

Knowledge Base System using LLM-Driven Retrieval

Cheng Chen (UAH)¹, Hwan Yoon (UA)²

**¹Industrial & Systems Engineering and Engineering
Management Department**

²Mechanical Engineering Department

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.

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



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

Capturing Expertise Before It's Gone:

- Companies implementing comprehensive knowledge management systems report a **216% ROI** with **~6-month** payback
- Retiring engineers are creating knowledge gaps, particularly around legacy SCADA and bespoke PLC systems unfamiliar to younger staff.
- 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**.



- **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

• **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

Project Goals

Components of Expert System

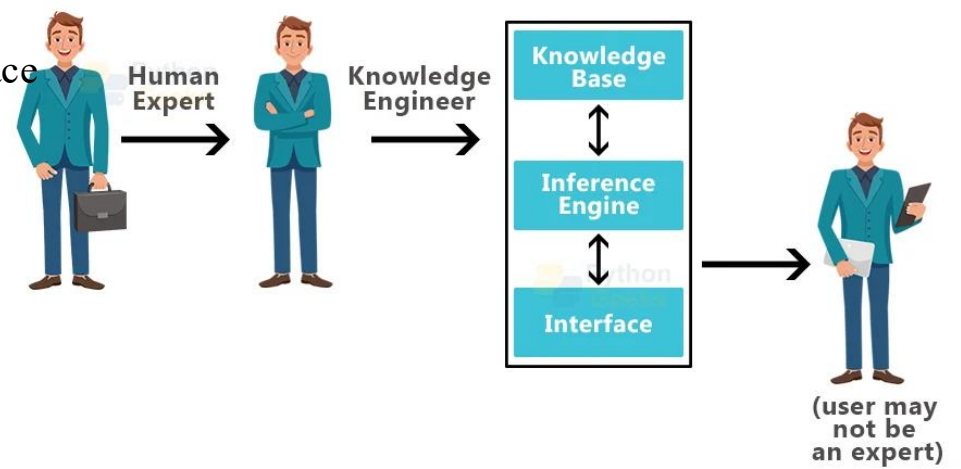
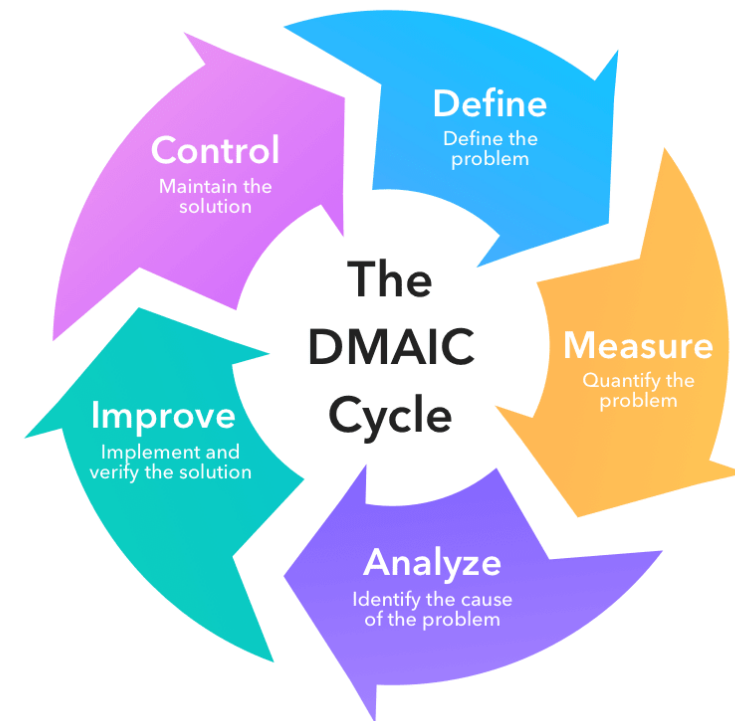


