Azure Developer Community

Tamil Nadu

What we do?

Meetups

Network of Beginners, Professionals & Experts

Azure Cert Study Jam

Tech-talks / workshops
In Colleges

#Azure Developer Community

RAG Is Not About Vector Databases - It's About Retrieval Thinking

Why Most GenAI Proof-of-Concepts Fail in Production

Azure Developer Community

Tamil Nadu

LEARN WITH SARVESH

COURSES. CLARITY. CONFIDENCE.

Self intro

If vector databases solved RAG... why do most RAG apps still fail in production?

Why LLMs Alone Are Not Enough

- Trained on general internet data
- No access to your private documents
- Hallucinates when uncertain

What is RAG?

RAG = Retrieval + Generation

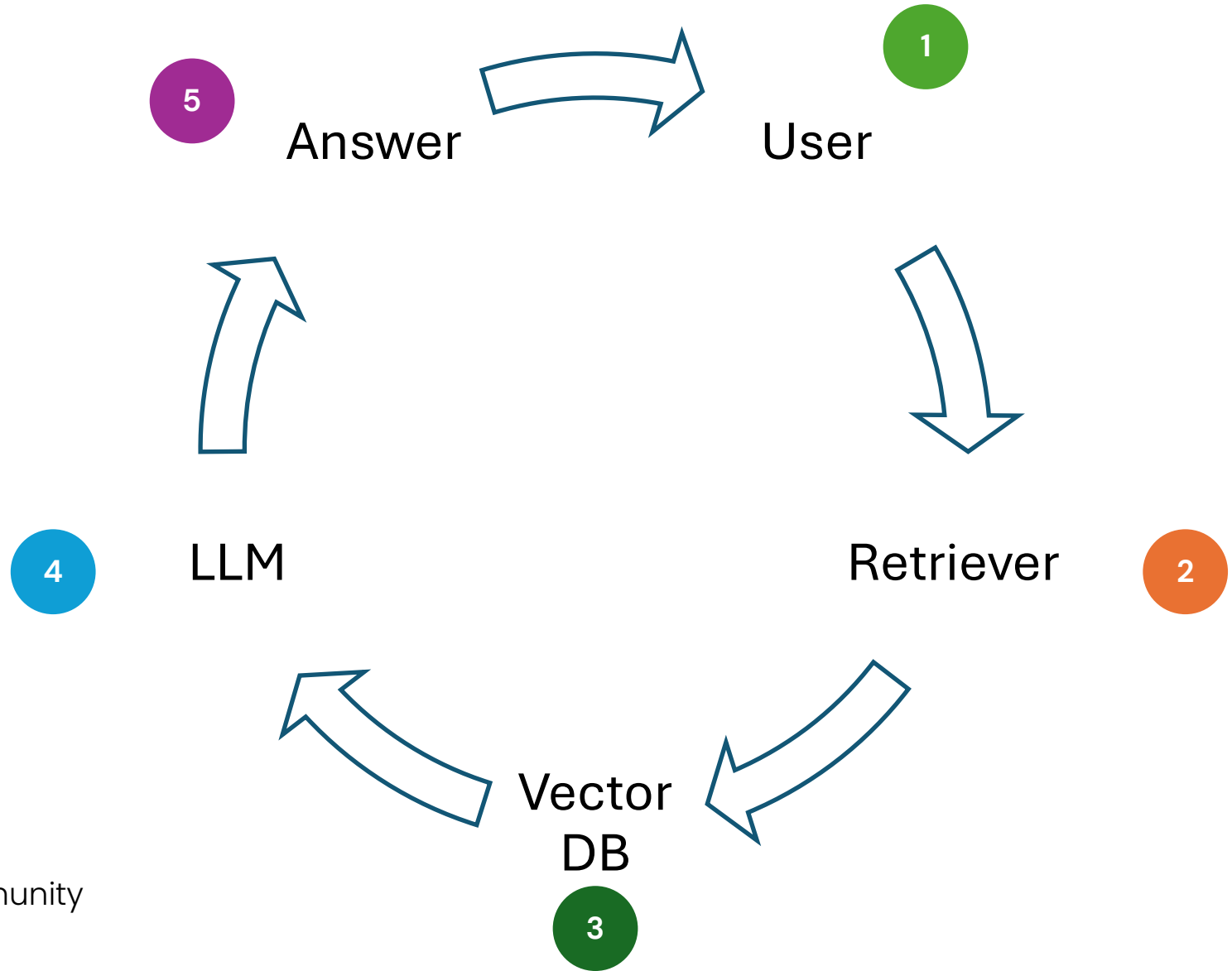
Retrieve relevant information

Generate answer using that information

Analogy:

