

# Agentic AI Workshop

September 2025

[abu.desai@readytensor.com](mailto:abu.desai@readytensor.com)



---

# About me

+



## **Abhyuday “Abu” Desai, Ph.D.**

- Founder, CEO – Ready Tensor, Inc., an AI/ML start-up
- 20+ years in analytics, data science roles
- Led R&D and consulting teams in the analytics space
- Collaborated with teams at many of the largest corporations in the US and internationally
- Ph.D. from Texas Tech University in Operations Research



# Agentic AI Workshop 1: What We'll Learn Today

---

## 1. Basic LLM Interactions

Text generation, image creation, audio synthesis

## 2. Embedding & Similarity

- Generate embeddings for documents
- Calculate cosine similarity
- Find most relevant documents to queries

## 3. Complete RAG System

- Ingest documents into ChromaDB
- Build semantic search pipeline
- Create RAG-powered chatbot

# What we will use - OpenAI Python SDK

---



## OpenAI Python SDK

- Official Python library for OpenAI's APIs
- **Chat Completions** - GPT-4, GPT-3.5 for text generation
- **Embeddings** - Convert text to vectors (text-embedding-3-small)
- **Images** - DALL-E for text-to-image generation
- **Audio** - Text-to-speech and speech-to-text
- **Streaming** - Real-time response generation

# What we will use - ChromaDB

---



**ChromaDB**

- Open-source vector database for AI applications
- Stores embeddings and enables fast similarity search
- In-memory or **persistent** - Choose based on your needs
- Metadata filtering - Search with both vectors and metadata
- Collections - Organize documents into groups
- Query interface - Simple API for similarity search

# What we will use - Gradio

---



- **Open-source** Python library from HuggingFace for building ML demos and UIs
- Create web interfaces with just **a few lines of code**
- **ChatInterface** - Pre-built chat UI with history management
- **Interface** - Custom input/output components (text, images, audio)
- **Sharing** - Generate public links or deploy to Hugging Face Spaces

# Setup for the Coding Exercises

---

## Follow these steps:

1. Clone repo:
  - <https://github.com/readytensor/agentic-ai-workshop-1>
2. Create .env file in root of your project
  - Use ``.env.example`` file as reference
3. Setup your OpenAI API Key.
  - Don't have one? Ask program coordinator for a temporary key to use during workshop
  - This temp key will be deleted after the workshop
4. Create virtual env
5. Install dependencies
6. Open jupyter notebook
7. Navigate to the 3 notebook files in ``../notebooks``

# What is Agentic AI?

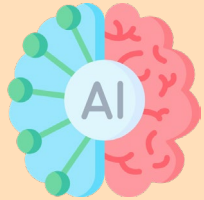
Agentic AI moves beyond simple input-output to create systems that can **reason, plan, and act independently**.

- AI systems that can **act autonomously to achieve goals**
- Systems that can **use tools, make decisions, and take actions**



# LLM Workflows vs Agentic AI

## LLM Workflows



- Simple input-output
- No real autonomy
- Static, predictable path

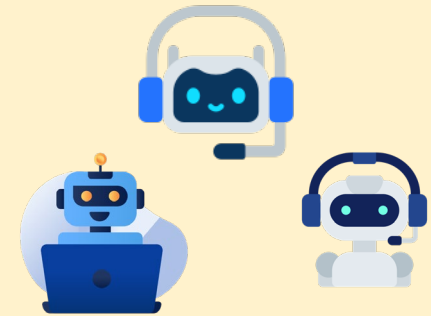
## Agentic AI

### Single Agent



- Can use tools and APIs
- Makes decisions based on context
- Can take actions to achieve goals

