Digital Twin Testing and AI-Assisted Search

Jet Propulsion Laboratory
California Institute of Technology

Gerald Parish
JPL Group 345C: Advanced Electronics Systems & Technology

May 1st, 2025

Mentors: **Dr. Thomas Lu** & **Dr. Edward Chow**

Itinerary

- About Me
- Project Goals
- Web App
- Latency Testing
- AI-Assisted Search
- Lessons Learned

About Me

About Me

- UARK 2025
  - Comp. Sci. & Poli. Sci.

- Previously at:

ConocoPhillips

Tyson Foods

EmployWell

About Me

About Me

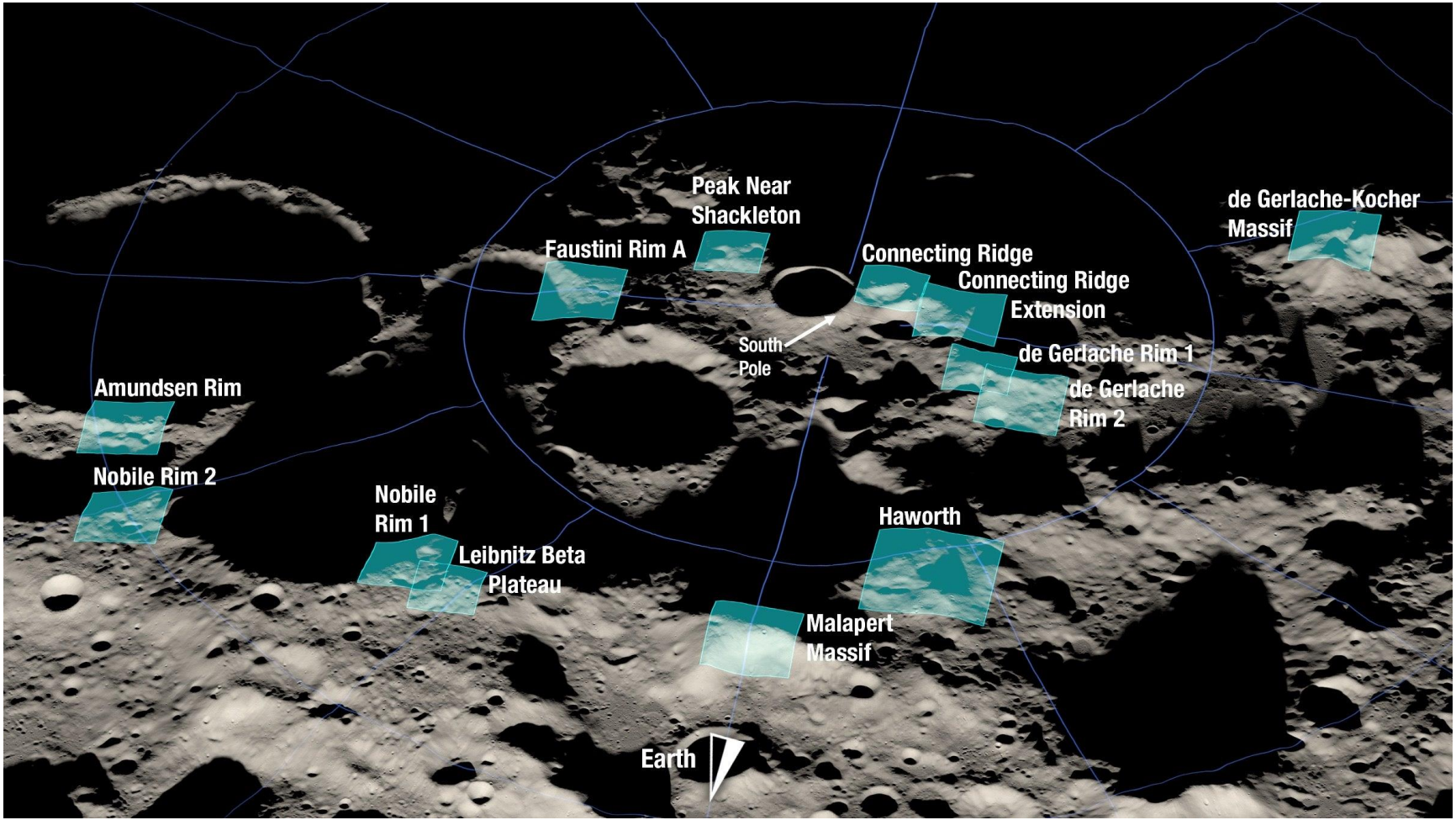
About Me

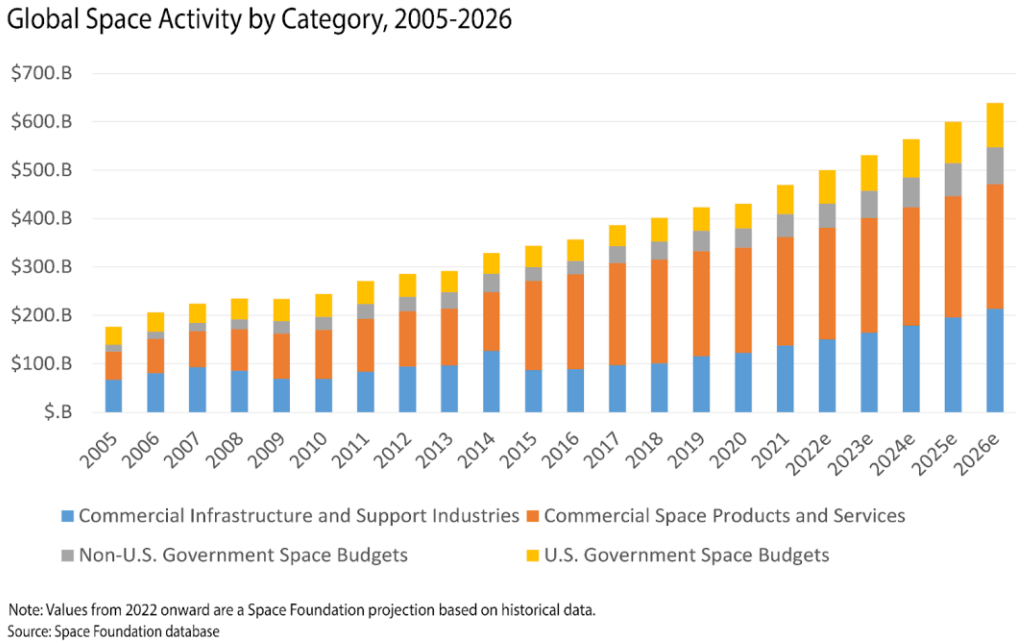
About Me

Future Plans

- Emerging Digital Technologies Team @ ConocoPhillips