Open-book exam for the model

Typical RAG Architecture



The Problem

The Myth

“Just add a vector database”

“Better embeddings = better answers”

“More documents = better results”

Why PoCs Look Impressive

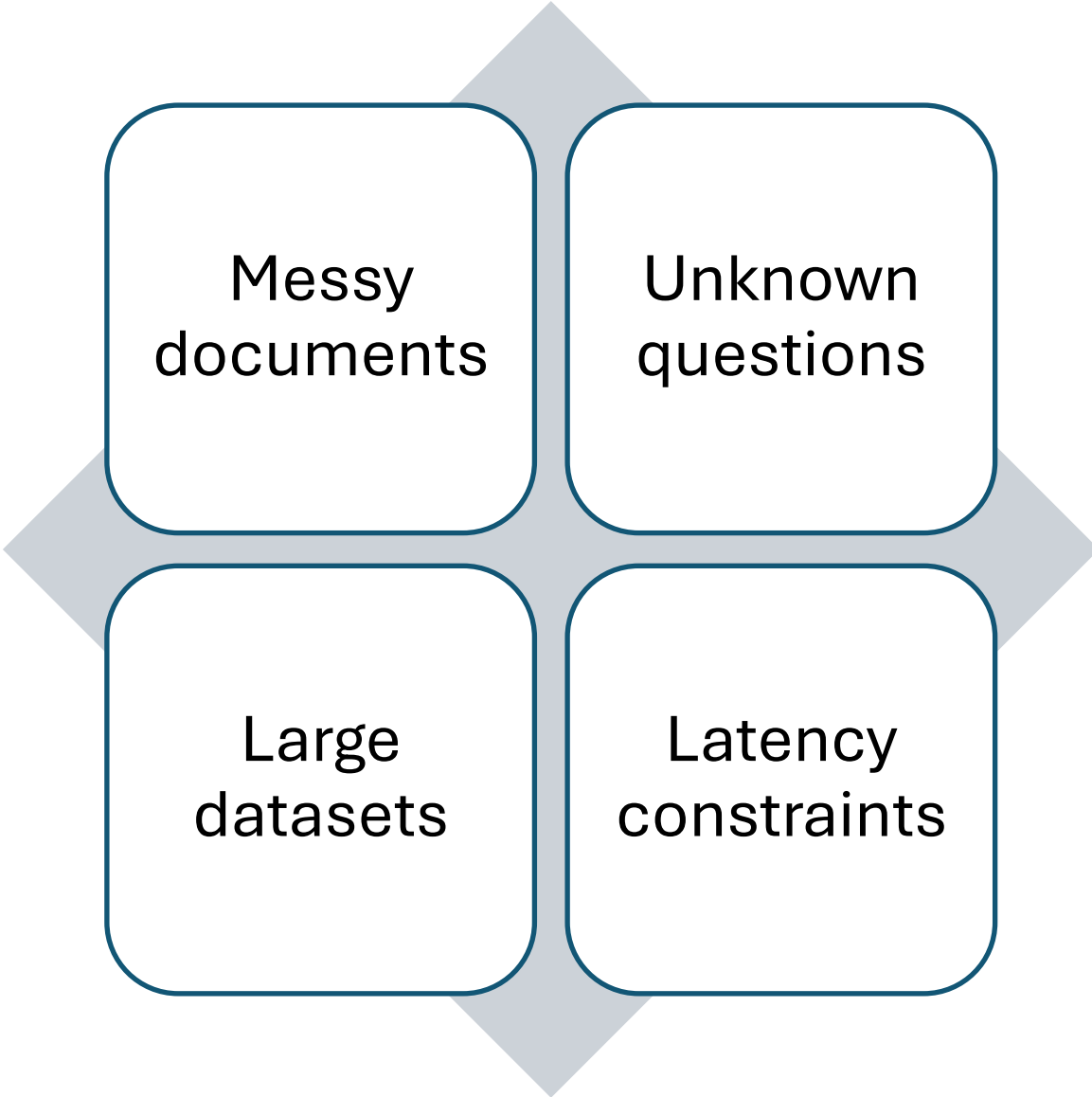
Small dataset

Known questions

Clean documents

Manual testing

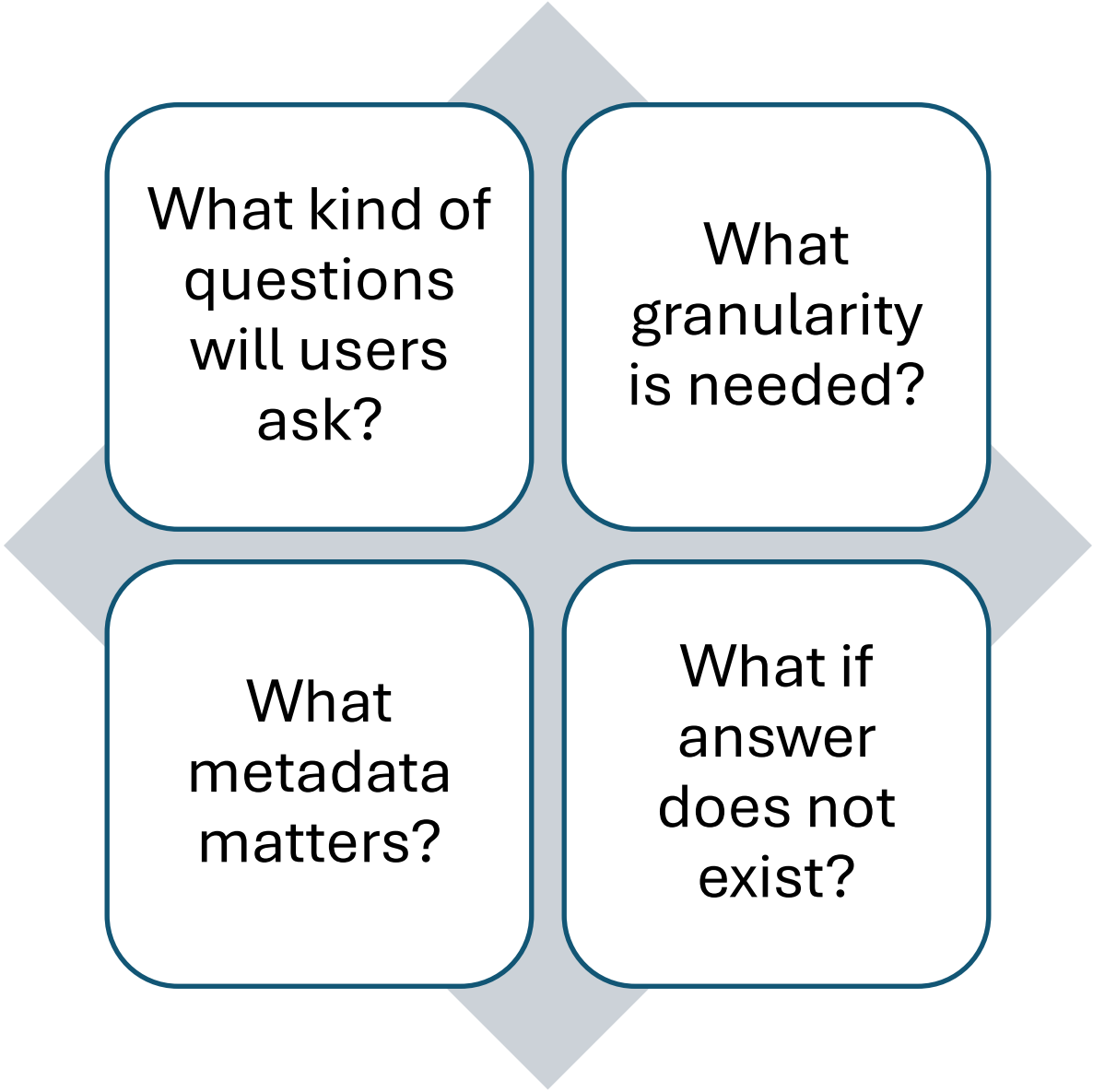
Reality in Production



Retrieval Thinking

Retrieval Thinking

Before storing documents, ask:



Chunking Is a Design Decision

Bad:

- Large 2000-token chunks

Better:

- Logical sections
- Policy-based segmentation
- FAQ-style chunks

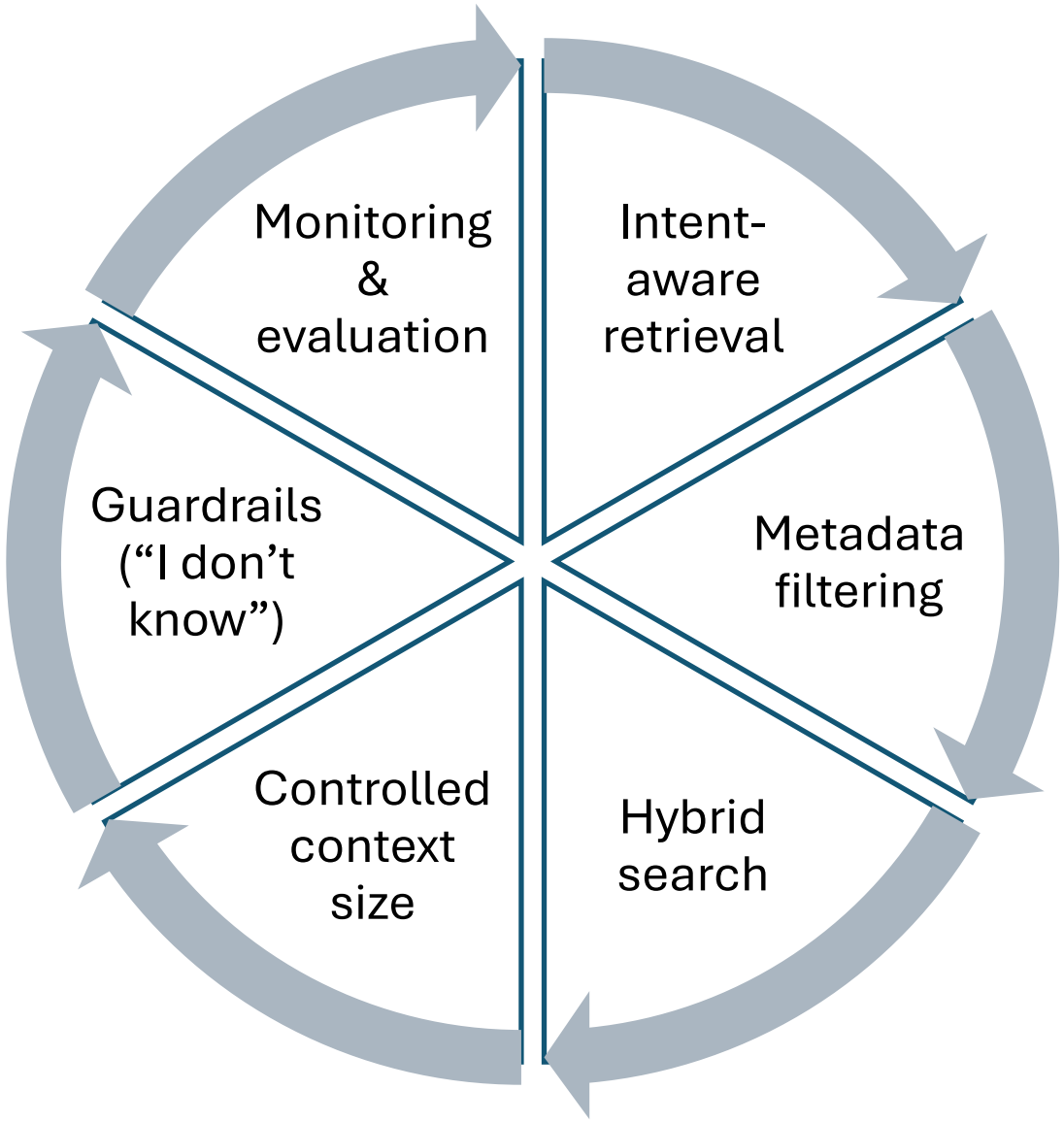
The Demo

After Basic RAG Failure

After Improved Retrieval

Production-Ready RAG

Production-Ready RAG Needs



RAG Patterns in Production

Hybrid RAG

Vector search + Keyword search
Better precision

Multi-Stage Retrieval

Retrieve broadly
