

Accelerator



From Data to Done:

GenAI infused
Application development for Data Pros

Andrea Vasco
andrewankenobi@google.com



Google for Startups



Before getting started, I appreciate if:

- DND your devices
- Respect each other, defer judgement
- Say “Please” and “Thank You” to your fellows
- One conversation at a time
- Encourage wild ideas
- Build on top of others’ ideas (yes, and....)
- When giving feedback, share:
 - ✓ One thing that went well
 - ✓ One thing that you’d have done differently
- Relax and **have fun!**



Your speaker today



Andrea Vasco
EMEA Analytics Lead



- Talk to people, do things and (occasionally) break stuff
- At Google I advise on Analytics, Machine Learning and AI
- Volunteer at Google's Innovation SIG and Startup Accelerator
- Incubated Google Cloud's Football Industry in EMEA
- Before Google, I mostly worked/launched startups, across the UK, Russia and the US
- Supporter of Save the Children, UNWFP, Kurzgesagt, and more
- Married with 2 kids, 6 bikes, 8 guitars, an E-class AMG and a Mini Cooper S

Long. Live. The Startups.



Intro



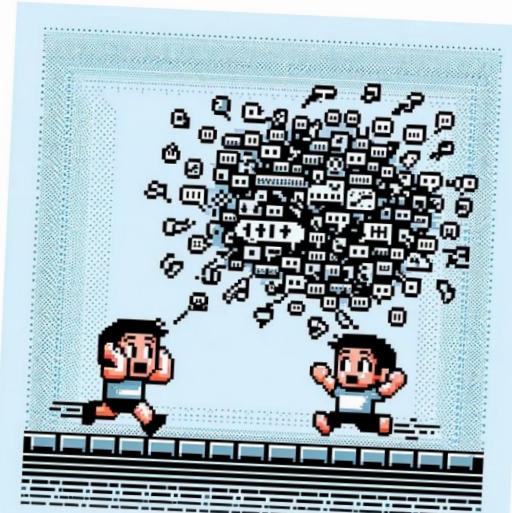
Never ever you ever?

Anyone been
here?

- Spent weeks analyzing data, only to struggle explaining your insights to non-technical stakeholders?
- Created amazing visualizations that got lost in a 50-page report no one read?
- Wished you could turn your complex model into a simple app that anyone could use?
- Felt limited by off-the-shelf tools that don't quite fit your unique data story?
- Dreamed of an AI assistant that could explain your findings as well as you can?

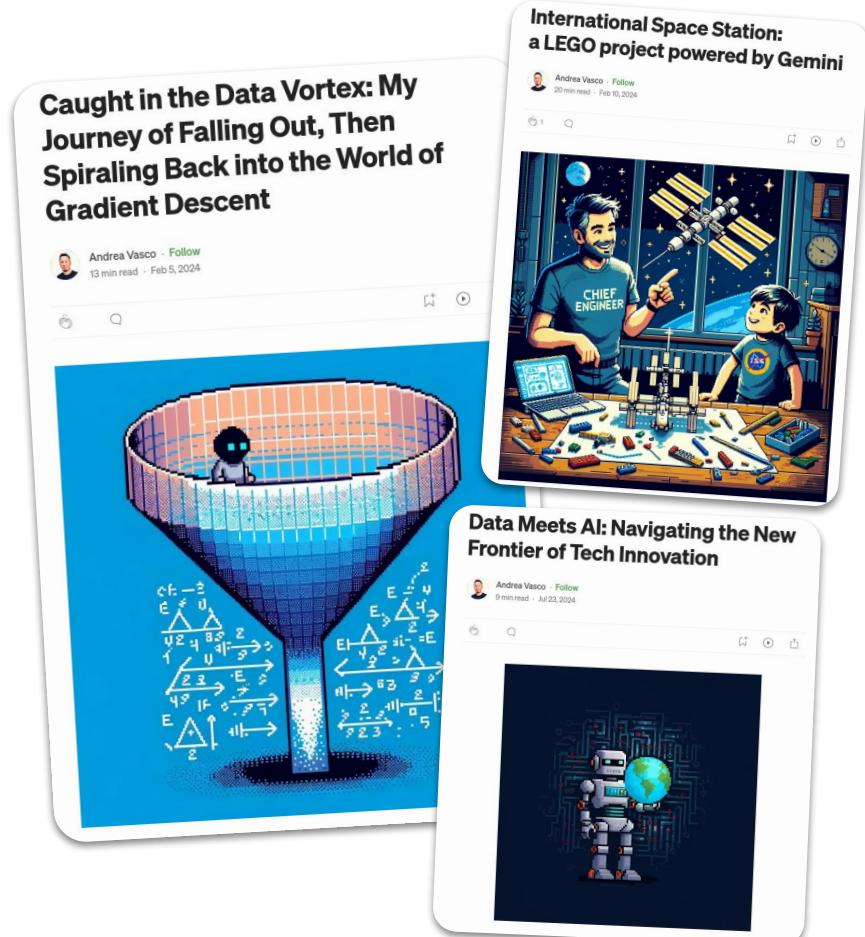
When a Simple Question Becomes a Therapy Session: Understanding the Deep Dive Virus

 Andrea Vasco · Follow
9 min read · Jan 27, 2024



What if...