  - Gen. AI, Digital Twins, IoT
- "AI Analyst"

Project Goals/Web App


Peak Near Shackleton
Faustini Rim A
Connecting Ridge
Connecting Ridge Extension
de Gerlache-Kocher Massif
South Pole
de Gerlache Rim 1
de Gerlache Rim 2
Amundsen Rim
Nobile Rim 2
Nobile Rim 1
Leibnitz Beta Plateau
Haworth
Malapert Massif
Earth



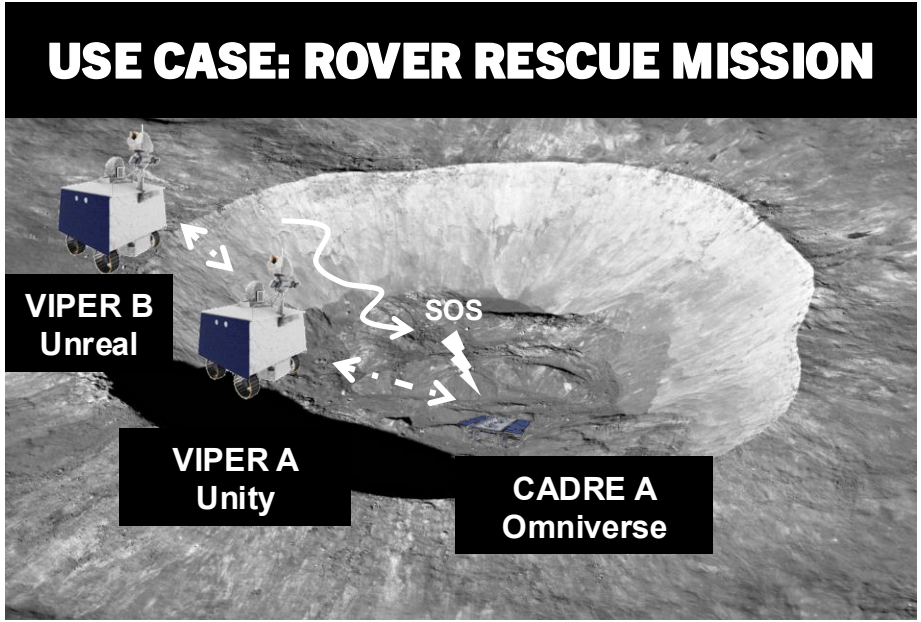
"Space Industry Growth: Where Are the Opportunities in 2022?"
(Space Foundation, 2022)

And now a video...

Project Goals

- Spearhead a new "common language", HSML, for Web 3.0
- Provide an easy way to generate HSML-compliant JSON objects
- Prove the HSML API doesn't compromise latency

