



# SOTA RAG

State-of-the-art AI architectures  
on Cloudflare

By  **LiquidMetal AI**



# Matt Wirth

Engineer at  **LiquidMetal AI**

Background:

- Systems Architecture
- Devops
- Big Data
- Sales Engineering

[matt@liquidmetal.ai](mailto:matt@liquidmetal.ai)

mattw\_98492 on Discord



# What is LiquidMetal AI

SEQUOIA



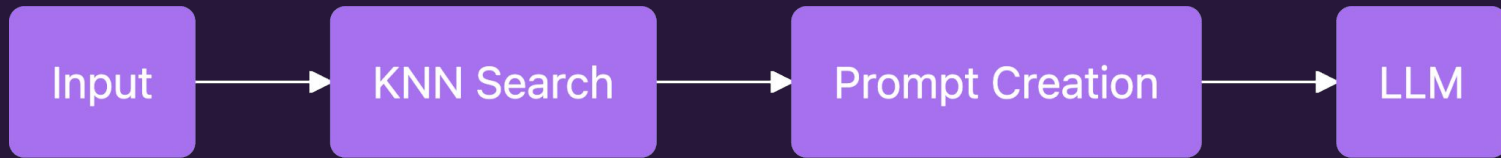
HARPOON



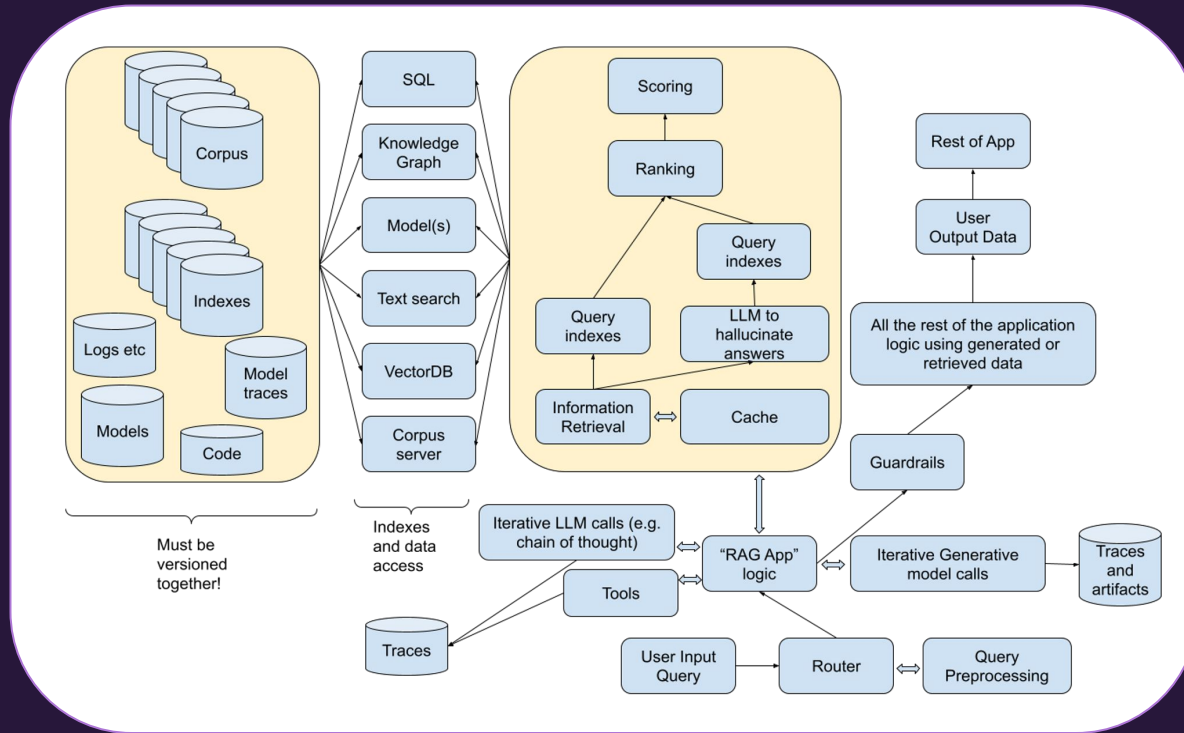
ATLANTIC  
BRIDGE

8VC

# Functional RAG



# State of the Art (SOTA) RAG



# LiquidMetal SOTA RAG

- Indexes PDFs, code, webpages, images, tables, wikipedia and any other data source
- Finds context based on embeddings, entity relationships, topic and topic relationships etc.
- Data management (governance, lineage, metadata)
- Uses a cognitive architecture to model the behaviour