Re-rank
Then generate

Agentic RAG

Planner decides which retriever to use
Tool-based retrieval

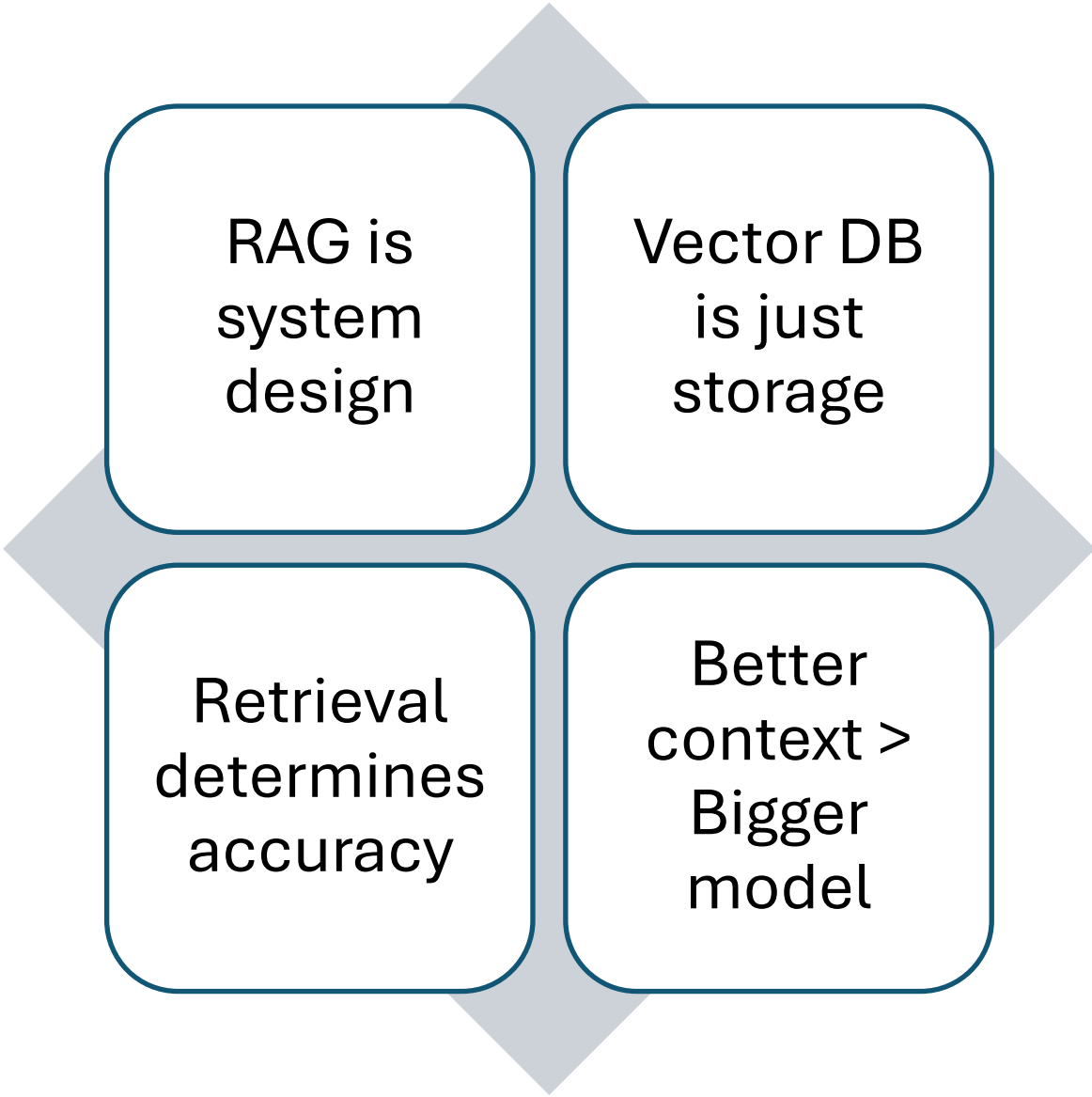
RAG + Evaluation

Automatic answer scoring
Feedback loops
Monitoring hallucinationas

Taking RAG to Production

- Latency constraints
- Cost control (token usage)
- Caching strategies
- Embedding refresh pipelines
- Document update handling
- Access control (who can see what)