Use Case Demonstration

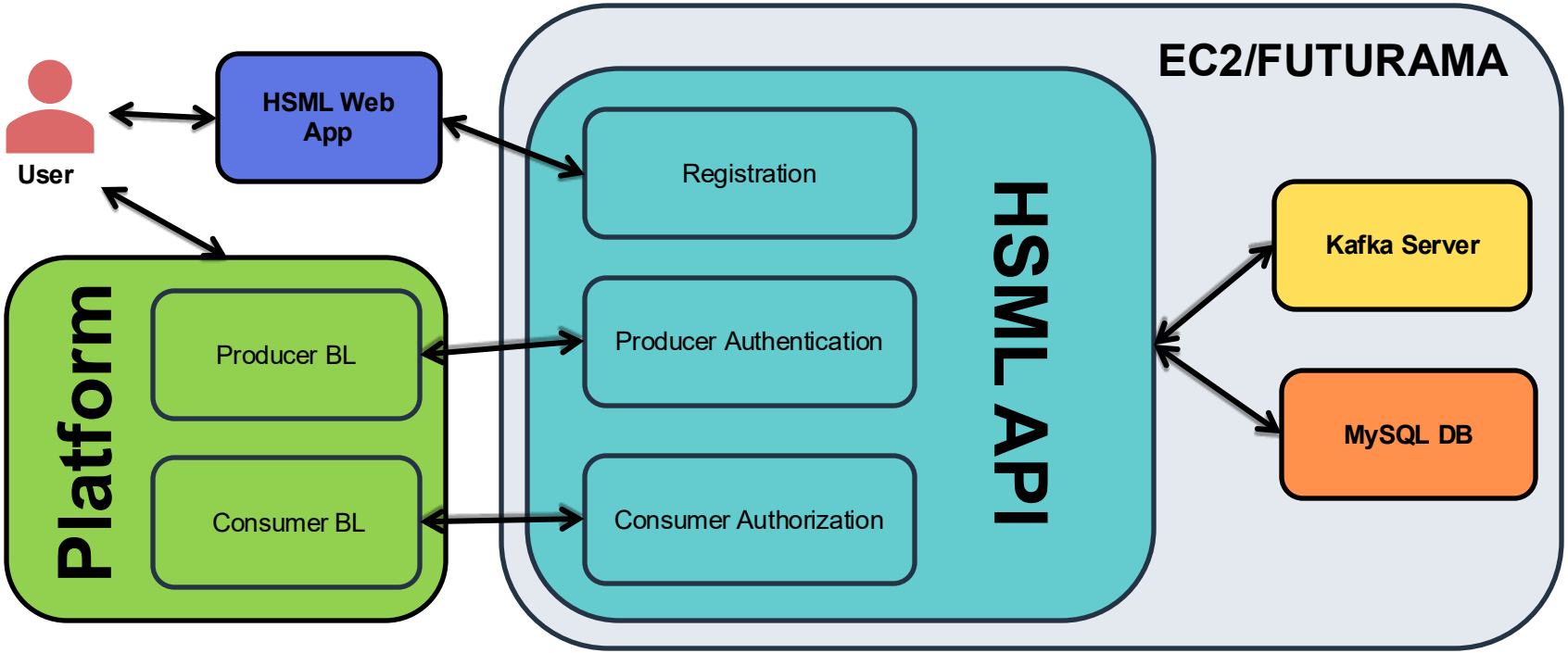


Simulation experiment to be carried out to evaluate the feasibility of the Spatial Web standard in achieving DT interoperability:

- Simulating **multi-agent robotic systems** in a lunar environment
- Testing **interoperability, data trust, and automation** across platforms.
- Demonstrating collaboration possibilities on the Moon, such as **rover assistance scenarios** using Digital Twins