- You could break free from technical limitations and create your own narratives?
- GenAI could help you communicate your insights as effectively as you analyze them?
- You could design exactly what you need to tell your data story, without being held back by technical debt?
- Your data could speak directly to your audience, in their language?





A game changer



The Convergence of Data *Engineering* and Generative AI

Data engineering and generative AI were once seen as distinct fields. Today, they are rapidly converging. This convergence is transforming the way we **collect, process, understand, and generate value** from data.

Generative AI as a consumer

- Generative AI models are incredibly powerful, but they are fundamentally data-hungry.
- Data engineering provides the mechanisms to clean, organize, and transform data into a format that AI models can understand.
- **High quality data is the foundation for accurate, insightful, and unbiased AI outputs**

Generative AI as a supercharger

- GenAI is not just a consumer of data. It also supercharges traditional data engineering tasks
- Examples: Automating code generation, improving data quality checks, intelligently recommending data sources for analysis, streamline complex semantic logic.
- Data engineers gain efficiency and can focus on strategic issues and complex use cases.



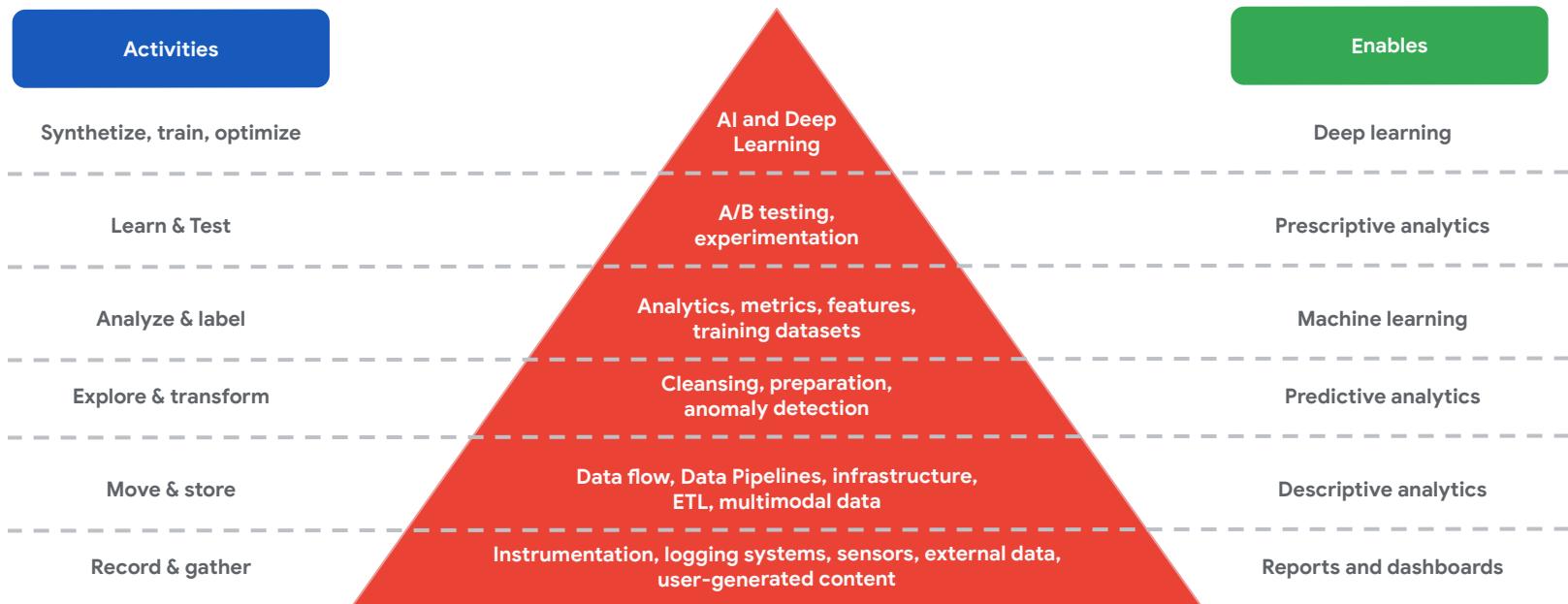
AI Hierarchy of needs



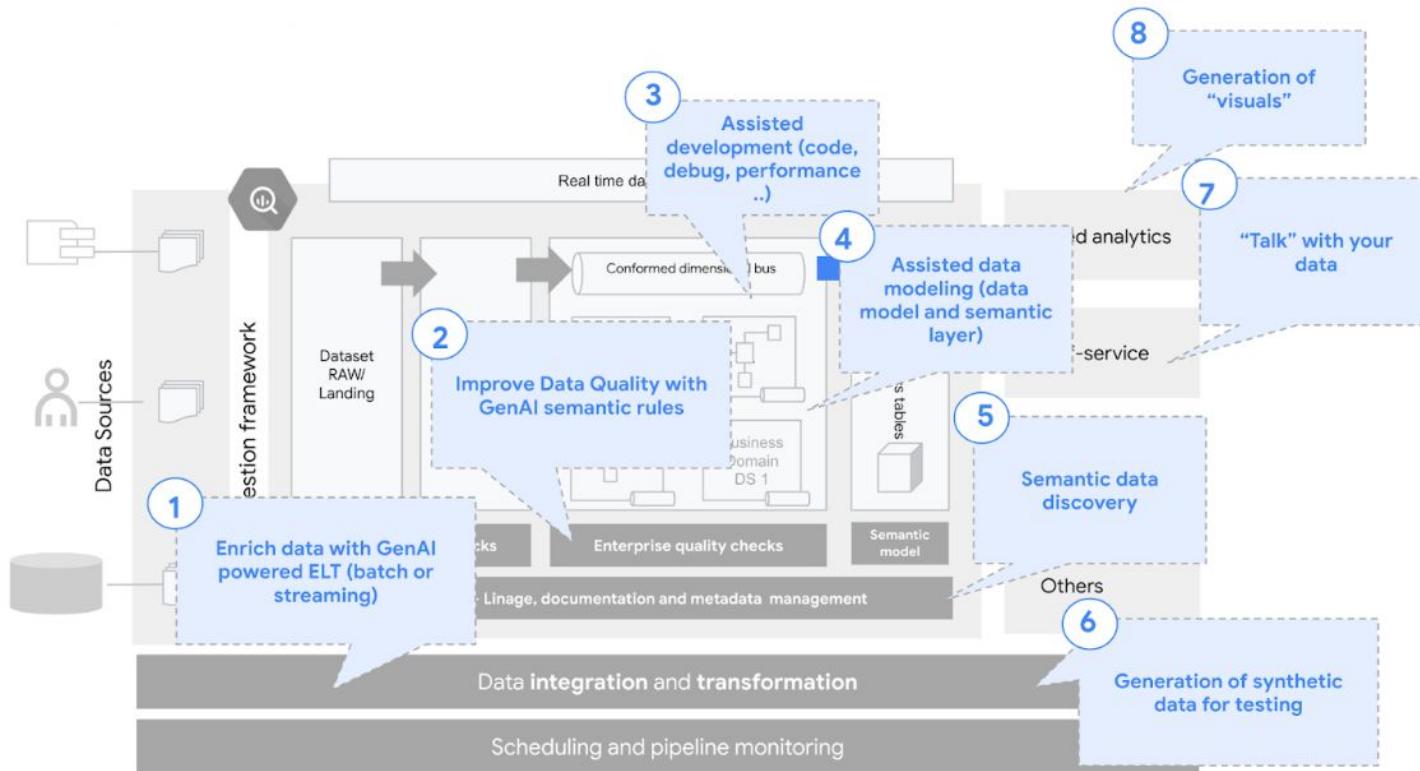
"Think of AI as the top of a pyramid of needs. Yes, self-actualization (AI) is great, but you first need food, water and shelter (data literacy, collection and infrastructure)."

[Monica Rogati](#)

Have you seen
this before?



GenAI can supercharge the whole stack



A few use cases to start from:

- Sentiment Analysis (e.g. give me the sentiment of the online reviews in col x and put the result in col y)
- Data Enrichment (e.g. for each city in col x, give me the country name in col y)
- Entity Extraction (e.g. extract entities from a piece of text)
- Metadata Enrichment (e.g. infer from schema, table content, documentation, lineage information)
- Table and Column content inference (e.g. describe the content of these columns, this table)
- Metadata Semantic Search (e.g. identify the tables necessary to respond to this enquiry)
- Content Generation (e.g. create a customized marketing email based on the customer information stored in a BQ table)
- Test Data (e.g. generate dummy data given a description)
- Code Documentation (e.g. redact a README.md file for this codebase)
- Data Quality (e.g. generate missing information for a specific data quality check)
- Customer Segmentation (e.g. explain the outcome of a K-means analysis in plain english)
- *And many more!*



