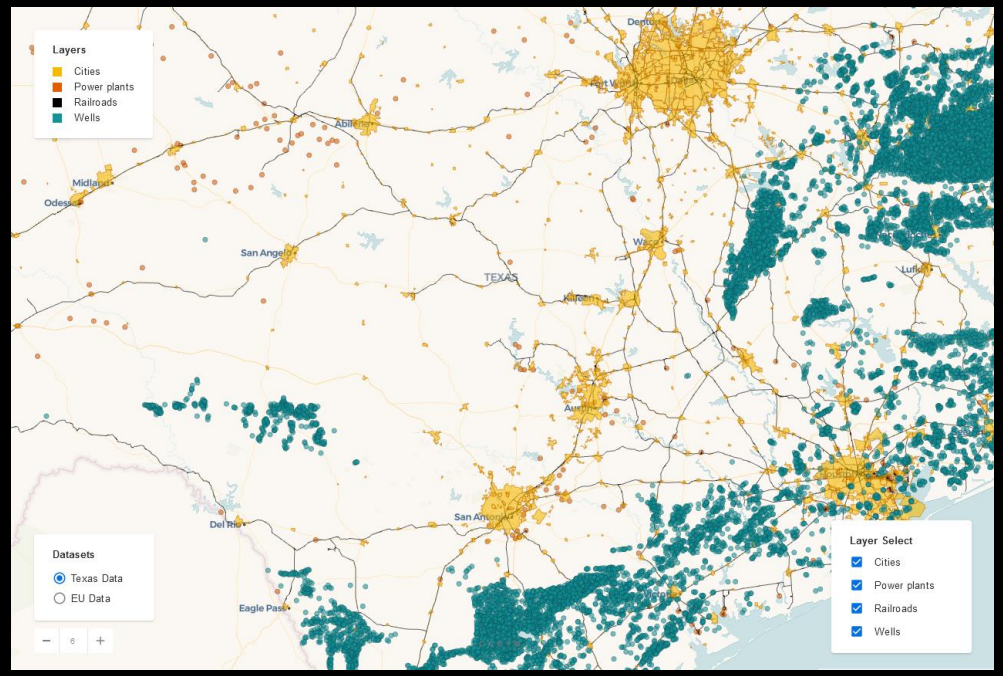
Features:

- Filtering of search results, updating map on the fly
- Define an area for search on the map as custom polygon
- Retrieves information from mouseover at the point of interest

Reacts to a human-like requests:

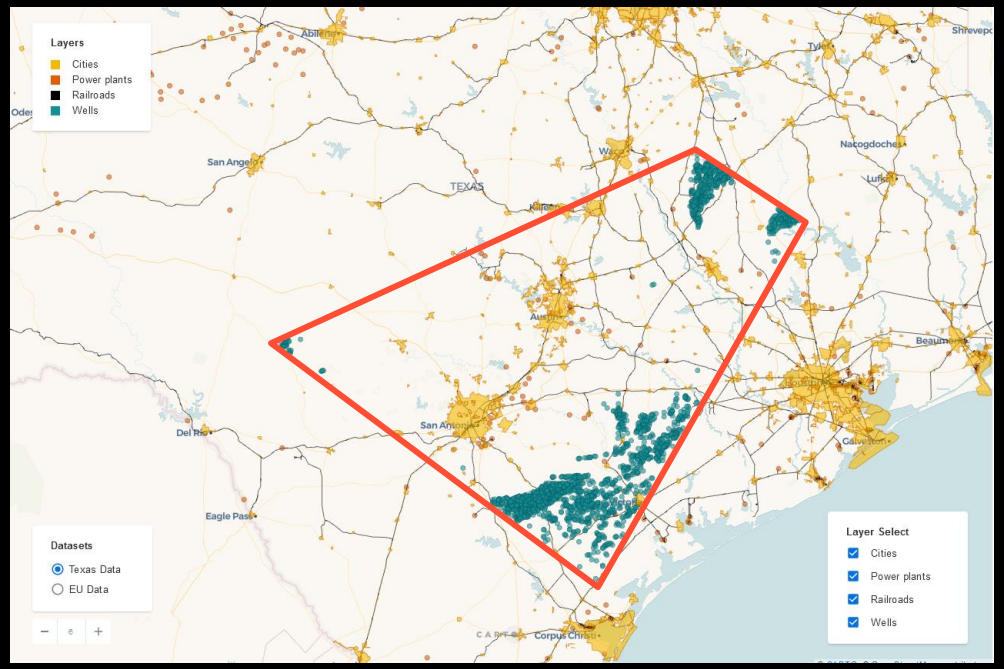
  - Show all the objects from specific administrative area
  - Show all objects of given type
  - Find object with minimum/maximum property
  - Show top N objects by property
  - What is the average property value of all objects?