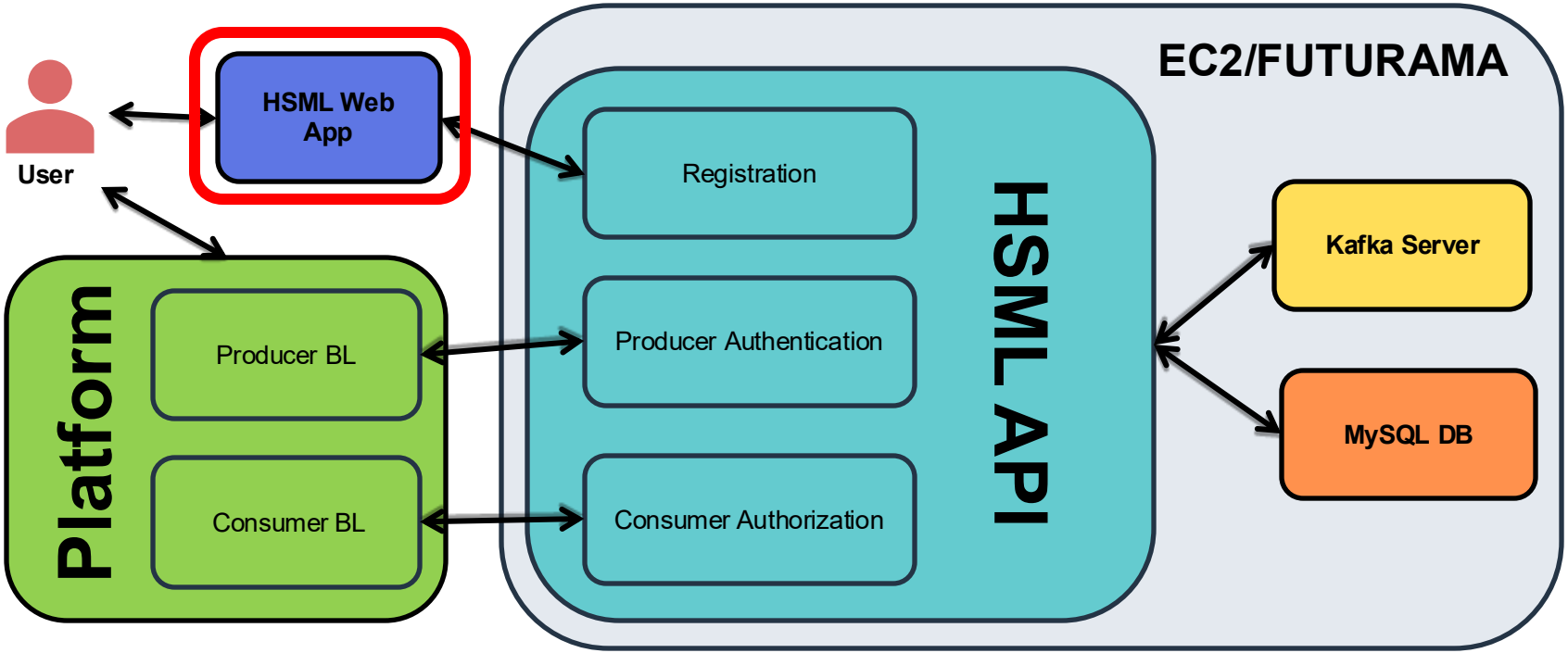
Methodology – HSML API

HSML API ARCHITECTURE

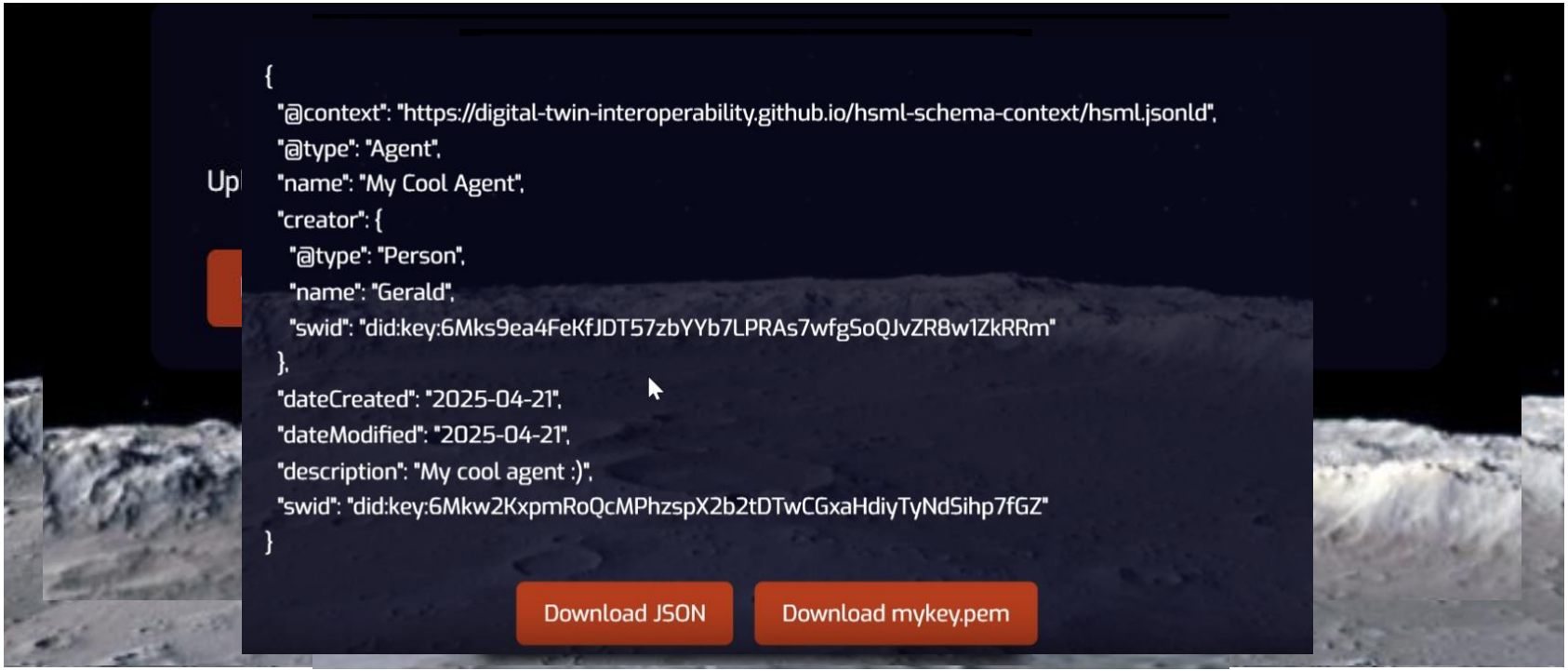


Methodology – HSML API

HSML API ARCHITECTURE



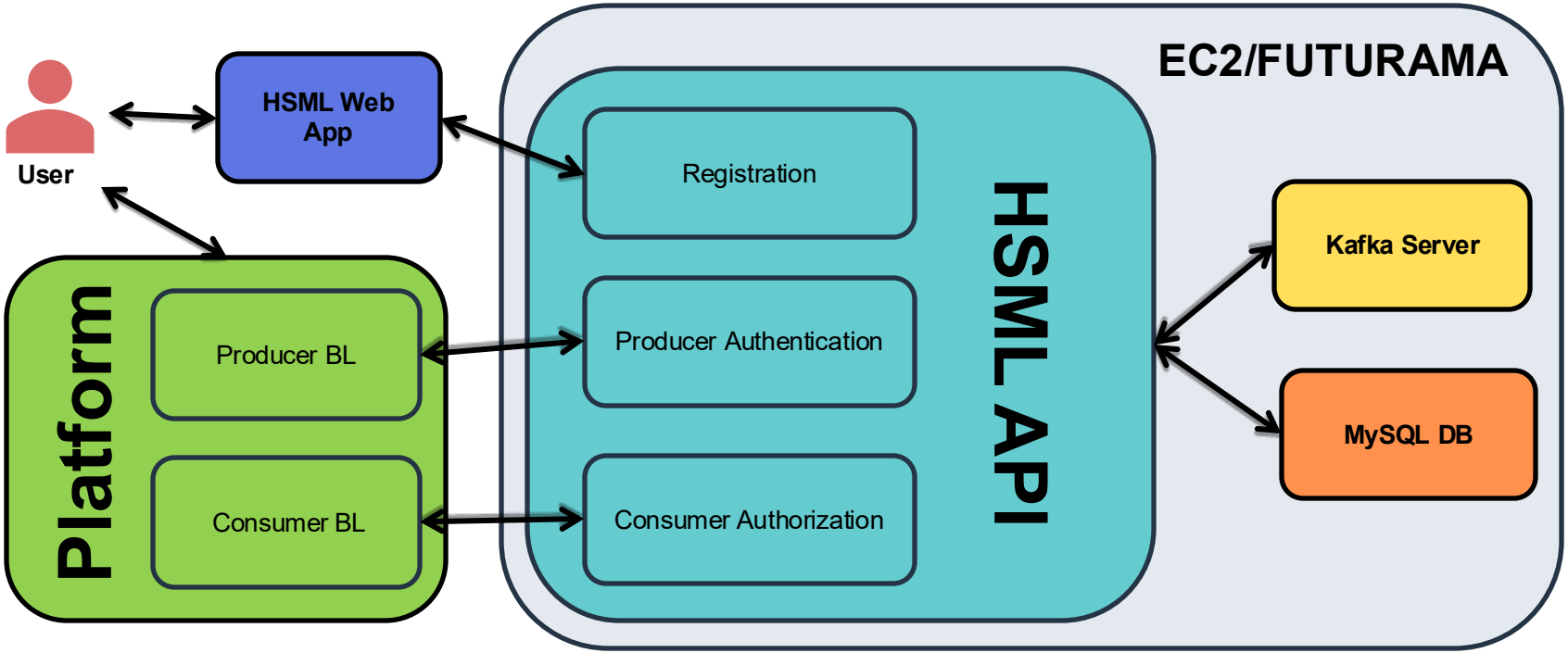
HSML Web App


{
"@context": "https://digital-twin-interoperability.github.io/hsml-schema-context/hsml.jsonld",
"@type": "Agent",
"name": "My Cool Agent",
"creator": {
"@type": "Person",
"name": "Gerald",
"swid": "did:key:6Mks9ea4FeKfJDT57zbYYb7LPRAs7wfgSoQJvZR8w1ZkRRm"
},
"dateCreated": "2025-04-21",
"dateModified": "2025-04-21",
"description": "My cool agent :)",
"swid": "did:key:6Mkw2KxpmRoQcMPhzspX2b2tDTwCGxaHdiyTyNdSihp7fGZ"
}
Download JSON
Download mykey.pem

What I used...

Methodology – HSML API

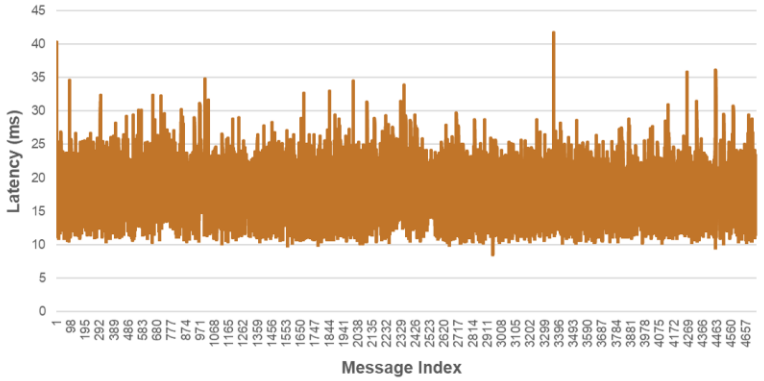
HSML API ARCHITECTURE



Latency Testing

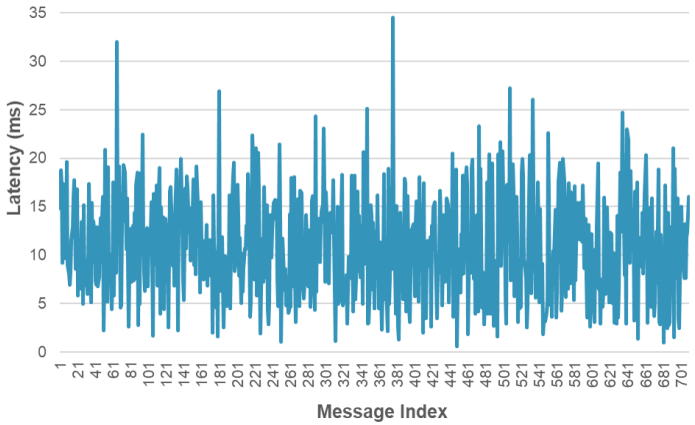
Key Results

LATENCY DIRECTLY THROUGH KAFKA



Average latency = 18.20 ms , Max =41.67 ms, Min = 8.53 ms, Median = 18.02 ms, Std. dev. = 4.42 ms.

LATENCY WITH HSML API



Average latency = 10.83 ms, Max = 34.54 ms, Min = 0.60 ms, Median = 10.57 ms, Std. dev. = 5.39 ms.

AI-Assisted Search

PROBLEM: We have a **LOT** of documentation.

Alicia Sanjurjo Barrio, 349-B Affiliate
02/03/25

HSML Schema

HSML & the Spatial Web