Democratizing Data Insights with GenAI

- Enabling a wider range of people to get value from data without necessarily having deep coding experience
- Explainable Results: some GenAI tools can explain why an insight or prediction has been made, increasing trust
- Better collaboration between technical and business users on data projects
- The Future of Data-Driven Decision Making: GenAI ushers in an era where data isn't just accessible but it's actively used by a wider range of people across businesses

The image shows the front cover of a book titled "Data-driven decision making: A crash introduction to Decision Intelligence Engineering" by Andrea Vasco. The cover is white with yellow and grey decorative elements. At the top right, there is a small video thumbnail showing a man speaking into a microphone. On the left side, there is a small video thumbnail showing a woman smiling. The bottom right corner features a photo of a diverse group of people in a conference setting. The "Google for Startups" logo is at the bottom center.

Accelerator

Data-driven decision making

A crash introduction to
Decision Intelligence Engineering

Andrea Vasco
andrewankenobi@google.com

Google for Startups



Pitfalls and Success Factors

Common Pitfalls (Slowing Adoption)

- ✖ Overemphasis on technology without cultural change
- ✖ Siloed approach to governance
- ✖ Inadequate focus on data literacy
- ✖ Rigid, inflexible frameworks
- ✖ Neglecting business value
- ✖ Poor communication

Success Factors (Speeding Adoption)

- Strong executive sponsorship
- Phased, iterative approach
- Treating data as a strategic asset
- Robust metadata management
- Continuous improvement mindset
- Investment in automation and AI
- Focus on user experience
- Clear communication and change management
- Strong Partner engagement



BigQuery and Vertex AI

Any integration

Onprem Systems

Other Clouds

Partner Solutions

Other Systems

Any user

Data Practitioners

IT

Software Developers

AI Practitioners



Most powerful Data Platform, BigQuery Suite

AI brought to your data: Responsibly and efficiently do AI at scale by running AI directly on your data

Any data or storage: Structured, semi-structured JSON/function calls, unstructured (text, images, video, audio). All storage options in addition to BigQuery Storage including S3, Iceberg, Delta Lake, Hudi, Azure Storage.

Any engine or storage: Any type of workload whether batch, mini-batch or real time with any engine SQL, Spark, Dataflow and BI tools such as Looker



Most powerful AI Platform, Vertex AI

Models and APIs ready out of the box: Use the best models from Google and open source, no need to build or train yourself

Models customized for your business: Customize Google and open source models with your proprietary data or build your own models from scratch

AI based on your enterprise context: Augment model responses with your broader enterprise data, systems, and governance



Democratization



Productivity



Collaboration



Governance



Scale



Seeing is believing

Welcome to Google's Climate Change Accelerator

What we will cover

- ✓ Ingesting carbon footprint data into BigQuery
- ✓ Analyzing data with SQL and BigQuery ML
- ✓ Creating predictions with traditional ML and GenAI
- ✓ Deploy, monitor and operationalize the model via VertexAI
- ✓ Create a Gemini-based virtual Agent using Vertex AI Agents

Why it matters

- ➔ Transform raw data into user-friendly applications
- ➔ Bridge the gap between data science and software development
- ➔ Create impactful tools for climate change awareness and action

30%

Portion of a household's
carbon footprint from food

Center for Sustainable Systems,
University of Michigan

25%

Maximum potential reduction
in an individual's carbon
footprint through dietary
changes

Climatic Change journal

Challenge

Create a chat app example in 15 mins or so

European Individual Carbon Footprint Dataset

Key Features

- 3000 individuals from 5 European countries
- Age range: 18-80 years old
- Lifestyle factors: Diet, Transportation, Housing

First few rows of the dataset:

```
['France', 19, 'Vegan', 'Public Transit', 'Apartment', 1.95]
['France', 61, 'Vegan', 'Bicycle', 'Apartment', 1.63]
['France', 19, 'Meat-based', 'Car', 'Apartment', 3.82]
['France', 53, 'Meat-based', 'Bicycle', 'House', 2.8]
['UK', 35, 'Meat-based', 'Car', 'House', 5.32]
```

Dataset Statistics:

Average Carbon Footprint: 3.65
Minimum Carbon Footprint: 1.44
Maximum Carbon Footprint: 8.45