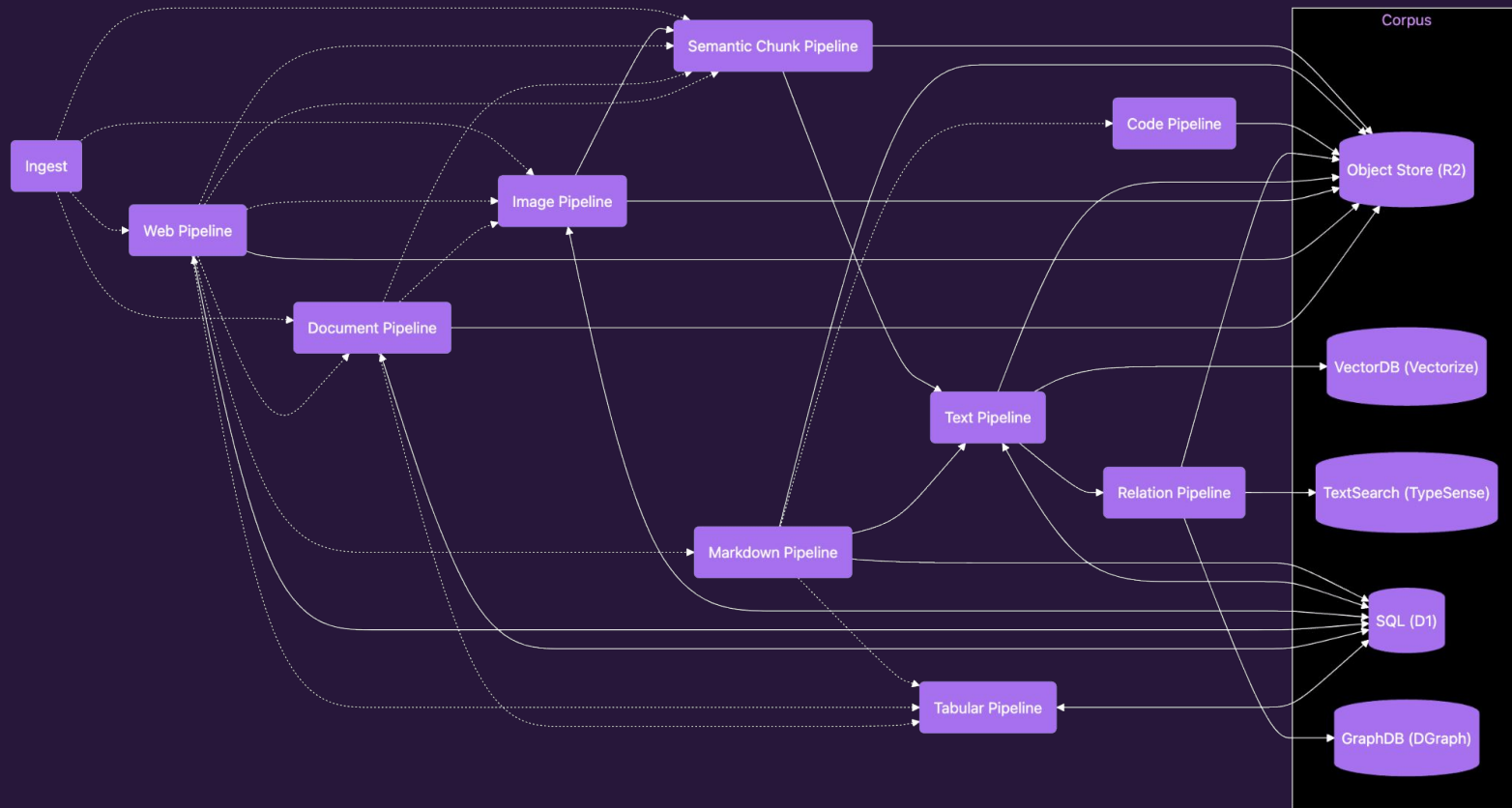


# SOTA RAG Data and Infra



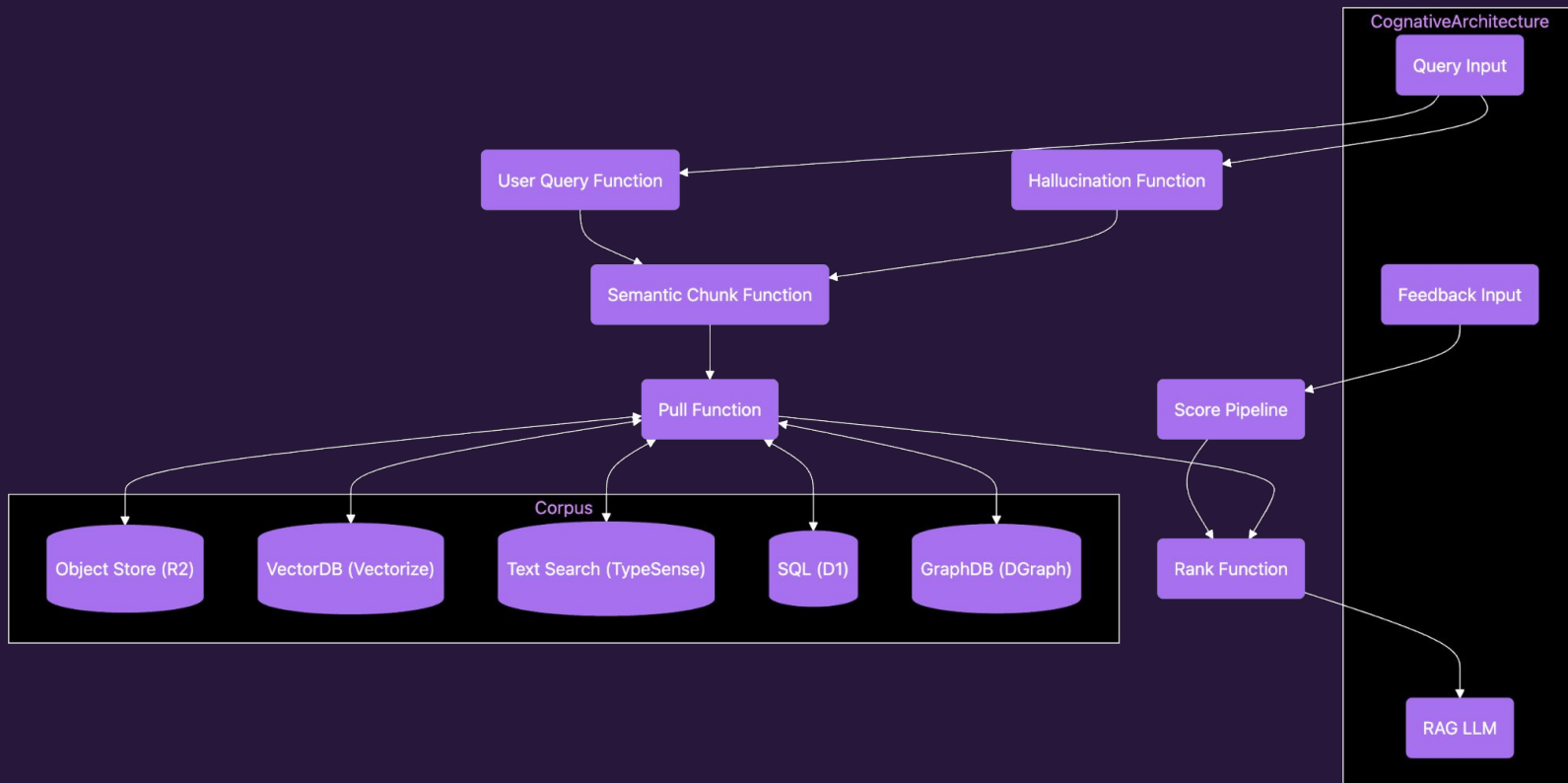
- **Cloudflare workers** - Compute
- **Cloudflare queues** - Pipes
- **Cloudflare durable object** - Cognitive architecture
- ~17 LLM calls (depending on the user's question), split between **Cloudflare Workers AI** and others
- Data storage:
  - Data lake in **R2**
  - **D1** for metadata and tabular data
  - **Vectorize** for vector embeddings
  - **DGraph** for graph relationship search
  - **Typesense** for text search

