Utility AI: Systems are Individually Excellent, but Collectively Broken

1 **Pilots everywhere, little in production.**

95% of enterprise AI projects deliver no measurable ROI. In utilities, chatbots and billing bots rarely scale past pilot. Only 1% of the 92% Agentic projects are in production.

2 **Every system is optimised; the whole business process is not.**

CIS, MDM, OMS and field service each excel locally. Move-ins, high-bill disputes and outages break at every hand-off.

3 **People absorb the non-happy path.**

Estimated bills, meter exceptions and storm surges land on back-office teams, unmeasured and unassisted, driving up cost-to-serve. Happy path covers only 18% of inbound service requests.

The gap **The failure is architectural, not technological.**

THE ANSWER

An agentic conveyor belt for executing Utility business processes

- Binds CIS, MDM, OMS and field service
- Resolves exceptions end to end
- Absorbs storm and outage surges
- Co-exists with SAP IS-U and CC&B
- Cuts cost-to-serve on every journey

Source: MIT NANDA, “The GenAI Divide” (2025); Quant

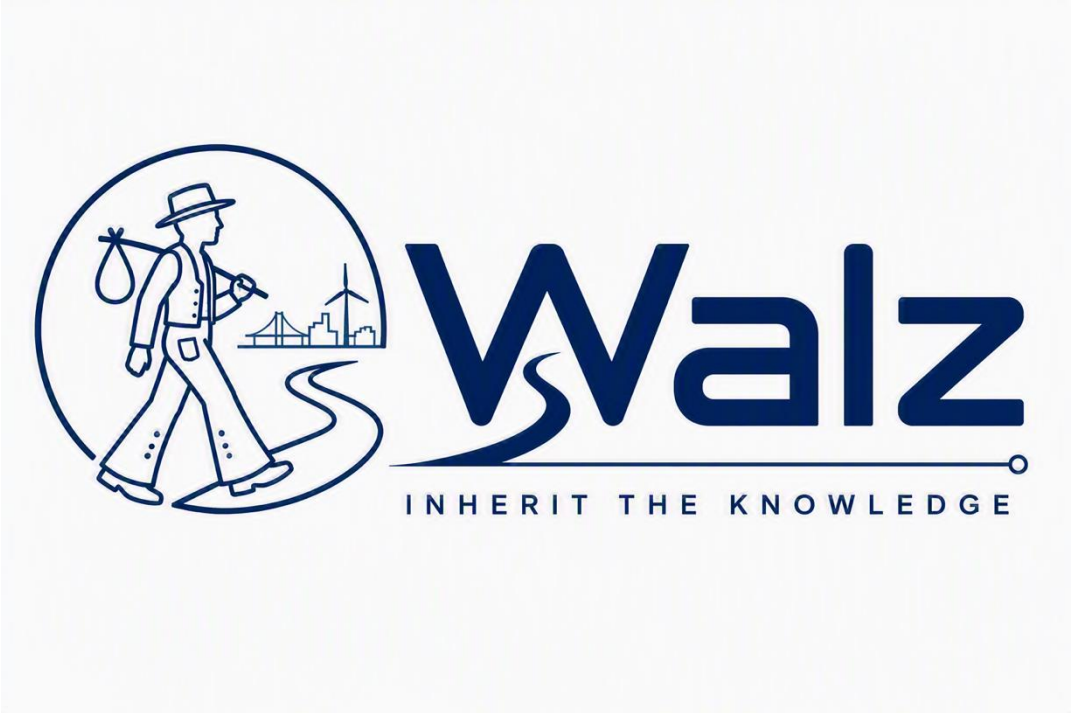
Walz - Oral History App

Knowledge Capture & Retention

Dr. Lea Boche lboche@epri.com
Avery Hodges ahodges@epri.com

© 2026 Electric Power Research Institute, Inc. All rights reserved.

Why Capture Oral History?



Problem: Subject Matter Experts (SMEs) hold hidden knowledge that is

- Not written in reports
- Not recorded in webcasts
- Only shared verbally

Solution: Create a formal process to

- Capture SME knowledge efficiently
- Preserve expertise across generations
- Share lessons learned across the power industry

© 2026 Electric Power Research Institute, Inc. All rights reserved.

Knowledge Capture & Retrieval Architecture

1. Capture

Oral histories are recorded and ingested into the app.

2. Transcribe

On-premises Whisper turns audio into accurate transcripts.