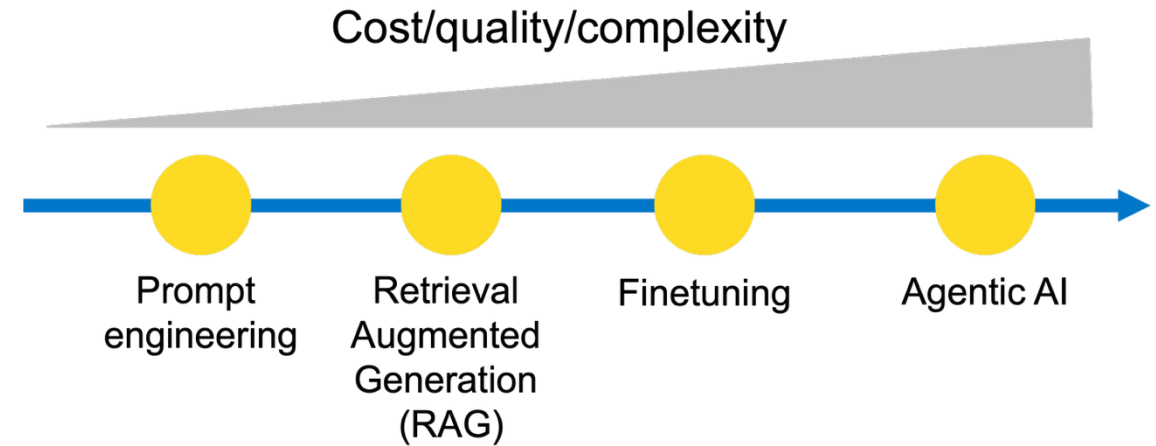
Image source

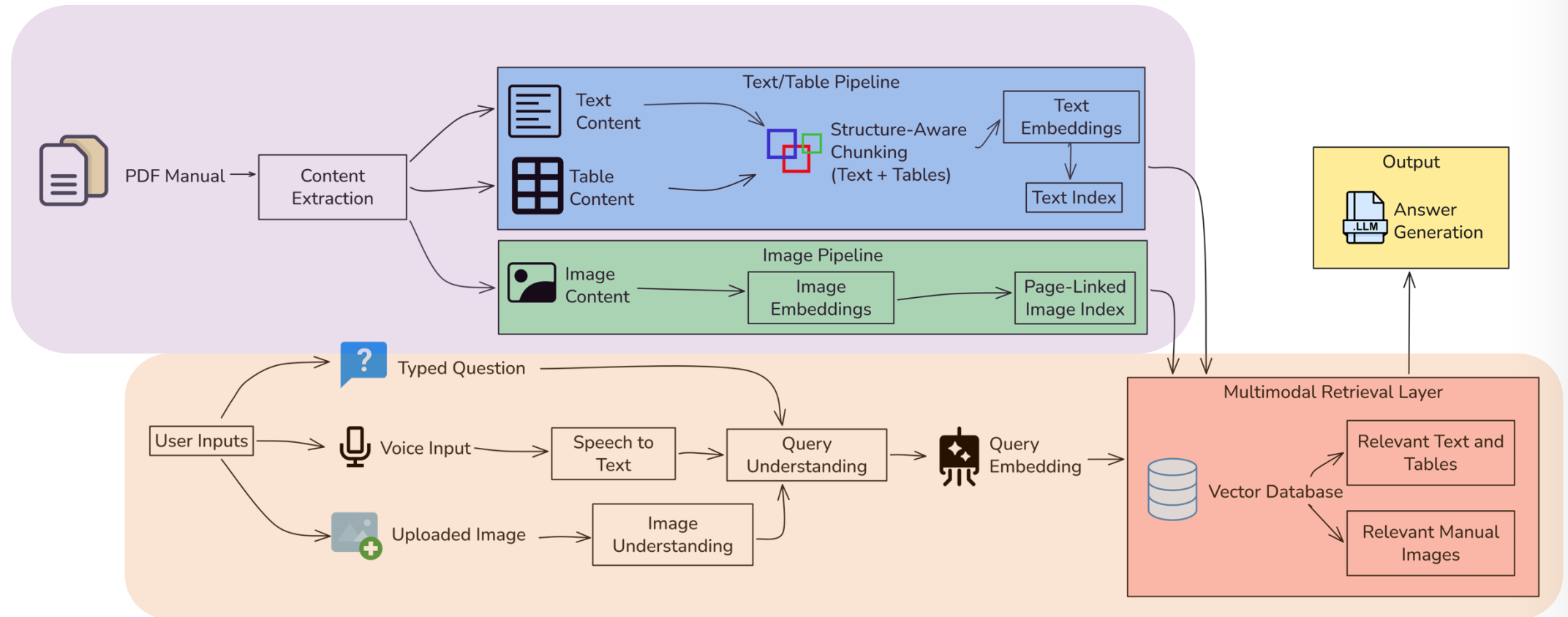
- Capture and preserve institutional knowledge
- Reduce production downtime and errors
- Improve knowledge accessibility and reuse
- Unify fragmented data sources
- Enhance decision support with LLM-RAG
- Improve operational efficiency and ROI