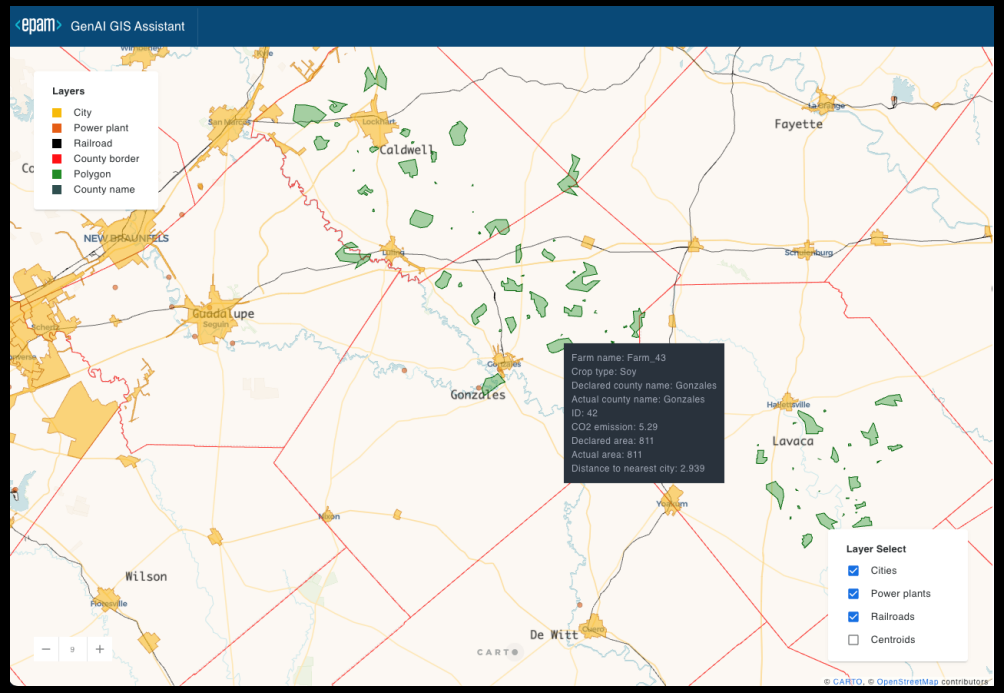
- Define an area for search on the map as custom polygon
- Multiple data and visualizing layers:
  - Cities and counties (administrative boundaries)
  - Power plants
  - Railroads
  - Other points of interest