Uploading to BigQuery

Uploading Data to BigQuery

- Go to BigQuery console:

```
console.cloud.google.com/bigquery
```

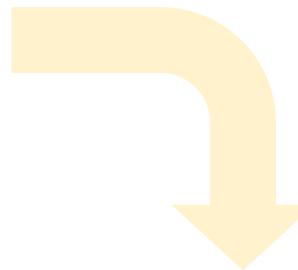
- Create a new dataset:

```
climate_change_accelerator
```

- Create a new table:

```
individual_carbon_footprint
```

- Upload the CSV file we generated



Basic Query

```
SELECT *  
FROM  
`climate_change_accelerator.individual_carbon_footprint`  
LIMIT 5;
```

Running some Insightful Queries

Average Carbon Footprint by Country and Diet

```
SELECT
  Country,
  Diet,
  AVG(Carbon_Footprint) as Avg_Footprint
FROM `climate_change_accelerator.individual_carbon_footprint`
GROUP BY Country, Diet
ORDER BY Country, Avg_Footprint DESC;
```

Impact of Transportation on Carbon Footprint

```
SELECT
  Transportation,
  AVG(Carbon_Footprint) as Avg_Footprint
FROM `climate_change_accelerator.individual_carbon_footprint`
GROUP BY Transportation
ORDER BY Avg_Footprint DESC;
```

Introducing BigQuery ML for Predictive Analysis

What is BigQueryML?

- Machine learning models in SQL
- No need to export data from BigQuery
- Supports LOADS model types
(supervised, unsupervised, deep nn,
automl, external models, LLMs)

Creating a Carbon Footprint Prediction Model

```
CREATE OR REPLACE MODEL
`climate_change_accelerator.carbon_footprint_model`
OPTIONS(model_type='linear_reg',
        input_label_cols=['Carbon_Footprint']) AS

SELECT
*
FROM
`climate_change_accelerator.individual_carbon_footprint`;
```

No feature
encoding needed!

Evaluating the model and making predictions

Evaluating the model

```
SELECT
  *
FROM
  ML.EVALUATE(MODEL
  `climate_change_accelerator.carbon_footprint_model`);
```

Making predictions

```
SELECT
  *
FROM
  ML.PREDICT(MODEL
  `climate_change_accelerator.carbon_footprint_model` ,
  (SELECT
    'France' AS Country,
    55 AS Age,
    'Meat-based' AS Diet,
    'Bicycle' AS Transportation,
    'House' AS Housing));
```

Register the model in Vertex AI

```
ALTER MODEL IF EXISTS
`climate_change_accelerator.carbon_footprint_model`
SET OPTIONS
(vertex_ai_model_id="carbon_footprint_model");
```

Testing the Vertex AI Endpoint

Endpoint Overview

- Our BigQuery ML model is now accessible via Vertex AI
- Scalable and managed by Google Cloud
- Enables real-time predictions for applications
- Supports batch predictions for large-scale analysis
- Allows for easy integration with various applications and services

JSON Payload Structure

```
{  
  "instances": [  
    {  
      "Country": "France",  
      "Age": 55,  
      "Diet": "Meat-based",  
      "Transportation": "Bicycle",  
      "Housing": "House"  
    }  
  ]  
}
```

CURL Command

```
curl -X POST \  
  -H "Authorization: Bearer $(gcloud auth  
print-access-token)" \  
  -H "Content-Type: application/json" \  
  https://us-central1-aiplatform.googleapis.com/v1/projects/222  
594707808/locations/us-central1/endpoints/5825618323747897344  
:predict \  
  -d <JSON Payload above>
```

Creating an Agent / 1



Search
Get quality results out-of-the-box and easily customize the engine

SELECT



Chat
Answer complex questions out-of-the-box

SELECT



Recommendations
Create a content recommendation engine

SELECT



Agent [PREVIEW](#)
Built using natural language, agents can answer questions from data, connect with business systems through tools, and more

SELECT



First, we create an application

Creating an Agent / 2

← Tools  Save

Using tools, you can connect agents to external systems. These systems can augment the knowledge of agents and empower them to execute complex tasks efficiently. [Learn more](#)

Tool name*
Carbon Footprint Predictor

Type*
OpenAPI

Connect to an external API using an OpenAPI tool.

Description

Provide a description of this tool. This description is provided to the model as context informing how the tool is used.

Description

Schema

Provide the [OpenAPI schema](#) defining this tool's API. YAML and JSON are supported. You can use the [OpenAPI schema builder](#) to assist with building your OpenAPI schema specification.

