National Science Foundation
Where Discoveries Begin

Knowledge Base System using LLM-Driven Retrieval

Cheng Chen (UAH)[1], Hwan Yoon (UA)[2]

[1]Industrial & Systems Engineering and Engineering Management Department
[2]Mechanical Engineering Department

IUCRC
Industry-University Research Partnership

Capturing Expertise Before It's Gone:

September 17, 2025 • 11 min read

The Hidden $92 Billion Crisis: How Manufacturing's Knowledge Problem Blocks Growth (And How to Solve It)

Manufacturing loses $92 billion annually to knowledge management failures. Learn how manufacturers can achieve 216% ROI by solving their tribal knowledge crisis with AI-powered systems.

- Companies implementing comprehensive knowledge management systems report a 216% ROI with ~6-month payback

INNOVATION

Tech Retirement Crisis: Vulnerable Industries And What They Must Do

By Expert Panel®, Forbes Councils Member.
for Forbes Technology Council, COUNCIL POST | Membership (fee-based)

Published Jul 24, 2025 at 08:15am EDT

- Retiring engineers are creating knowledge gaps, particularly around legacy SCADA and bespoke PLC systems unfamiliar to younger staff.

McKinsey & Company

The social economy: Unlocking value and productivity through social technologies

By using social technologies, companies can raise the productivity of knowledge workers by 20 to 25 percent.

Jul 1, 2012

- The most powerful applications of social technologies in the global economy are largely untapped. By using social technologies, companies can raise the productivity of knowledge workers by 20 to 25 percent.

Project Goals

- **Define and prioritize target knowledge assets**
  - Conduct stakeholder analysis with engineers, operators, maintenance, and management
  - Identify which knowledge should be captured first:
    - Legacy PLC / SCADA documentation
    - SOPs, work instructions, and maintenance manuals
    - Historical incident reports, root-cause analyses, and shift logs
    - Tacit expert knowledge from senior engineers
  - Focus initial scope on high-impact, frequently used, and hard-to-replace knowledge