From point to polygon

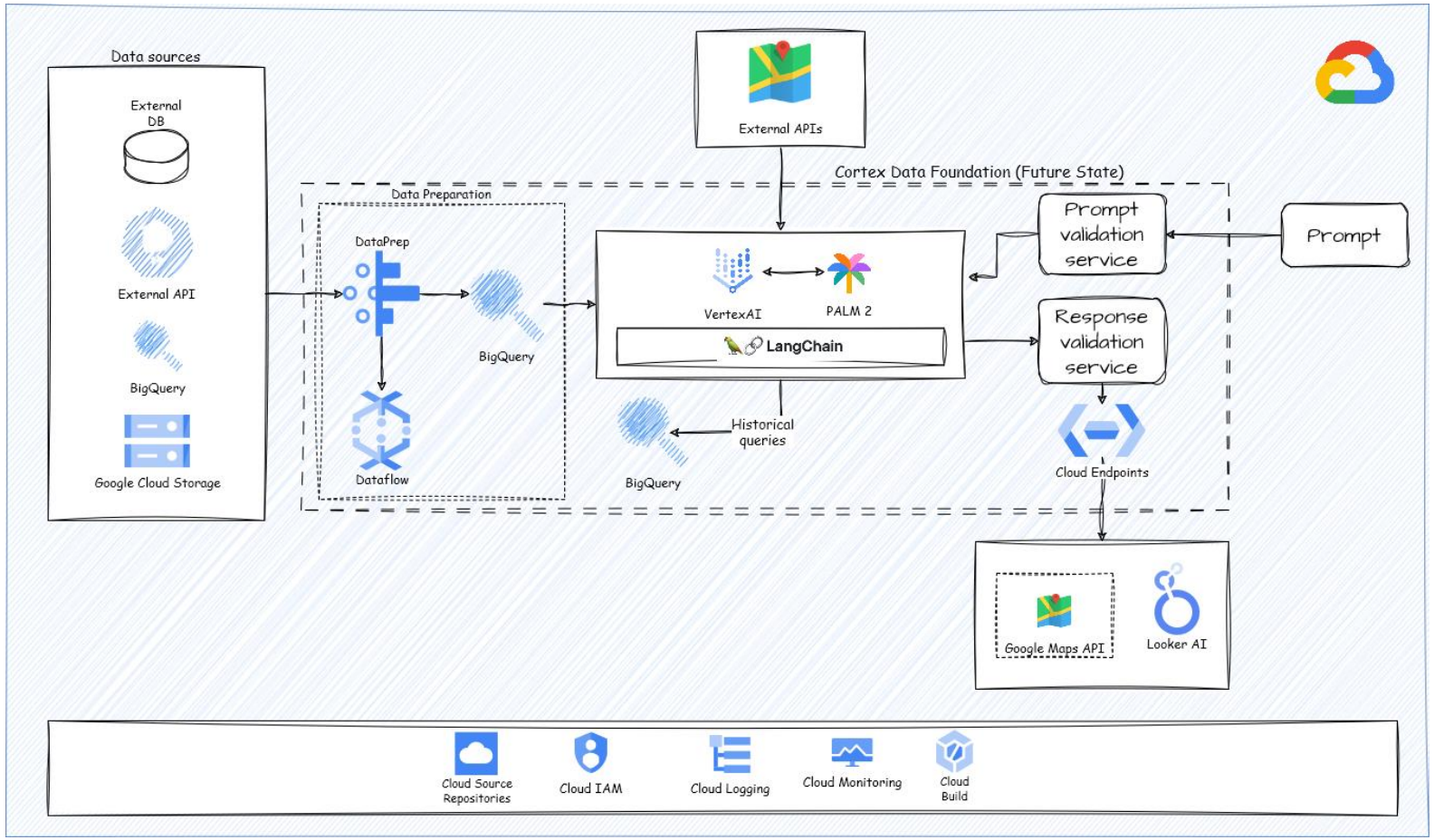
- The next phase of development was focused on evolving from a singular point of interest to arbitrary areas defined by geo coordinates
- This allows to perform search of the single PoI's (points of interest) inside the polygon, or use them as objects themselves
- This enabled search based on the polygon's properties, such as:
  - Show me all polygons with area greater than
  - Show me three nearest polygon to a Town
  - Show me all polygons of the attribute value equal…





Proposed architecture

- Utilises the entire codebase and all services from the Google offering
  - BigQuery
  - Google Cloud Storage
  - Google Cloud Dataflow & Dataprep
  - Vertex AI & PaLM 2
  - Looker
- Connecting to external data sources and APIs (such as Google Maps)