JSON YAML

Samples

```
1 openapi: 3.0.0
2   info:
3     title: Carbon Footprint Predictor
4     version: 1.0.0
5     description: Predicts carbon footprint based on lifestyle factors
6
7   servers:
8     - url: https://us-central1-aiplatform.googleapis.com/v1/projects/222594707808/locations/us-central1/endpoints/5825618323747897344
9
10  paths:
11    /predict:
12      post:
13        summary: Predict Carbon Footprint
14        operationId: predictCarbonFootprint
15        requestBody:
16          required: true
17          content:
```

Authentication

Authentication type

Service agent token OAuth API Key Bearer Token

Service agent auth type

Controls the type of auth token to be generated using this project's [Dialogflow service agent](#). The generated token is sent in the Authorization header.

ID token Access token



Using Service Accounts

Second, we create a tool
to invoke the Vertex AI Endpoint

OpenAPI YAML schema?!?!

OpenAPI YAML

```
openapi: 3.0.0
info:
  title: Carbon Footprint Predictor
  version: 1.0.0
  description: Predicts carbon footprint based on lifestyle factors

servers:
  - url:
    https://us-central1-aiplatform.googleapis.com/v1/projects/222594707808/locations/us-central1/endpoints/5825618323747897344

paths:
  /predict:
    post:
      summary: Predict Carbon Footprint
      operationId: predictCarbonFootprint
      requestBody:
        required: true
        content:
          application/json:
            schema:
              $ref: '#/components/schemas/PredictionRequest'
      responses:
        '200':
          description: Successful prediction
          content:
            application/json:
              schema:
                $ref: '#/components/schemas/PredictionResponse'
```

```
components:
  schemas:
    PredictionRequest:
      type: object
      properties:
        instances:
          type: array
          items:
            type: object
            properties:
              Country:
                type: string
              Age:
                type: integer
              Diet:
                type: string
              Transportation:
                type: string
              Housing:
                type: string
      required:
        - Country
        - Age
        - Diet
        - Transportation
        - Housing
    instances
    PredictionResponse:
      type: object
      properties:
        predictions:
          type: array
          items:
            type: object
            properties:
              Carbon_Footprint:
                type: number
```

Easy-peasy



I have this API that works fantastic:

```
curl -X POST \
-H "Authorization: Bearer $(gcloud auth print-access-token)" \
-H "Content-Type: application/json" \
https://us-central1-
aiplatform.googleapis.com/v1/projects/222594707808/locations/us-
central/endpoints/5825618323747897344:predict \
-d {
"instances": [
{
"Country": "France",
"Age": 55,
"Diet": "Meat-based",
"Transportation": "Bicycle",
"Housing": "House"
}
]
}
```

and responds:

```
{
"predictions": [
[
3.216961212414454
]
],
"deployedModelId": "8123835120311336960",
"model": "projects/222594707808/locations/us-
central/models/carbon_footprint_model",
"modelDisplayName": "carbon_footprint_model",
"modelVersionId": "1"
}"
```

I have to specify the openapi schema in YAML for this api.

Creating an Agent / 3

Carbon Footprint Agent Version history Save

Agent name* Carbon Footprint Agent

An agent is the basic building block of a Vertex AI Conversation app. Each agent is defined to handle specific tasks. [Learn more](#)

Goal

Goal* To help users understand, calculate, and reduce their personal carbon footprint by providing personalized estimates based on lifestyle factors and offering actionable suggestions for more sustainable living

High level description of the goal the agent intends to accomplish. [Learn more](#)

Instructions

Instructions

- Greet the user and introduce yourself as a Carbon Footprint Assistant.
- Ask how you can help them understand or reduce their carbon footprint today.
- Summarize the user's request and confirm that you've understood correctly.
- If the user wants to calculate their carbon footprint:
 - Explain that you'll need some information about their lifestyle.
 - Ask for their country of residence, age, diet type, primary mode of transportation, and housing type.
 - If any information is missing, ask follow-up questions to gather all required details.
- Use \${TOOL:Carbon Footprint Predictor} to calculate the user's carbon footprint based on the provided information. Make sure to ask for:
 - Housing (map the answer as either Apartment or House)
 - Country (map the answer as either Spain, France, Italy, UK or Germany)
 - Age
 - Diet (map the answer as either Vegan, Meat-based or Vegetarian)
 - Primary Transportation (map the answer as either Bicycle, Car or Public Transit)
- Clearly explain the results of the prediction to the user, including what their carbon footprint means in context.
- Offer personalized suggestions for reducing their carbon footprint based on the prediction results.
- Ask if the user would like more specific information or advice on any aspect of reducing their carbon footprint.
- If the user has additional questions, provide relevant information or use the tool again if needed.
- Summarize the key points of the conversation, including the user's current carbon footprint and top recommendations for reduction.
- Thank the user for their interest in understanding and reducing their carbon footprint.
- Encourage the user to implement the suggested changes and to check back in the future to recalibrate their carbon footprint.
- Say goodbye and offer to help again anytime they have questions about carbon footprint or sustainable living.