3. Anonymize

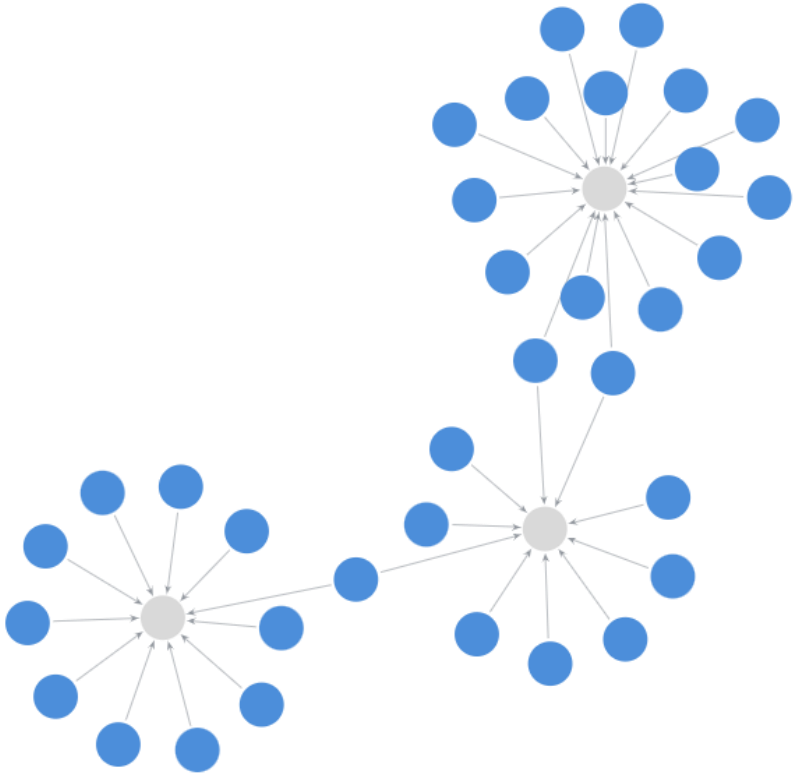
Sensitive details are stripped to protect privacy.

4. Connect

A knowledge graph links people, places, and themes.