Future uses

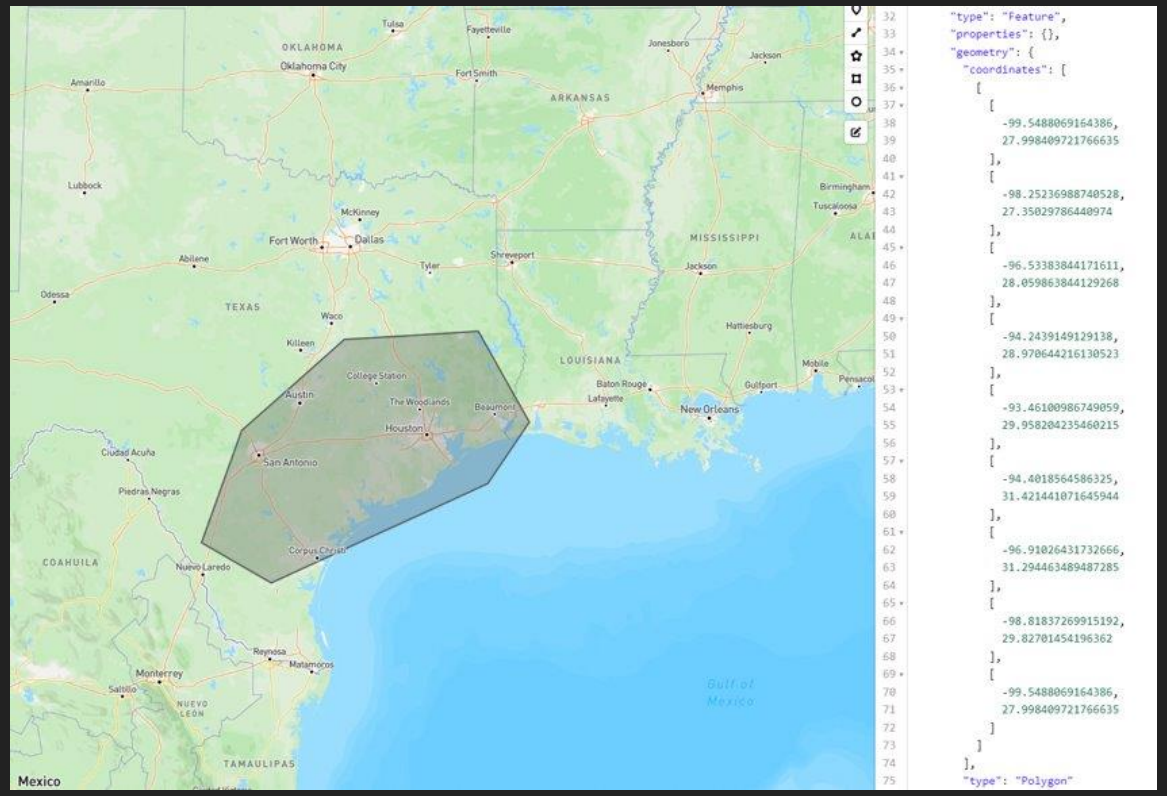
GPT-X technology enables tackling multiple angles of the problem and giving a neat, presentable solution:

- Translating various languages to create uniform definitions and allow for easier auditing
- Extracting the relevant information, even if placed in diverse places, styles, or formats
- Identification and extraction of specific entities, such as names, organizations, locations, or key terms creating unified input for further processing
- Comparing multiple filings and versions
- Reading definitions of land area boundaries (such as rivers, mountain ridges, roads, etc.) and transforming it into computer-processable entries