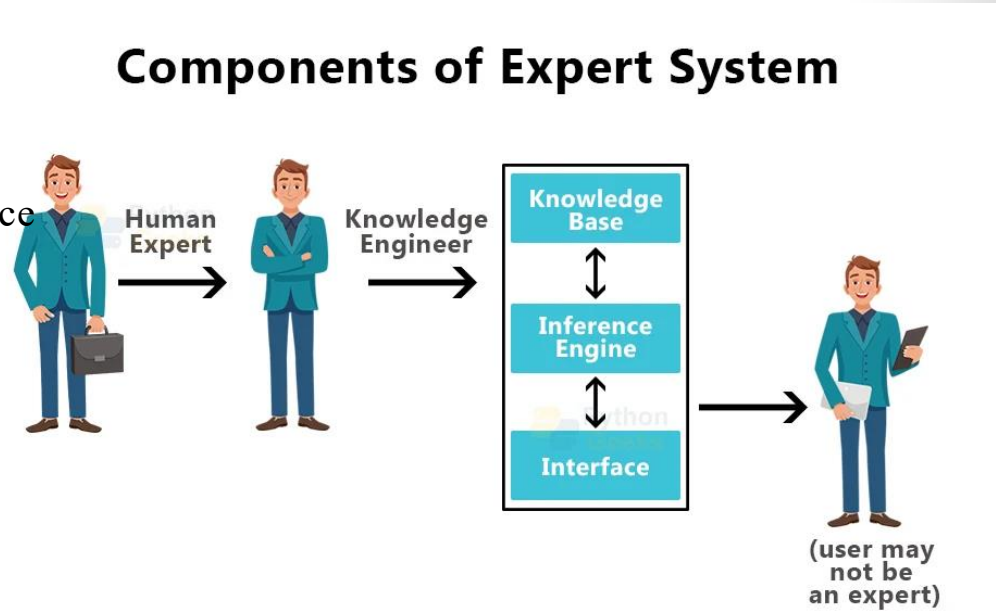


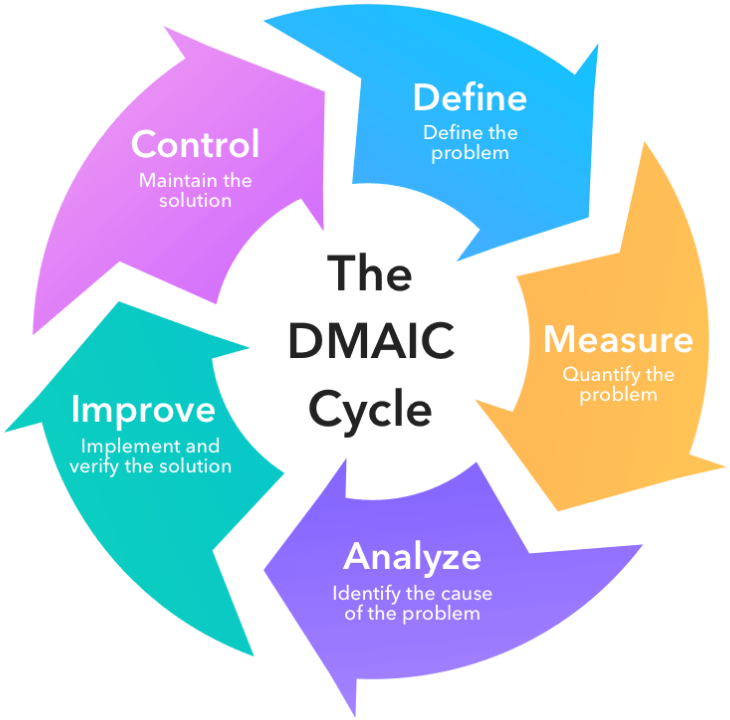
Image source

- **Design and validate a safe and reliable LLM-RAG architecture**
  - Ensure responses are grounded in approved internal sources
  - Implement citation, traceability, and access control
  - Minimize hallucinations and ensure auditability for engineering and compliance needs

- **Deploy a production pilot and establish operating**
  - Run user workflows with engineers, operators, and maintenance and iterate from feedback
  - Track usage, answer quality, and time saved

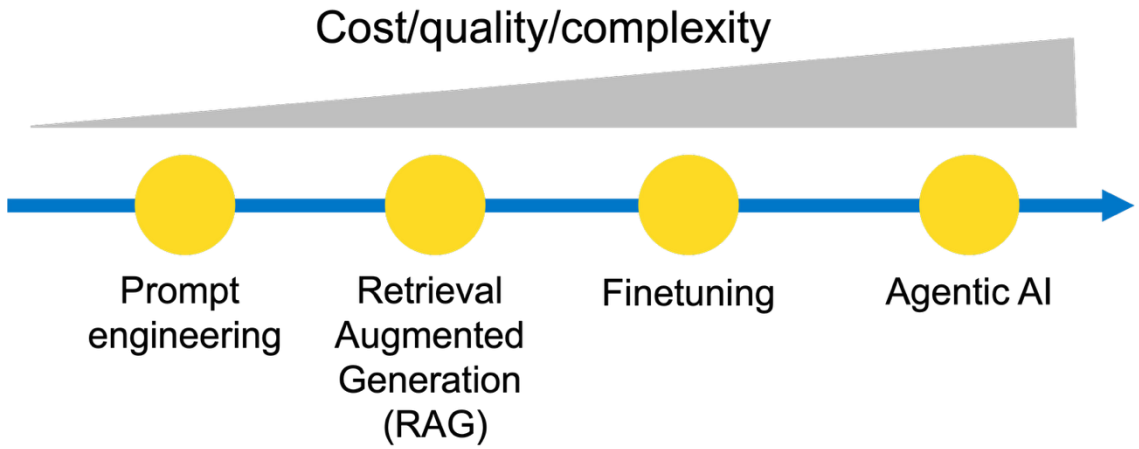
Objectives

- Capture and preserve institutional knowledge
- Reduce production downtime and errors
- Improve knowledge accessibility and reuse
- Unify fragmented data sources
- Enhance decision support with LLM-RAG
- Improve operational efficiency and ROI