Ordered list of step-by-step execution instructions to accomplish target goal. Specify instructions using [unordered markdown list](#) syntax. Instructions may be nested to specify substeps. Use the syntax \${TOOL: tool name} to reference a tool, and \${AGENT: agent name} to reference another agent. [Learn more](#)

Last, we create the agent

Guess how I created most of these?

Data Cleansing in the prompt

- If any information is missing, ask follow-up questions to gather all required details.
- Use \${TOOL:Carbon Footprint Predictor} to calculate the user's carbon footprint based on the provided information. Make sure to ask for:

- Housing (map the answer as either Apartment or House)
- Country (map the answer as either Spain, France, Italy, UK or Germany)
- Age
- Diet (map the answer as either Vegan, Meat-based or Vegetarian)
- Primary Transportation (map the answer as either Bicycle, Car or Public Transit)

Creating an Agent / 4

review agent: Carbon Footprint Agent
Max token count per LLM call in current turn: input 2361, output 32

Agent invocations

Carbon Footprint Agent

hello

Hi there! I'm your Carbon Footprint Assistant. How can I help you understand or reduce your carbon footprint today?

I want to reduce my footprint. Let's start with my current one. I live in the UK.

I can help you with that! To calculate your carbon footprint, I'll need to ask you a few questions about your lifestyle. Are you ready to get started?

yes

Great! First, what is your age?

33

Thank you. What is your primary mode of transportation?

Public Transportation

Got it. What type of diet do you follow?

Vegetarian

Thank you. Finally, what type of housing do you live in?

House

Carbon Footprint Predictor
Tool

predictCarbonFootprint Action

1 Input parameters

1 Output parameters

Based on the information you've provided, your current carbon footprint is estimated to be 3.15 metric tons of CO2 equivalent per year. This is slightly lower than the average carbon footprint in the UK, which is around 4.5 metric tons of CO2 equivalent per year.

We can now test it

integrations [preview]

Text Based

Suggested

- Dialogflow Messenger [Connect](#)
- Messenger [Connect](#)
- LINE [Connect](#)
- Google Chat [Connect](#)

- Slack [Connect](#)
- MMD Smart [Connect](#)
- Twilio (Text Messaging) [Connect](#)
- Discord [Connect](#)

- Spark (Webex) [Connect](#)
- Telegram [Connect](#)
- Viber [Connect](#)
- Azure Bot Service/Skype [Connect](#)

3D Avatar

- soul machines [Connect](#)

And integrate it

Creating an Agent / 5

Dialogflow Messenger

Dialogflow Messenger brings a rich UI for Dialogflow CX that enables developers to easily add conversational agents to websites. [View Documentation](#)

You can also, modify the interactive visual interface, by going to Agent settings then click on Multimodal, and then click on [UI settings](#)

 End-user interactions with the Dialogflow Messenger widget will be billed to your GCP account.

Add this agent to your website by copying the code below:

```
<link rel="stylesheet" href="https://www.gstatic.com/dialogflow-console/fast/df-messenger/prod/v1/themes/df-messenger-default.css">
<script src="https://www.gstatic.com/dialogflow-console/fast/df-messenger/prod/v1/df-messenger.js"></script>
<df-messenger
  location="us-central1"
  project-id="argon-zoo-413112"
  agent-id="63c7e607-a4ae-45f1-8096-688f2143901a"
  language-code="en"
  max-query-length="-1">
  <df-messenger-chat-bubble
    chat-title="Carbon Footprint Estimator">
  </df-messenger-chat-bubble>
</df-messenger>
```