### Multi-Agents

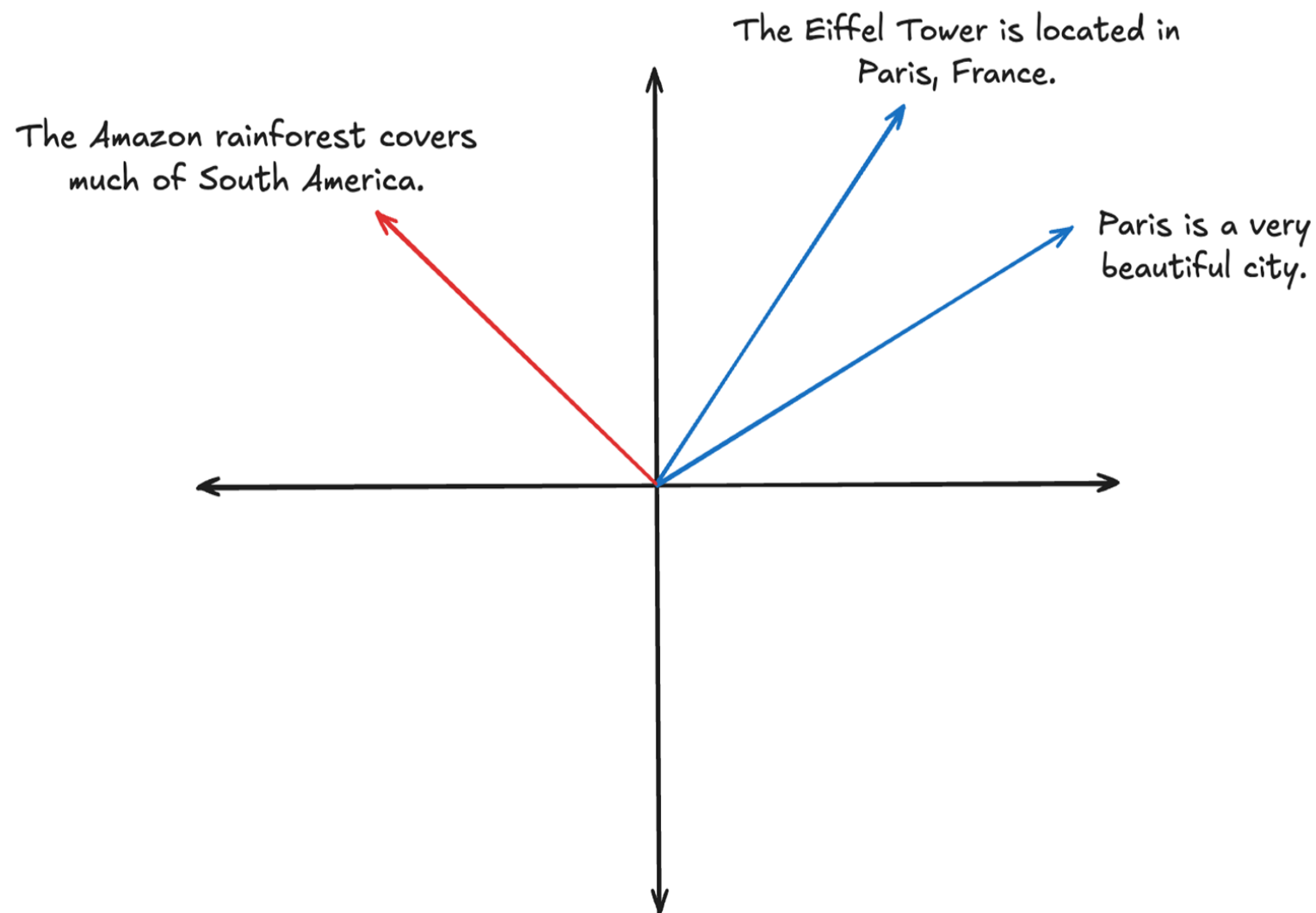


- Multiple specialized agents
- Agents communicate and coordinate
- Complex task decomposition

# What are Embeddings?

Numerical representations of text as vectors

- Similar texts have similar embeddings
- High-dimensional (e.g., 1536 dimensions)
- Capture semantic meaning, not just words



# Similarity Metrics

---

## Cosine Similarity

- Measures angle between vectors
- Range: -1 to 1
- Higher is better
- Most common

## Euclidean Distance

- Straight line distance between 2 vectors
- Range: 0 to infinity
- Lower is better

## Cosine Distance

- $1 - \text{Cosine Similarity}$
- Range: 0 to 2
- Lower is better

# Retrieval Augmented Generation (RAG)

---

**RAG is a technique that combines information retrieval with text generation to provide more accurate and contextual responses.**