The **Spatial Web** standard aims to set the guidelines that will define the Web 3.0, a computing environment that will blur the lines between the physical and digital worlds. The Spatial Web is an evolution of the internet, where we move from a 2D environment to a 3D virtual space that will allow people to interact with places, things, and each other (René & Mapes, 2019). To be able to place information spatially and contextually, we require a detailed map that explains how all the different data is related to each other in the system. This is where the **Hyperspace Modeling Language (HSML)** comes into play. HSML is a multi-dimensional ontology for encoding fundamental elements and the relationships between them (IEEE Computer Society, 2024). HSML does for the Spatial Web as HTML does for the WWW.

HSML Ontology Hierarchy

This ontology is composed of different types of elements, also known as **classes**. The highest class in the hierarchy is **Entity.** Entities are the base items used across the Spatial Web in HSML, and all the other classes are subtypes of the Entity base class. These include, but are not limited to, **Activity**, **Agent**, **Contract**, **Channel**, **Credential**, **Domain**, **Hyperspace/Space**, and **Time**

Deploy

Upload document

PDF or DOCX

Drag and drop file here

Limit 200MB per file • PDF, DOCX

Browse files

Model Settings

Provider

Local

Local model

Gemma 3 27B IT

Search Settings

Search Mode

Fine Search

Fine Search uses an AI router to select the most relevant document for your query.