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

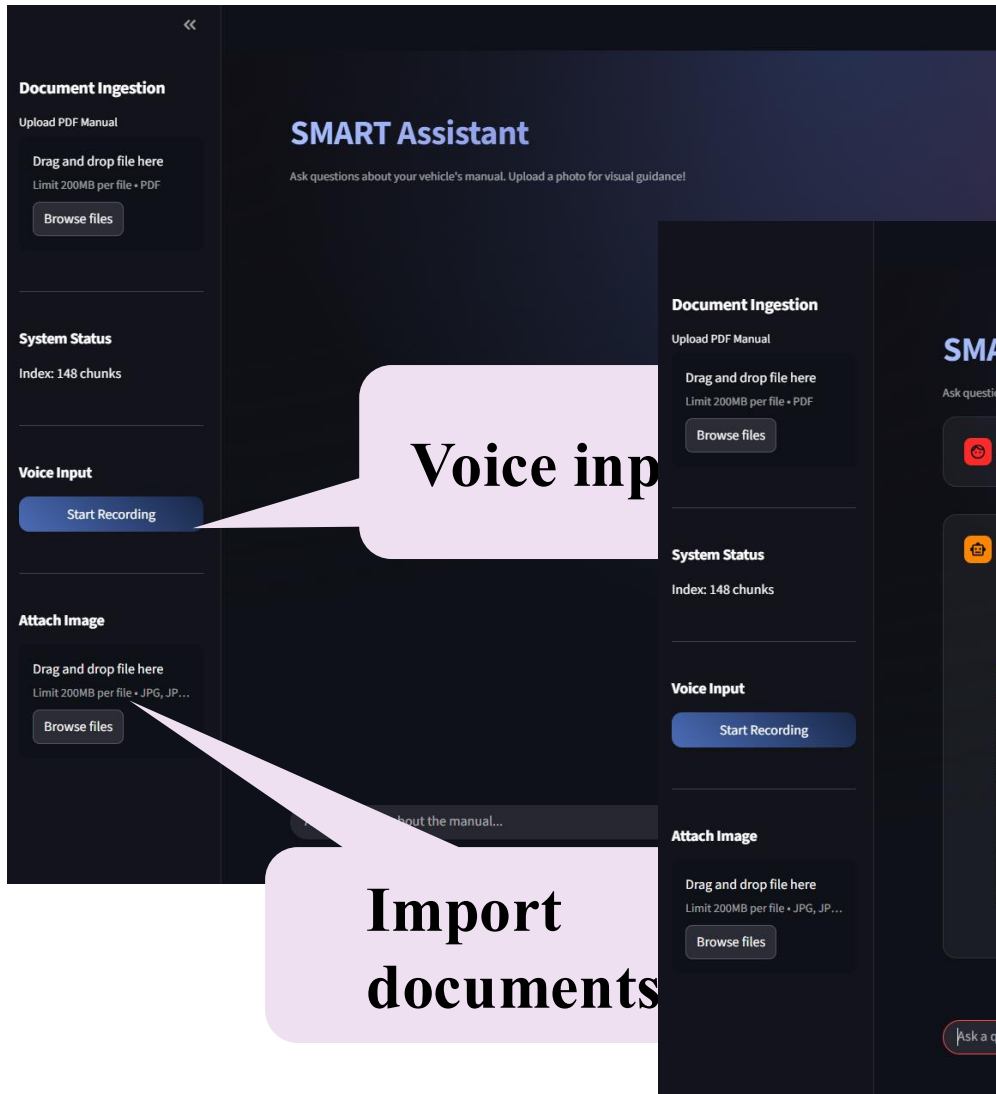




Architecture of a Multimodal Retrieval-Augmented Generation (RAG) Model

- Explainable and trustworthy
- Safe and grounded in facts

Outcomes / Deliverables



Document Ingestion
Upload PDF Manual
Drag and drop file here
Limit 200MB per file • PDF
Browse files

System Status
Index: 148 chunks

Voice Input
Start Recording

Attach Image
Drag and drop file here
Limit 200MB per file • JPG, JP...
Browse files

SMART Assistant
Ask questions about your vehicle's manual. Upload a photo for visual guidance!

Document Ingestion
Upload PDF Manual
Drag and drop file here
Limit 200MB per file • PDF
Browse files


System Status
Index: 148 chunks

Voice Input
Start Recording

Attach Image
Drag and drop file here
Limit 200MB per file • JPG, JP...
Browse files

Voice input

Import documents



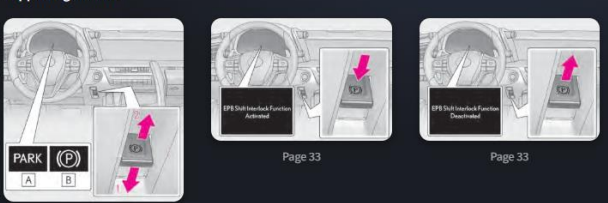
Vuzix M4000™ Smart Glasses