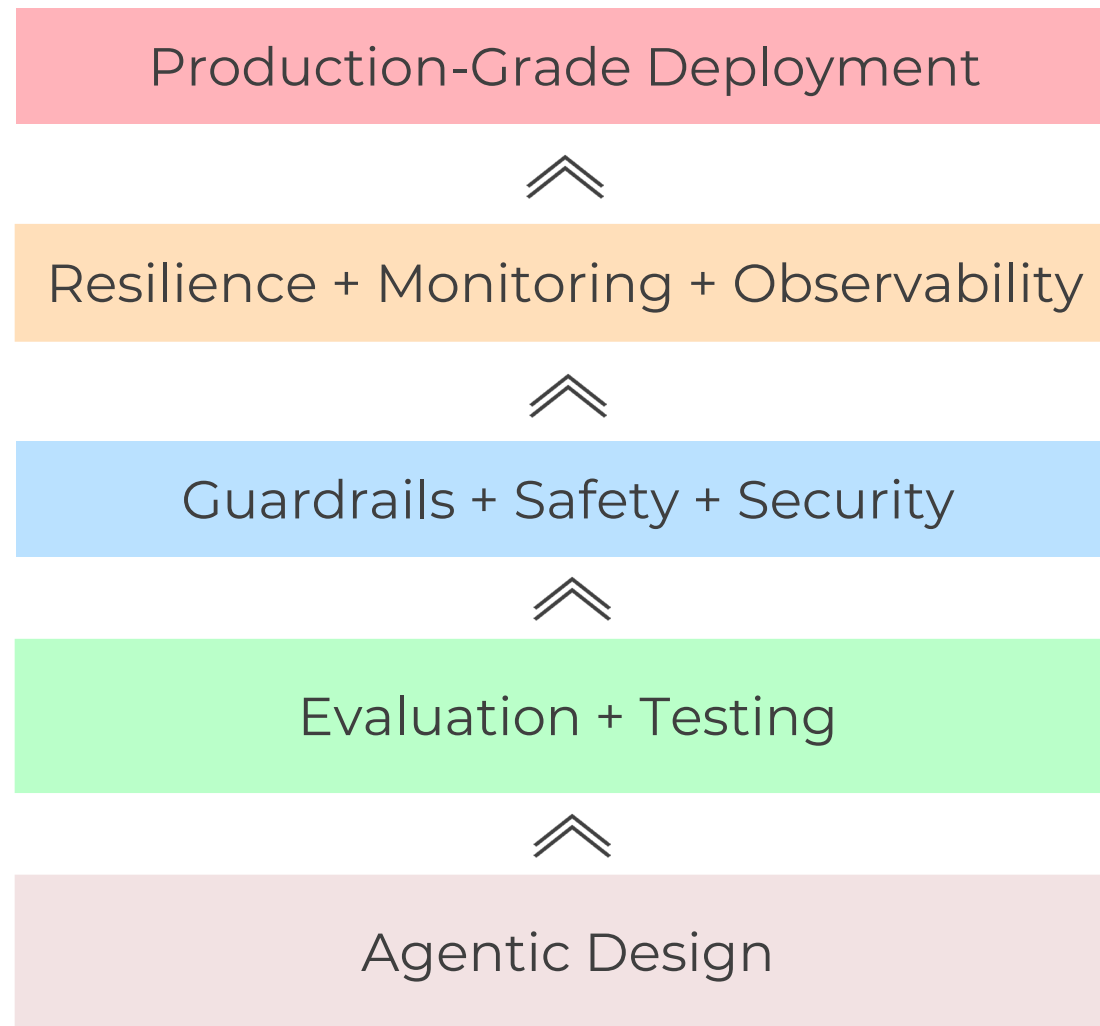
- Retrieves relevant documents using embeddings
- Uses retrieved context to generate better answers
- Combines search + generation for grounded responses

How it uses embeddings:

- Convert documents to embeddings → store in vector database
- Convert query to embedding → find similar documents
- Use retrieved documents as context for LLM

# Agentic Design is Just the Beginning!

---





## Agentic AI Developer Certification

# BECOME AN AGENTIC AI DEVELOPER



### MODULE 1

Prompt engineering  
Reasoning  
Tool use  
Memory  
Vector DBs & RAG

### MODULE 2

Multi-agent design  
Architectures  
Shared memory  
MCP  
Evaluations

### MODULE 3

Guardrails  
Safety  
APIs / Inference  
Monitoring  
Observability

Enroll for free: <https://www.readytensor.ai/agentic-ai-cert>



## **The Global Hub for AI Innovation and Discovery**

---

Abu Desai, Ph.D. – Founder, CEO

[Ready Tensor, Inc.](#)

[abu.desai@readytensor.com](mailto:abu.desai@readytensor.com)

+1 (858) 245-7739