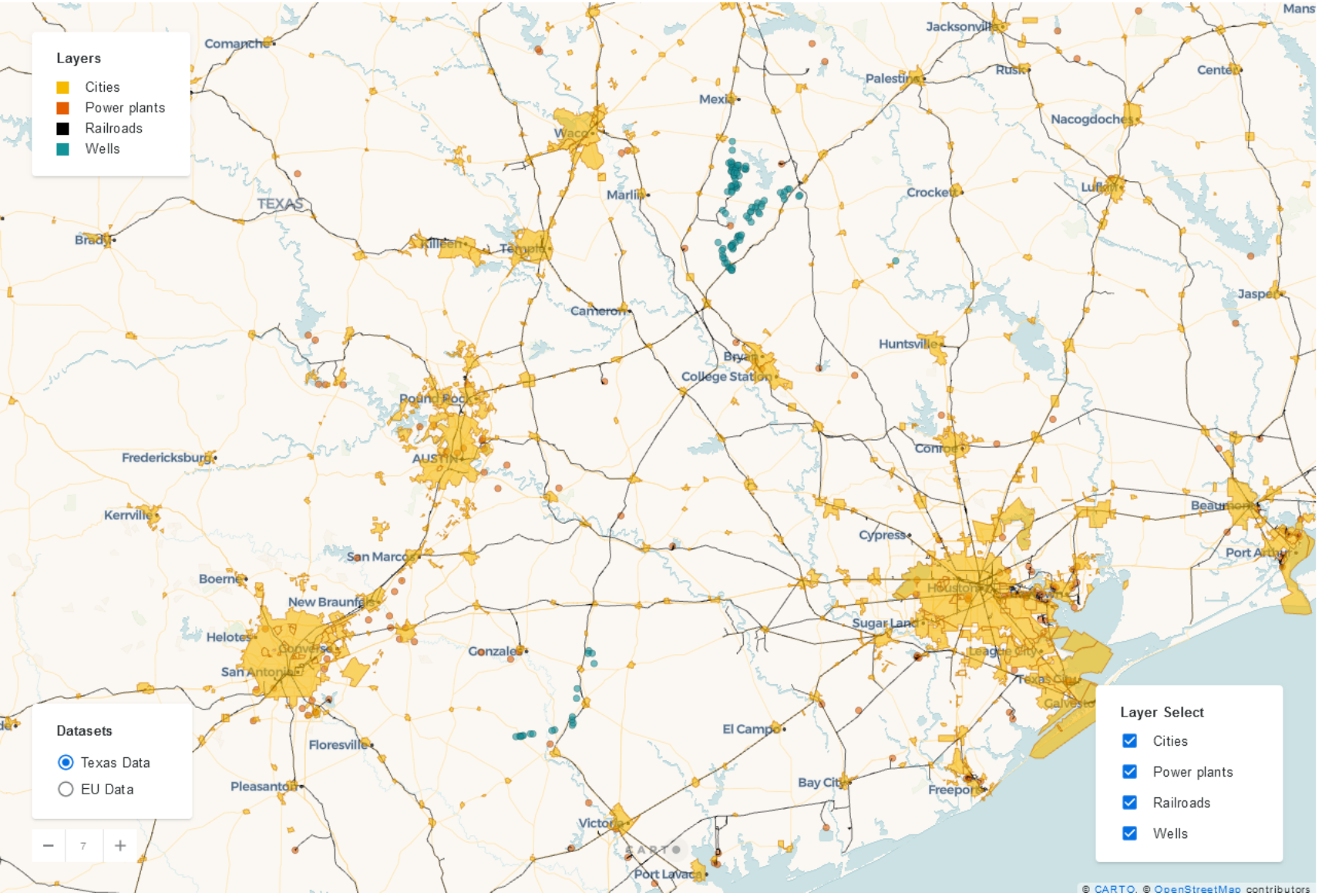
From there we can overlay it with satellite or aerial photos and perform series of checks, such as:

- General geography (middle of the ocean)
- Presence of mountains, national parks, airfields, military exclusion zones, etc.
- Presence of heavy industry, motorways, and railroads, etc.

This gives a universal tool allowing to work with complex challenges such as deforestation, emissions, city planning, population migrations, etc.



Example application for search and analysis of the geothermal wells for renewable energy production



Show all the wells inside the polygon with the following GPS coordinates: -98.4301769, 32.5506630 -100.2539074, 30.0552783 -97.3754894, 28.1944678 -95.5517589, 31.0199553 -98.4301769, 32.5506630

5821 wells have been found and shown on the map

Show all the wells with temperature more than 100 degrees at depth level of 2500 meters

5429 wells have been found and shown on the map

Show all the wells with thermal coefficient more than 2 and heat flow more than 75