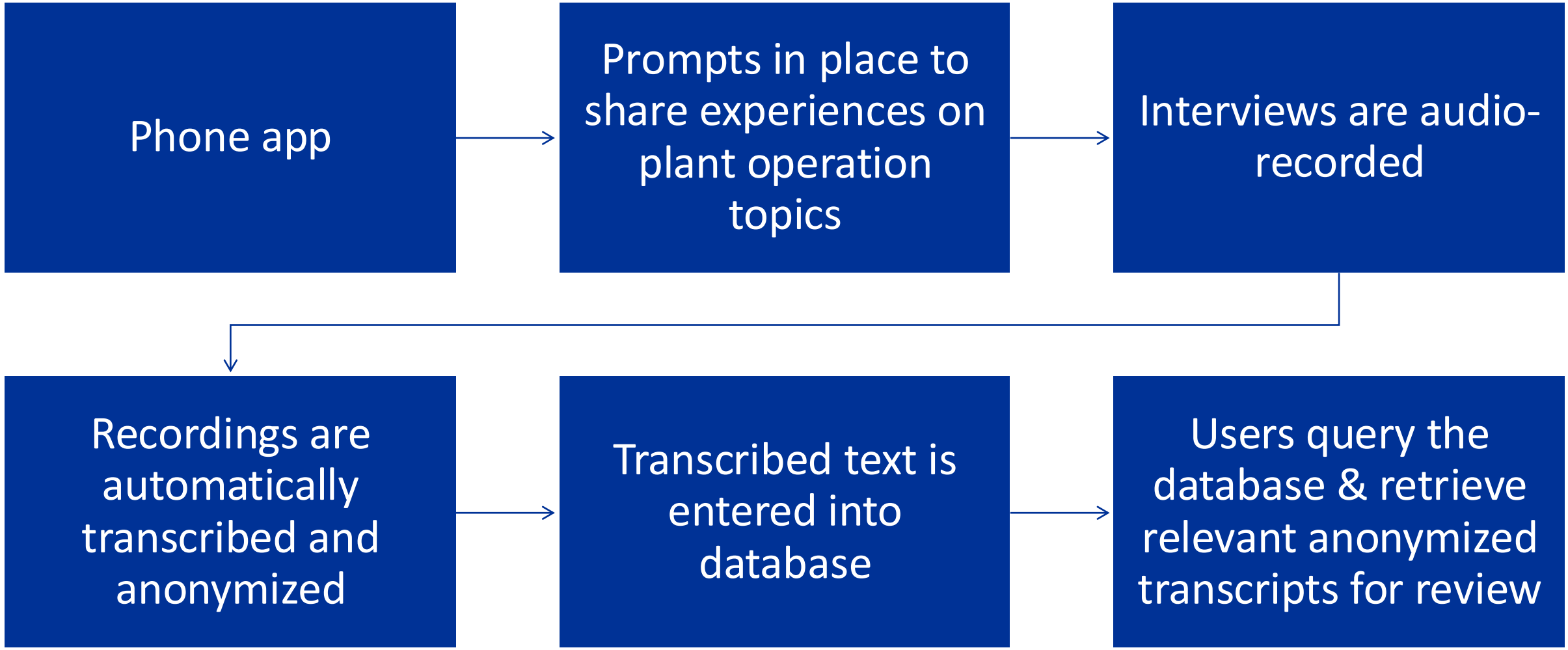
5. Retrieve

An LLM answers questions over the searchable archive.



© 2026 Electric Power Research Institute, Inc. All rights reserved.

User Experience Pipeline



Goal: Knowledge Retention, Storage, Searchability

© 2026 Electric Power Research Institute, Inc. All rights reserved.

Searching Through Knowledge: An Example

User Query → App Searches → Retrieved Results

"Show me interviews that mentioned startup issues for feedwater pumps at nuclear plants."

- Asset:
  Feedwater Pump
- Incident:
  Startup Event
- Plant type:
  Nuclear
- Source:
  All Transcripts

- Interview 011

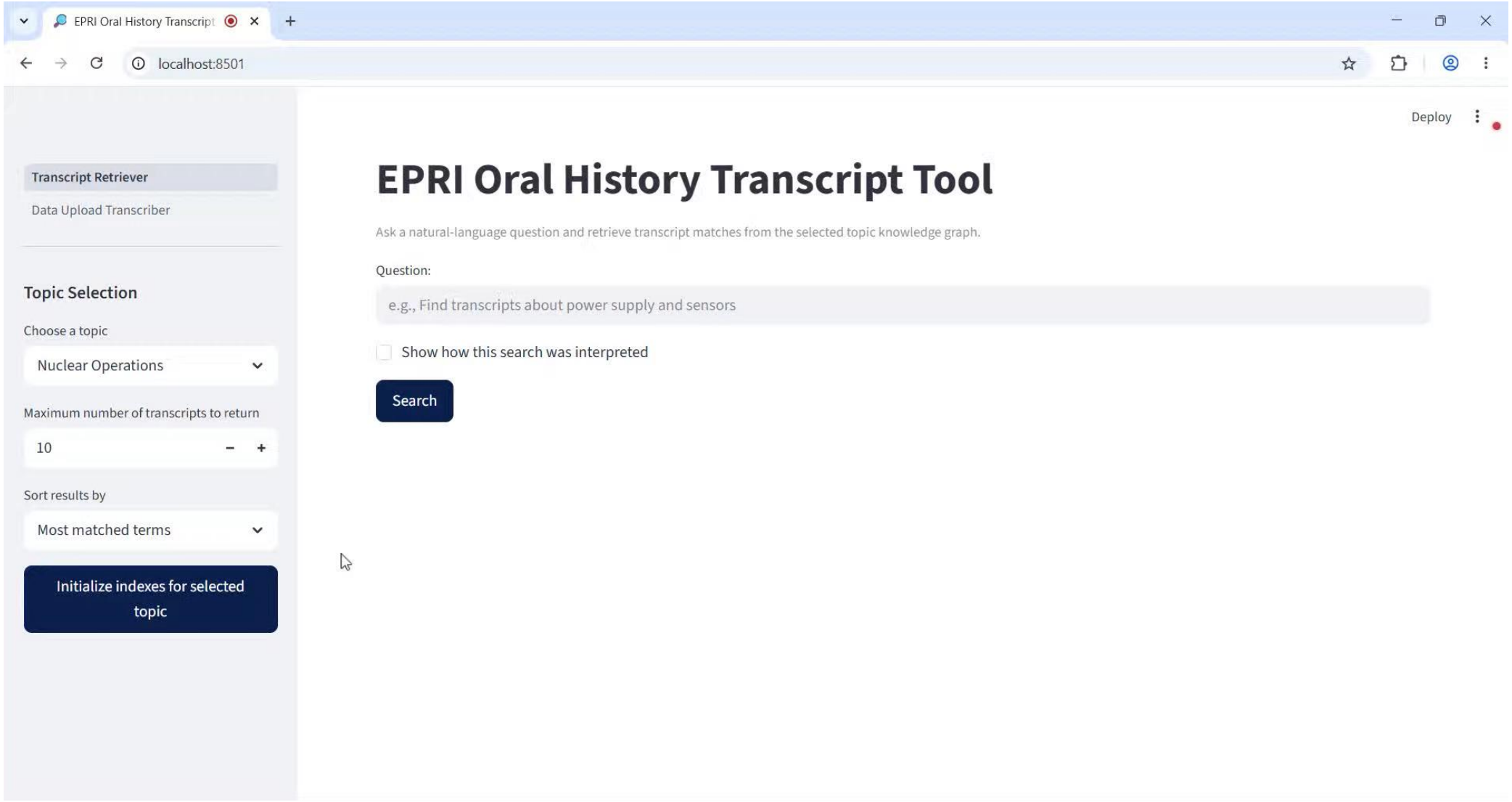
"During startup, we noticed vibration at <LOCATION>, and …"