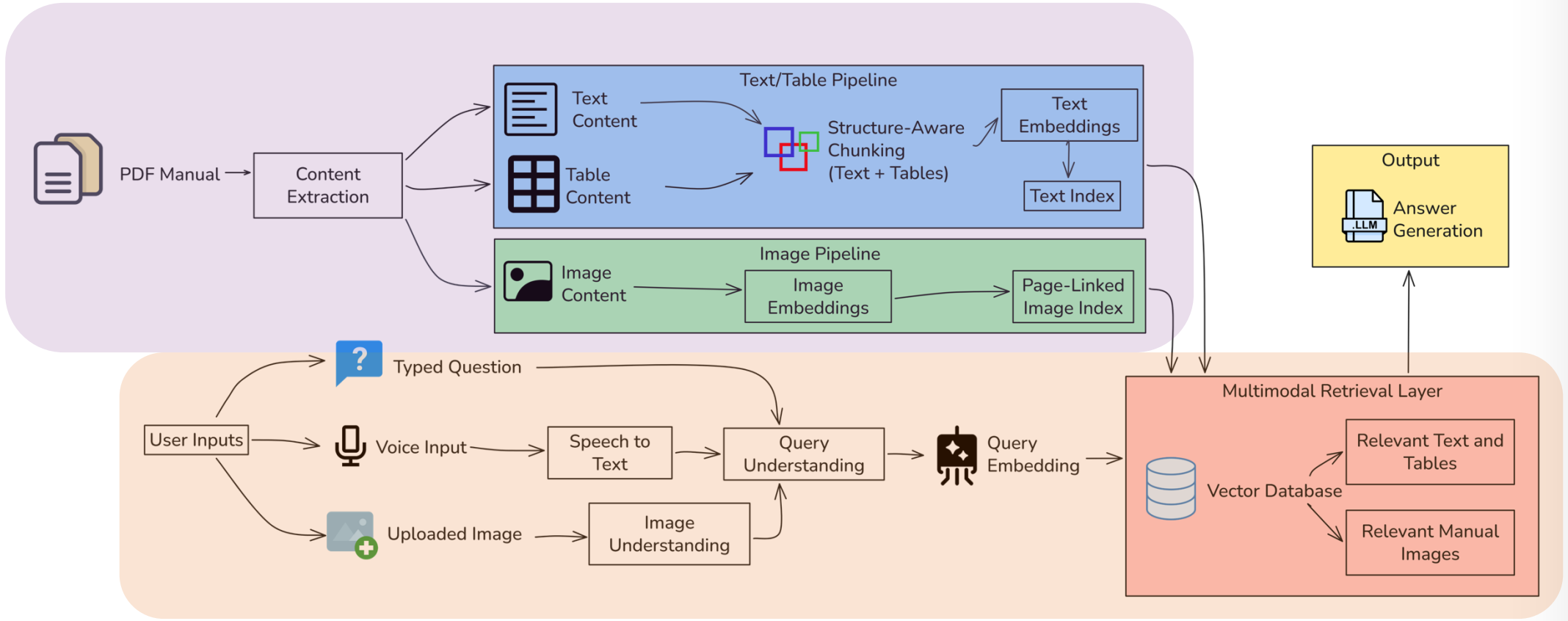


Approach / End-to-End LLM Integration

- **Data readiness and quality assessment**
  - Ground-truth set creation
  - Governance rules
- **LLM design and retrieval research**
  - Chunking experiments
  - Hybrid retrieval
  - Knowledge graph
  - Agentic AI
- **Prompting, grounding, and answer policies**
  - Answer format
  - Causal reasoning/Uncertainty-aware reasoning