236 wells have been found and shown on the map

Summarize key numeric features of the wells

The mean values of key numeric features of the wells are: Mean Depth: 12513.69 feet Mean Crude Oil Production/day: 3.48 barrels Mean Condensate Oil Production/day: 0.50 barrels Mean Tubinghead Gas Production/day: 26.69 MCF Mean Casinghead Gas Production/day: 10.62 MCF Mean Heat Flow: 97.63 BTU/hr-ft Mean Thermal Coefficient: 2.10 BTU/hr-ft-°F Mean Temperature at 1000m: 75°F Mean Temperature at 1500m: 76.06°F Mean Temperature at 2000m: 100°F Mean Temperature at 2500m: 125°F Mean Temperature at 3000m: 125°F

+ New request | Prompts | Use PaLM only | Send

Data used in the PoC phase

1) Wellbore Query Data

https://mft.rrc.texas.gov/link/650649b7-e019-4d77-a8e0-d118d6455381

This data set provides data related to oil and gas well bores used in the RRC Online Oil & Gas Data Well Bore Query. The Well Bore Query data set provides a way for users to search for wells by district, lease, county, field, operator, drilling permit number, API number, on/off schedule, or well type. The file is run at the beginning of each month.

2) Well Layers by county

https://mft.rrc.texas.gov/link/d551fb20-442e-4b67-84fa-ac3f23ecabb4

Contains latitude & longitude for each well from Wellbore Query dataset

3) Production data

https://mft.rrc.texas.gov/link/1f5ddb8d-329a-4459-b7f8-177b4f5ee60d

This data set contains production rates from 1993 to current. Note: for demo purposes only first 4 months of 2023 were kept

4) Base Layers by county

https://mft.rrc.texas.gov/link/8b375643-f251-40d0-936d-b16f21f38ded

It contains several shapefiles for each county, from which information about airports, railroads, roads, cities, and water could be extracted.

5) Power Plants

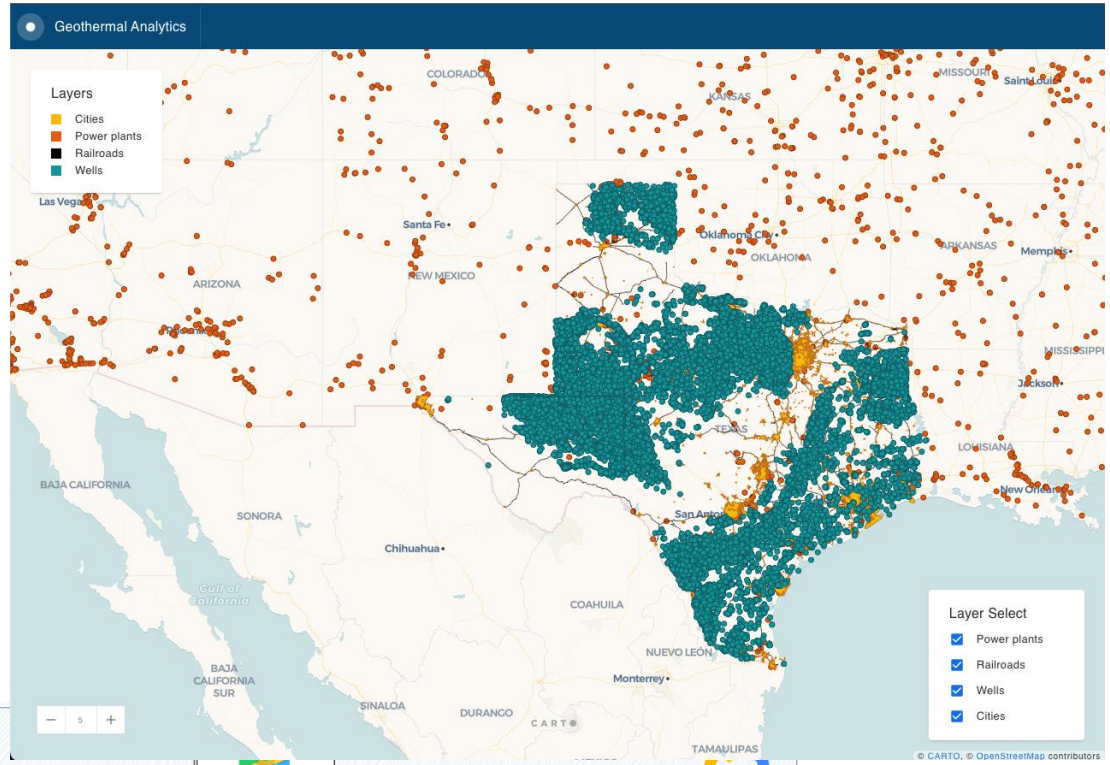
https://geodata.lib.utexas.edu/catalog/stanford-cc957ty2116