Embeddings Management

Documents

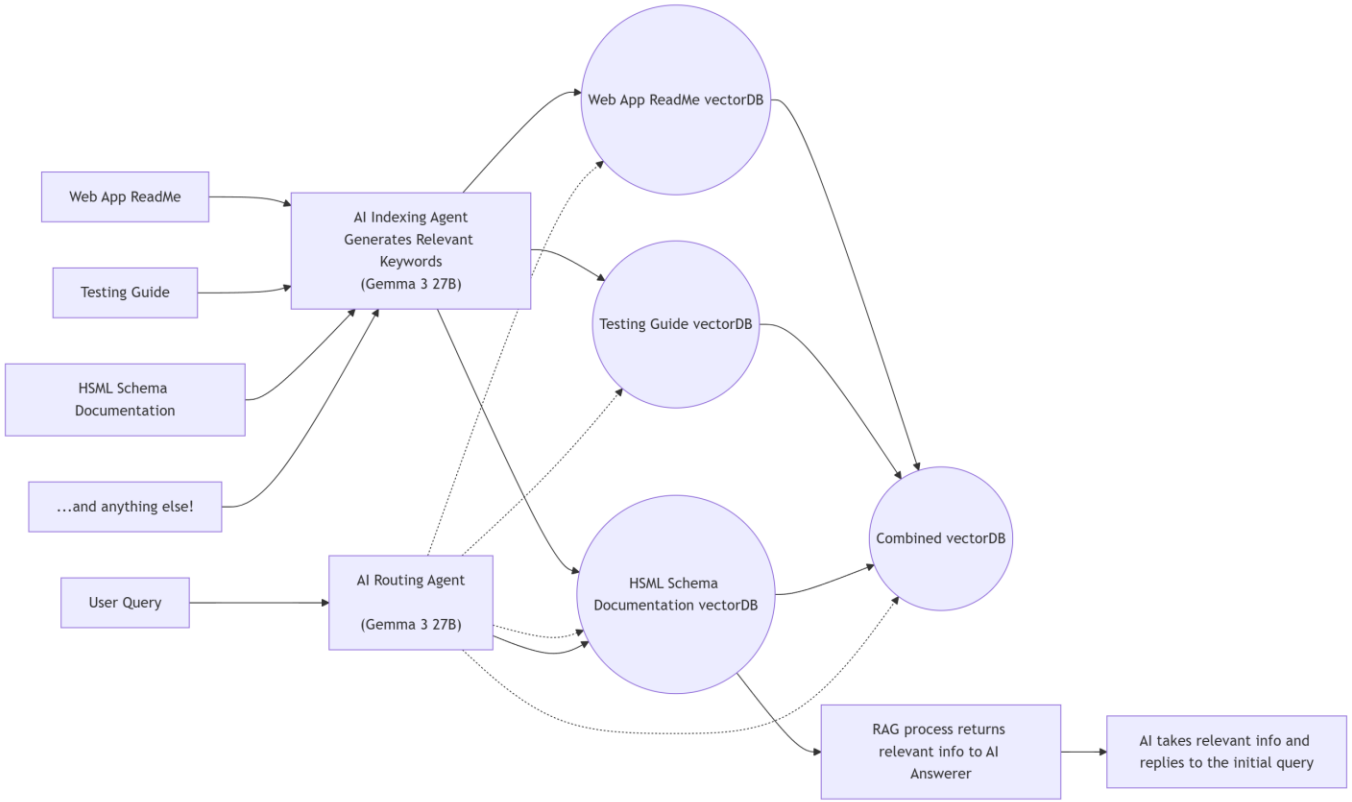
Using file: README.docx

How many pages are there in the web app, and what do they each do?

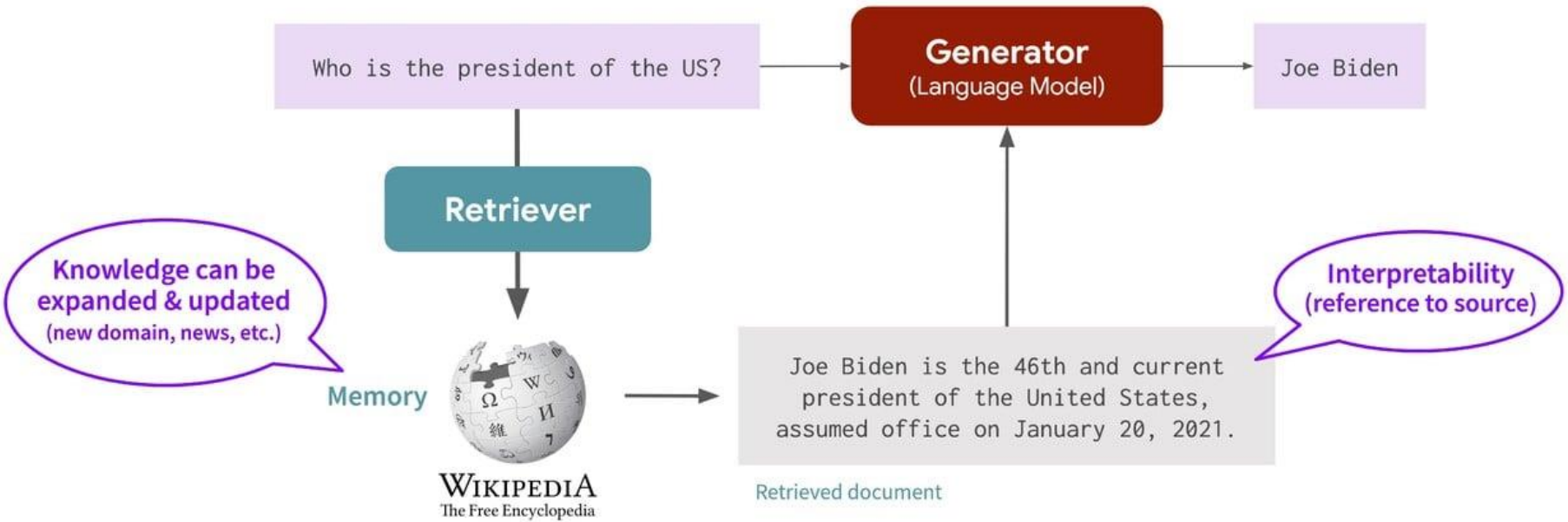
There are 7 pages in the HSML Web App:

- **index.html:** Allows the user to select between Entity, Credential, and Agent, then routes them to the relevant page.
- **landing.html:** The first page a user sees, prompting them to login or register.
- **login.html:** The login page, which prompts the user to upload a .pem file.
- **register.html:** Prompts the user to create a new Person or Organization object.
- **result.html:** Shows the formatted JSON result and prompts the user to download the associated key.pem file and HSML JSON file.
- **agent.html:** Prompts the user for relevant fields, then sends them to `result.html`.
- **credential.html:** Prompts the user for relevant fields, then sends them to `result.html`.
- **entity.html:** Prompts the user for relevant fields, then sends them to `result.html`.

Ask me anything...

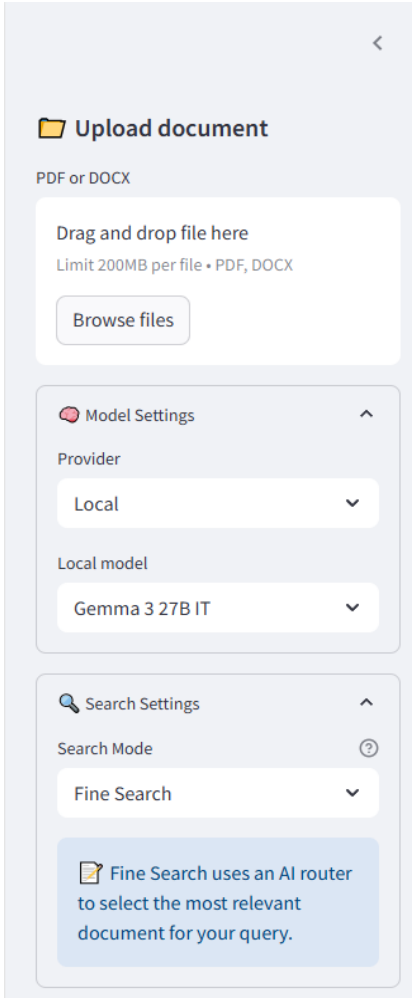

Web App ReadMe
Testing Guide
HSML Schema Documentation
...and anything else!
User Query
AI Indexing Agent Generates Relevant Keywords (Gemma 3 27B)
AI Routing Agent (Gemma 3 27B)
Web App ReadMe vectorDB
Testing Guide vectorDB
HSML Schema Documentation vectorDB
Combined vectorDB
RAG process returns relevant info to AI Answerer
AI takes relevant info and replies to the initial query

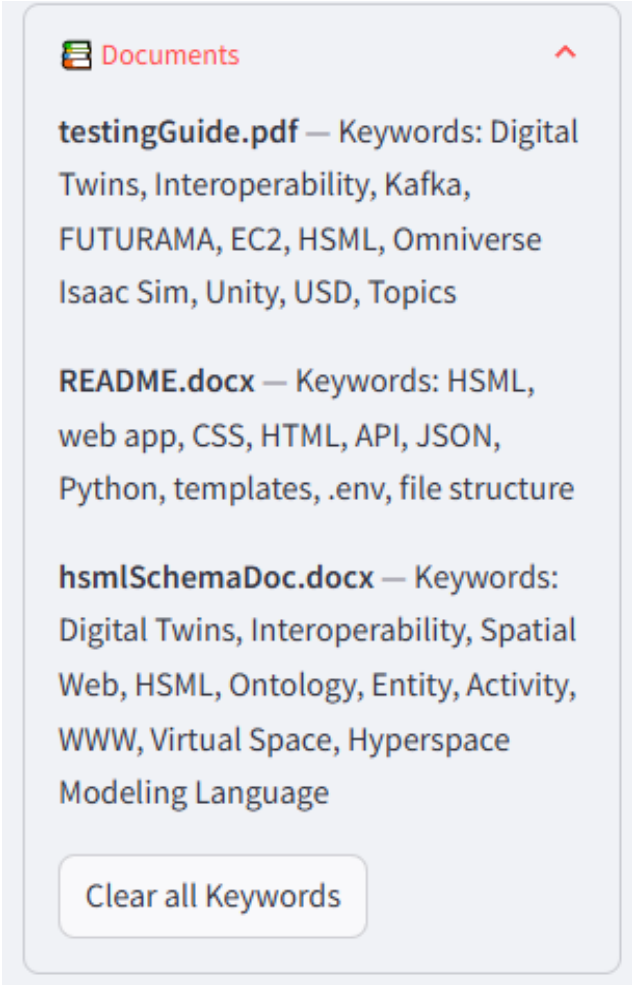
Retrieval augmentation



Features

- Model Agnostic
  - Natively supports: Gemma 3 27B, Gemma 3 4B, LLaMa 3.2 3B, and OpenAI Suite (ex. GPT 4.1)
- Bring your own files
  - **AI Indexing Agent:** Assigns keywords to each new document
- "Broad Search" vs "Fine Search"
  - **Broad:** Search the combined DB
  - **Fine:** Search the doc-specific DB
    - **AI Routing Agent:** Determines most relevant DB


Upload document
PDF or DOCX
Drag and drop file here
Limit 200MB per file • PDF, DOCX
Browse files
Model Settings
Provider
Local
Local model
Gemma 3 27B IT
Search Settings
Search Mode
Fine Search
Fine Search uses an AI router to select the most relevant document for your query.