How do I engage/disengage the parking brake?

- To engage the parking brake manually: a. In U.S.A.: Push the switch towards you. The parking brake indicator light will turn on, indicating that the parking brake is set. b. In Canada: Pull the switch towards you.
- To disengage (release) the parking brake manually: a. Operate the parking brake switch while depressing the brake pedal. The parking brake indicator light will turn off, indicating that the parking brake is released. b. Using the parking brake automatic release function, the parking brake can be released by slowly depressing the accelerator pedal. Make sure the parking brake indicator light turns off after releasing the parking brake.

WARNING: It's important to always ensure that the vehicle is fully stopped before setting or releasing the parking brake.

Supporting Visuals:



Page 32 Page 33 Page 33

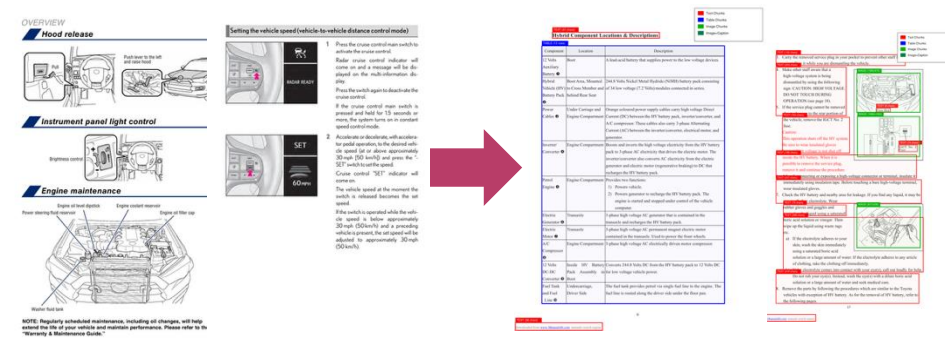
Ask a question about the manual...