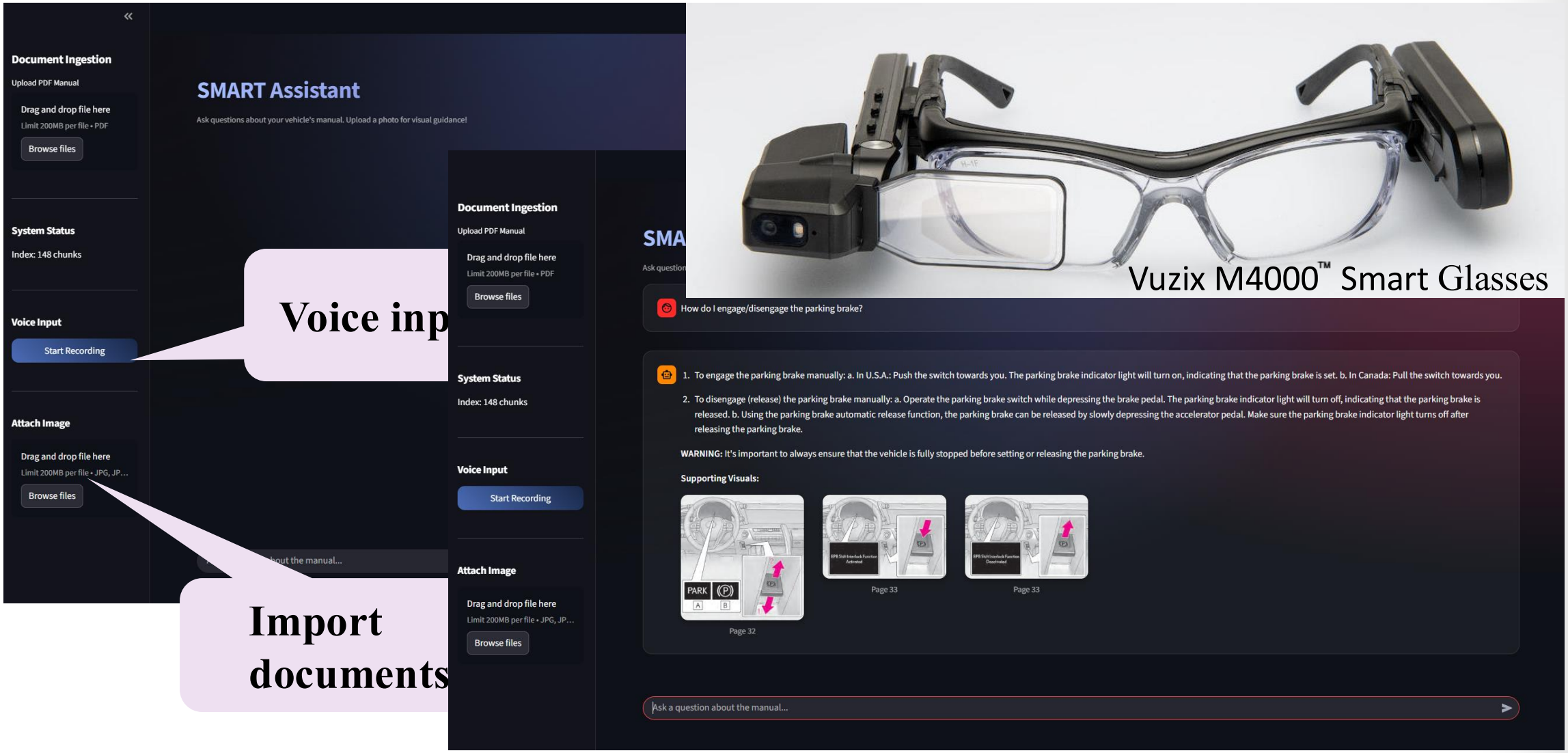
Approach / Research Methods



Architecture of a Multimodal Retrieval-Augmented Generation (RAG) Model

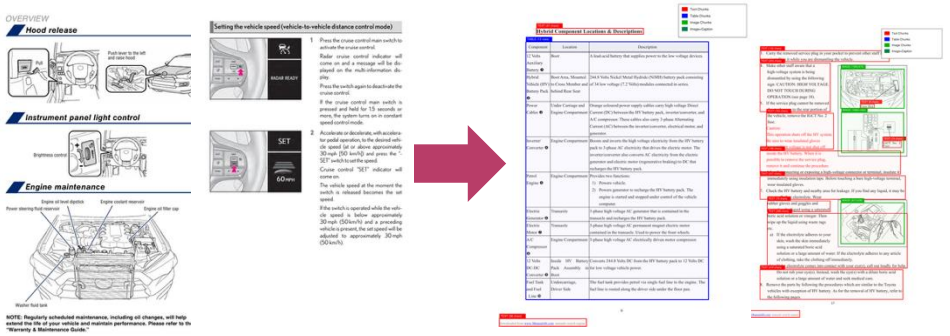
- **Explainable and trustworthy**
- **Safe and grounded in facts**