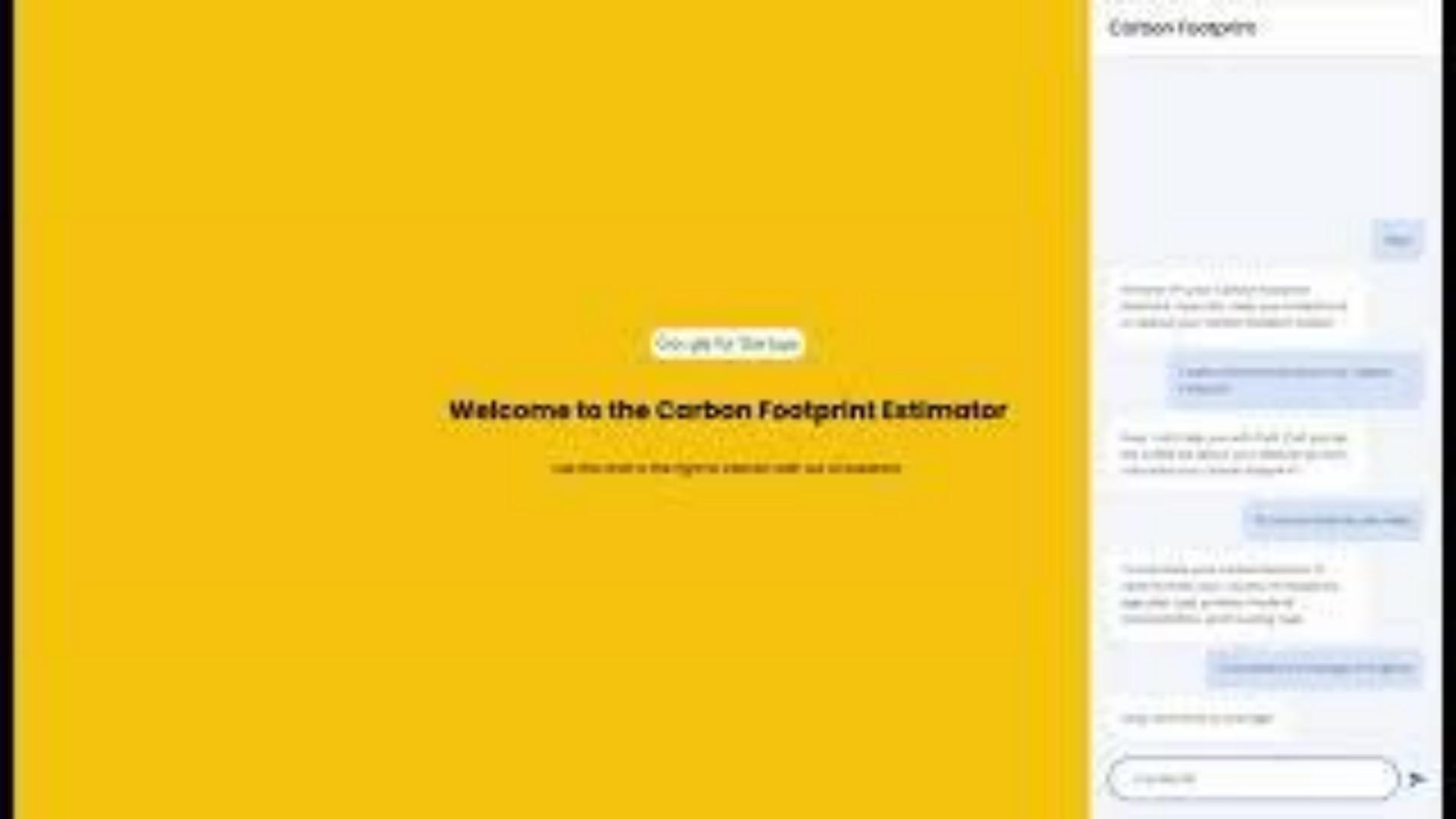
Environment: **Draft**

[Disable the unauthenticated API](#)

[Try it now](#)

[Done](#)

Cut and paste HTML code

A large yellow rectangular area covers the majority of the page. Inside this area, there is a white rounded rectangle containing the following text:

Calculate your carbon footprint
Discover opportunities to reduce emissions
And track your reduction progress over time

Calculate Your Footprint

Welcome to the Carbon Footprint Estimator

Use this online tool to estimate your carbon footprint.

How much does your electricity cost per month? Use our free service to calculate your carbon footprint.



Find out how much you can save by reducing your energy use. Our calculator will help you make informed decisions about your energy use.



Get started today!

Get Started



github.com/andrewankenobi/Climate-Change

Screenshot of the GitHub repository page for Climate-Change.

The repository is public and has 1 branch and 0 tags. The main branch is selected. A commit from andrewankenobi was made 4 minutes ago, with the commit hash being 329f423.

The commit message is "Committed". The files listed in the commit are:

File	Commit Status	Time Ago
demo files	Committed	4 minutes ago
static	Committed	4 minutes ago
.DS_Store	Committed	4 minutes ago
GFSA Climate Change demo.webm	Committed	4 minutes ago
README.md	Committed	4 minutes ago
app.py	Committed	4 minutes ago
requirements.txt	Committed	4 minutes ago



Key Takeaways

Welcome to Google's Climate Change Accelerator

- Maximize your use of GCP's tools and resources – we are here to help you every step of the way.
- Unlock the potential of BigQuery, Vertex AI, and other GCP services for your climate projects.
- Leverage Google to navigate the GCP ecosystem and accelerate your development process.
- Experience the accelerated development and iteration that LLMs bring to prototyping
- Together, we'll harness the power of GCP to create real-world impact in the fight against climate change.



Questions?



Thank you!