Documents
testingGuide.pdf — Keywords: Digital Twins, Interoperability, Kafka, FUTURAMA, EC2, HSML, Omniverse Isaac Sim, Unity, USD, Topics
README.docx — Keywords: HSML, web app, CSS, HTML, API, JSON, Python, templates, .env, file structure
hsmlSchemaDoc.docx — Keywords: Digital Twins, Interoperability, Spatial Web, HSML, Ontology, Entity, Activity, WWW, Virtual Space, Hyperspace Modeling Language
Clear all Keywords

Under the Hood

- Python
- Streamlit
  - Easy to style, easy to write, quick turnaround times
- LangChain
  - Retrieves 3 (fine) or 5 (broad) chunks, then sends relevant info to AI Answerer
  - **Chunks:** Sections of user-submitted text (1000 tokens = approx. 500 words, with 200 token overlap)
- LM Studio

```
st.sidebar.header("📁 Upload document")
uploaded = st.sidebar.file_uploader("PDF or DOCX", type=["pdf","docx"])
if uploaded:
    name     = uploaded.name
    raw      = uploaded.getbuffer()
    new_hash = hashlib.sha256(raw).hexdigest()
    prev     = registry.get(name, {})

    if prev.get("hash") != new_hash:
        with open(name, "wb") as f:
            f.write(raw)

        loader   = Docx2txtLoader if name.lower().endswith(".docx") else PyPDFLoader
        splitter = RecursiveCharacterTextSplitter(chunk_size=1000, chunk_overlap=200)
        docs     = splitter.split_documents(loader(name).load())
        for d in docs:
            d.metadata["source"] = name

        embedder = DedicatedEmbeddings()
        faiss_db = FAISS.from_documents(docs, embedder)
        save_faiss_index(faiss_db, name)
```

Status: Running | Settings | Reachable at: http://127.0.0.1:1234

READY

llm gemma-3-27b-it-qat | cURL | Idle TTL 60 min | Size 16.43 GB | Eject

Supported endpoints (OpenAI-like)

GET /v1/models

POST /v1/chat/completions

POST /v1/completions

POST /v1/embeddings

Clear logs

Developer Logs


```
[INFO] [gemma-3-27b-it-qat] Generated prediction: {
  "id": "chatcmpl-cxddj3uwg0de1s1jq5k7g9",
  "object": "chat.completion",
  "created": 1746119989,
  "model": "gemma-3-27b-it-qat",
  "choices": [
    {
      "index": 0,
      "logprobs": null,
      "finish_reason": "stop",
      "message": {
        "role": "assistant",
        "content": "There are 7 pages in the HSML Web App:\n\n*   **index.html:** Allows the user to select between Entity, Credential, and Agent, then routes them to the relevant page.\n*   **landing.html:** The first page a user sees, prompting them to login or register.\n*   **login.html:** The login page, which prompts the user to upload a .pem file.\n*   **register.html:** Prompts the user to create a new Person or Organization object.\n*   **result.html:** Shows the formatted JSON result and prompts the user to download the associated key.pem file and HSML JSON file.\n*   **agent.html:** Prompts the user for relevant fields, then sends them to `result.html`.\n*   **credential.html:** Prompts the user for relevant fields, then sends them to `result.html`.\n*   **entity.html:** Prompts the user for relevant fields, then sends them to `result.html`."
      }
    }
  ],
  "usage": {
    "prompt_tokens": 501,
    "completion_tokens": 214,
    "total_tokens": 715
  },
  "stats": {},
  "system_fingerprint": "gemma-3-27b-it-qat"
}
```


Info | Context | Inference | Load

Model Information

Model: lmstudio-community/gemma-3-27B-it-qat-GGUF

File: gemma-3-27B-it-QAT-Q4_0.gguf

Format: GGUF

Quantization: Q4_0

Arch: gemma3 | Vision Enabled

Domain: llm

Size on disk: 16.43 GB

API Usage

This model's API identifier

gemma-3-27b-it-qat

The local server is reachable at this address

http://127.0.0.1:1234

What's Next

- Follow Transitive Data
  - "See Page 4 of Document XYZ"
- AI Chunk Citation
  - AI can leave an in-text citation, telling you exactly what paragraph it received it's info from with a link to the relevant document
- Embeddings Optimization
  - Currently: Nomic
- Prompt Engineering
  - Little changes can mean big performance boosts
  - Anecdote: 50% reduction in hallucination from "My best guess is..."

Lessons Learned

- Test while you develop
  - Digital Twins: Developed the left half only to find out it didn't connect with the right half
- Hallucinations Set You Back
  - AI-generated code can lose/break functionality once the codebase is large enough
- Communicate frequently
  - Large team (seven interns!), primarily remote, requires frequent communication
- Appreciate the moment
  - LA is a once-in-a-lifetime experience, there's still so much I want to do
- Be there for family
  - Ambition must be balanced with presence

Acknowledgements

- My mentors: **Dr. Tom Lu**, and **Dr. Ed Chow**
- My coworkers: Alicia, Jared, Sub, Diego, Gabriel, Niki, Josh, Sydni, Sohee, and Aaron
- Arkansas Space Grant
- My roommates: Jake, Jerome, and Maanav

Thanks everyone!

Questions?