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

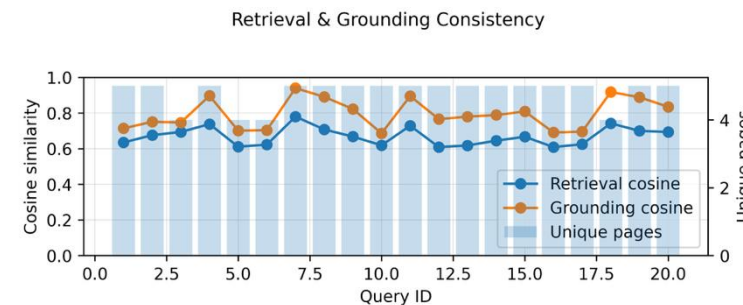
- Benchmark model performance: accuracy vs. inference speed: Deliver a validated model, dashboard, and concise deployment rationale so teams can trade accuracy, cost, and grounding with confidence.

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

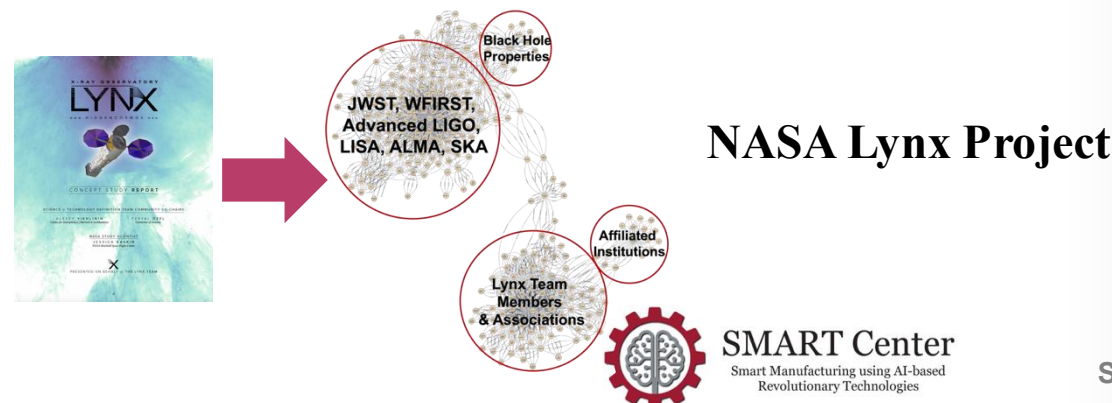
Expected Impacts



Validation Example: PDF Page and Its RAG Chunk



Retrieval QA Benchmark

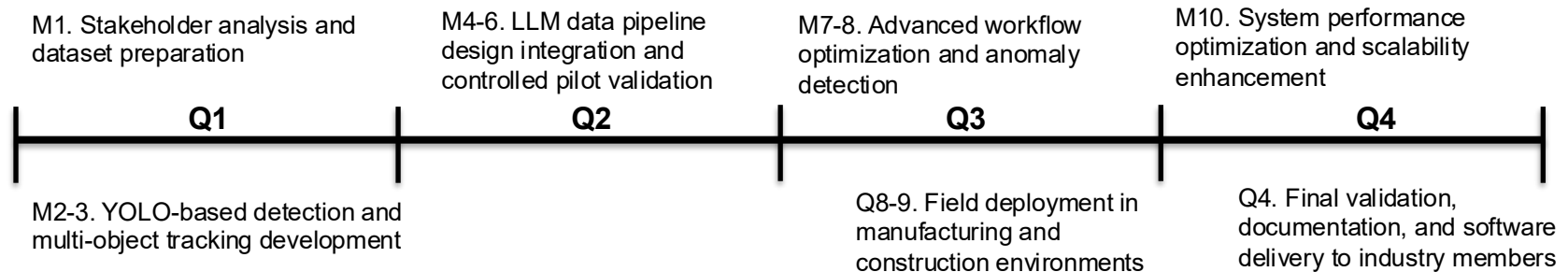


Project Administration

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**



Questions?

National Science Foundation

Where Discoveries Begin



IUCRC



Industry-University
Research
Partnership

SMART Center Planning Meeting - Confidential
The University of Alabama, May 5-6, 2026



SMART Center
Smart Manufacturing using AI-based
Revolutionary Technologies

Slide 10