Outcomes / Deliverables



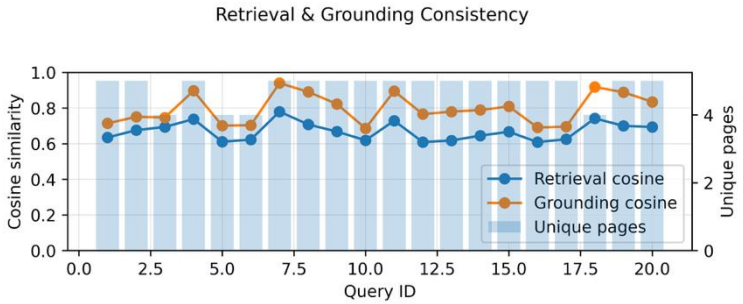
Expected Impacts

- Intelligent chunking for real-time explainability visualization, supporting segment documents by semantic boundaries & relevance scores



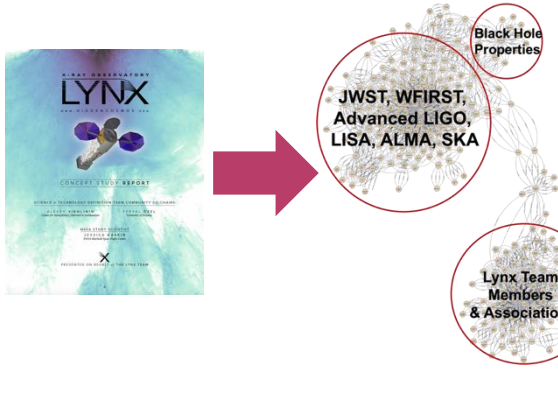
Validation Example: PDF Page and Its RAG Chunk

- Benchmark model performance: accuracy vs. inference speed: Deliver a validated model, dashboard, and concise deployment rationale so teams can trade accuracy, cost, and grounding with confidence.



Retrieval QA Benchmark

- Analytics-driven decision-support - use a knowledge graph to interpret model outputs and drive action



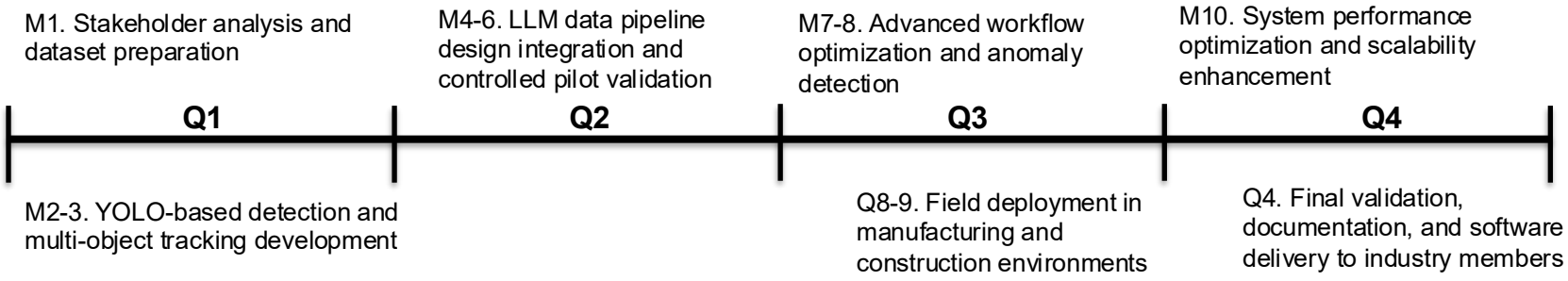
NASA Lynx Project

Project Administration

Budget

Item	Year 1
Salaries and Stipends	
Faculty – H. Yoon (0.1/1.0)	3000
Faculty – C. Chen (0.1/1.0)	3000
Graduate Research Assistant	31,200
Fringe Benefits	
Faculty and Staff (32%)	1,920
Student Research Assistants	1,664
Other Direct Costs	
Travel	3,416
Total Direct Costs	44,200
Indirect Costs (10% Request)	4,420
Items Not Charged F&A	
GRA Tuition	11,380
Budget Totals	**$60,000**

Duration: 12 Months



National Science Foundation
Where Discoveries Begin

Questions?

IUCRC
Industry-University Research Partnership