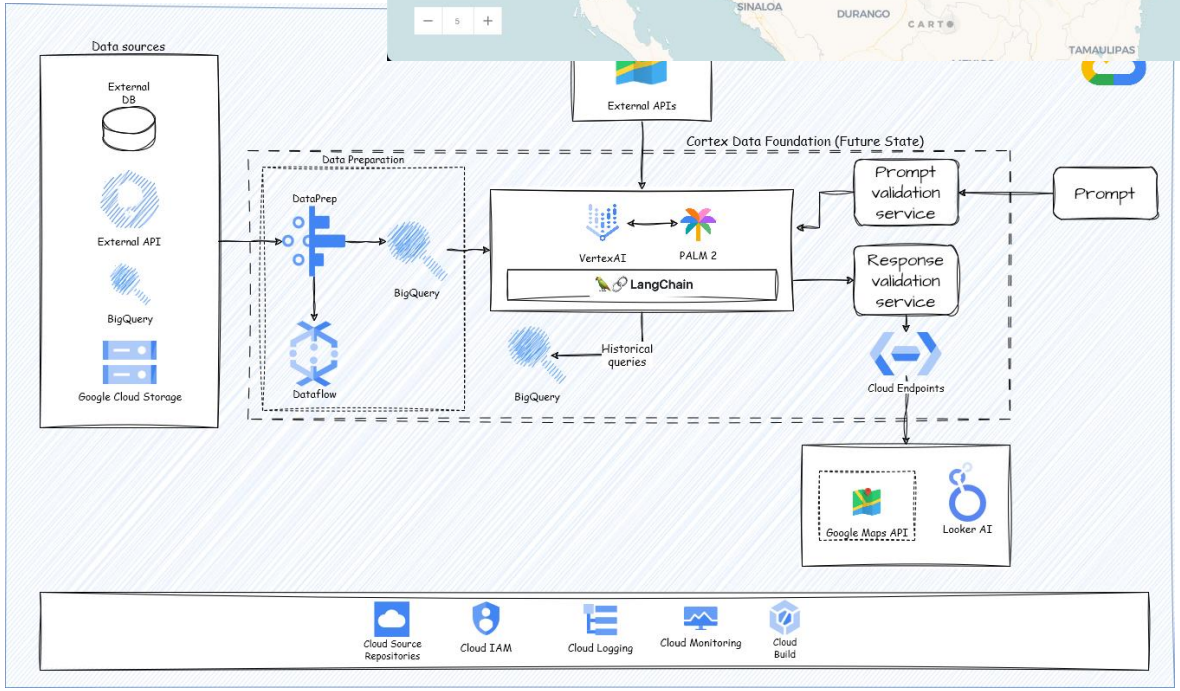
The data represents electric power generating plants in the United States by energy source.

GenAI based GIS accelerator

Using “Hub-and-spoke” model and LangChain as a process orchestrator, this tool can access and analysing GIS data on massive scale in real-time

- Built on VertexAI and PaLM2 LLM, BigQuery, and Google Maps API
- Using CARTO front-end
- Allows for GIS data visualization
- Filtering of search results, updating map on the fly
- Define an area for search on the map as custom polygon
- Retrieves information from mouseover at the point of interest





GenAI based GIS assistant

- Multi-purpose tool designed to help connecting and rapidly analysing geospatial data
- Designed for ESG applications and accurate tracking of the multi-source emissions
  - Using EPA Flight Greenhouse Gas Emissions data & Ren Energy database
- Summarizes and maps electronics supply chain
- Allocates emissions based on Google's spend vs. total spend for a vendor
- Capable of working with normal prompt:
  - Show me the summary emissions from suppliers for Google's Henderson data center
  - Show me a map of the suppliers for the Loudoun County data center