# Indexer

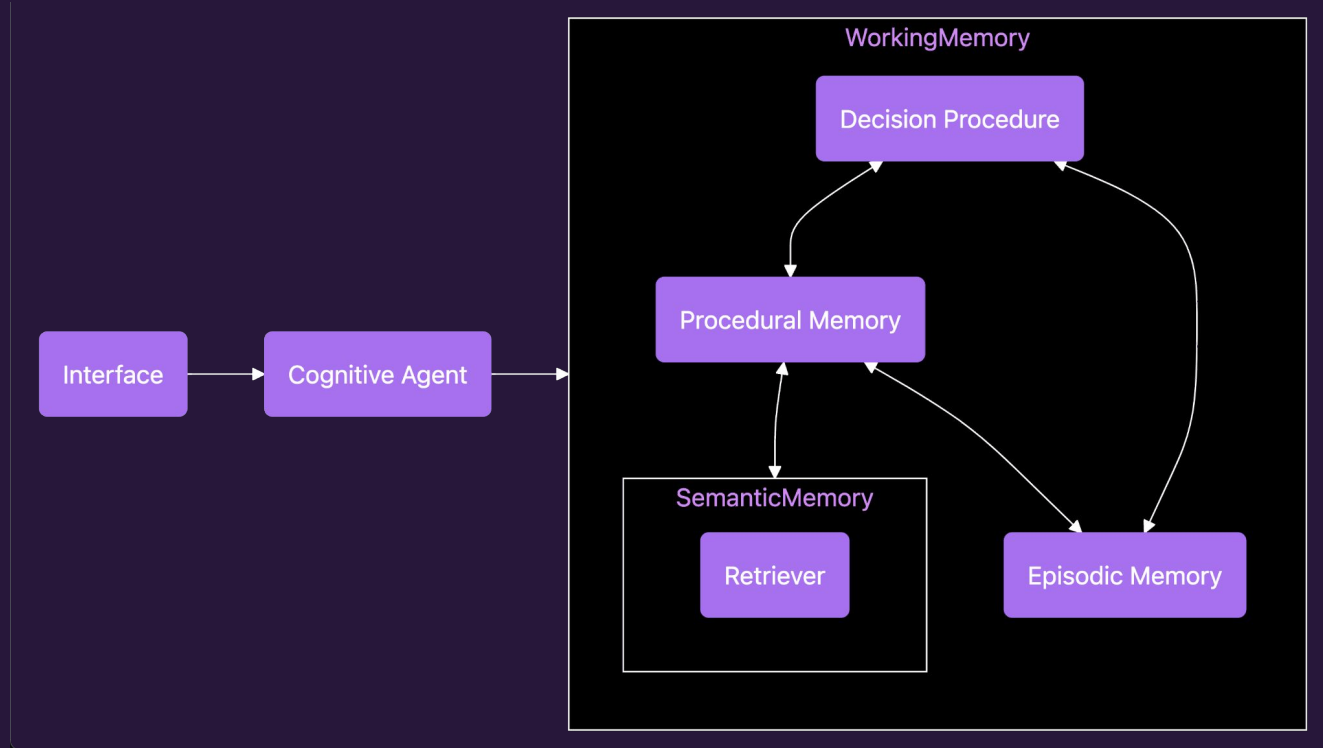




# Retriever



# Cognitive Architecture





# Meet Edgy!

Demo



# Day 2 challenges

- Managing data generated at the edge
  - Data governance
  - Data discovery
  - Data lineage and metadata
- Managing the application lifecycle and resources
  - Stand up, tear down large web of resources
  - Version and manage both data and code



# LiquidMetal Raindrop Build

- A LiquidMetal solution, built on Cloudflare
- Declarative build, deploy, destroy
- Version and manage large deployments and the resources/data they interact with



**Check it out!**





# Thank You!

Matt Wirth

[matt@liquidmetal.ai](mailto:matt@liquidmetal.ai)  
mattw\_98492 on Discord