Final Takeaways



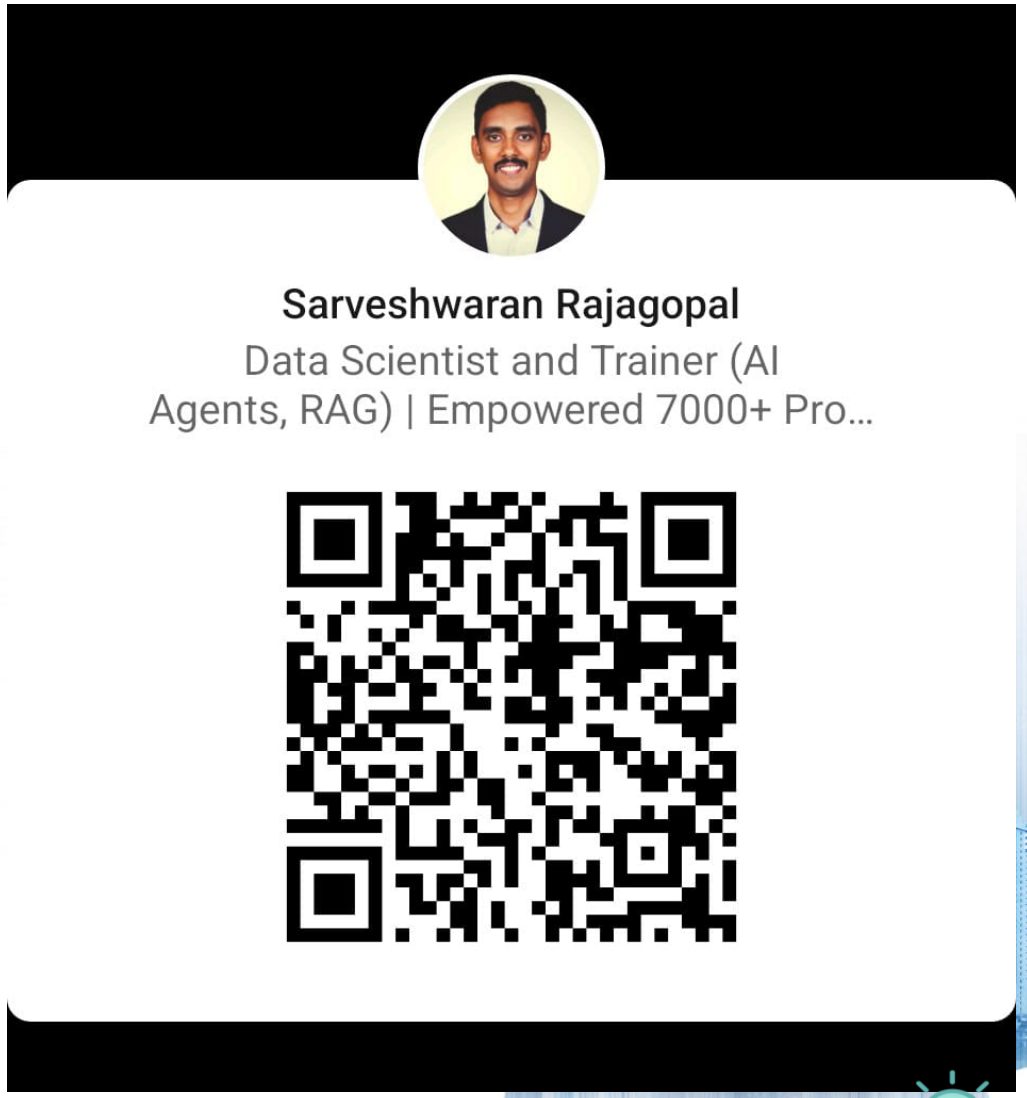
Cohort

https://www.learnwithsarvesh.com/

#Azure Developer Community

Open for Q&A

@WITHSARVESH.AI



#Azure Developer Community

Thank you!!

Thank You!