- Interview 148

"In a situation like this, <NAME> and I went to double-check valve…"

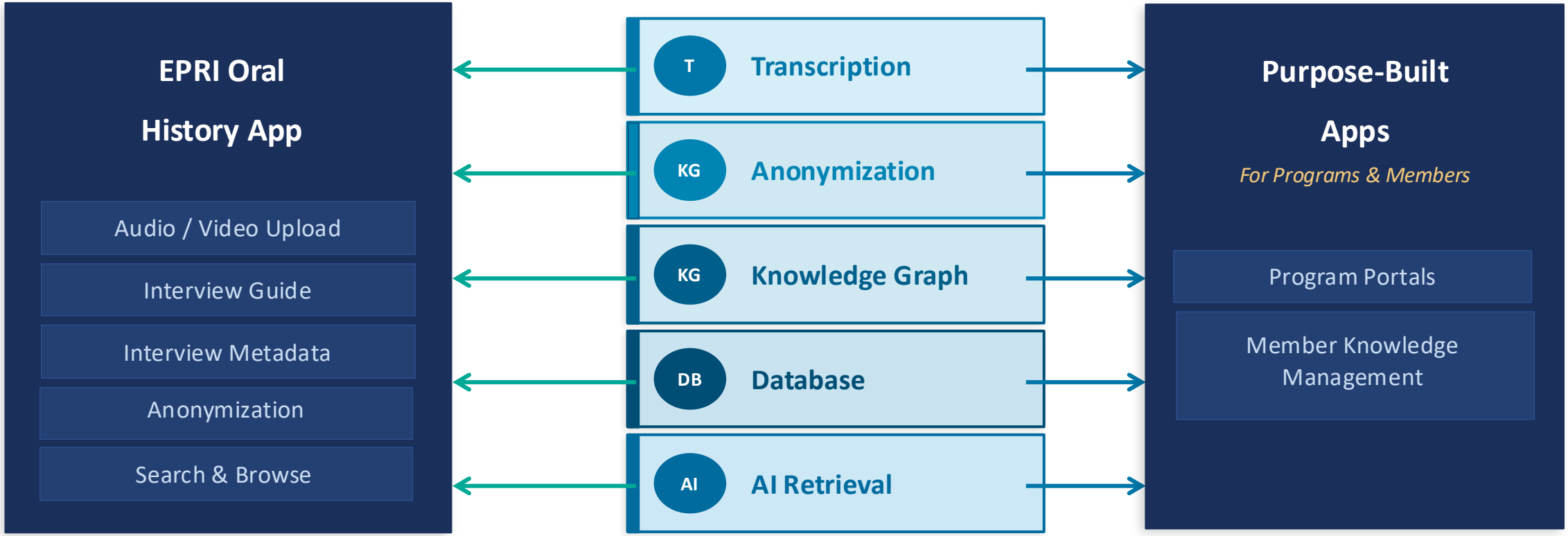
© 2026 Electric Power Research Institute, Inc. All rights reserved.

App Demo



© 2026 Electric Power Research Institute, Inc. All rights reserved.

Vision for the Tool(s)



© 2026 Electric Power Research Institute, Inc. All rights reserved.

WALZ

INHERIT THE KNOWLEDGE

Secure Knowledge Capture and KAG-Powered Discovery

Guided interviews • Privacy-first processing • Structured evidence • EPRI-hosted

1 KNOWLEDGE CAPTURE & PROCESSING

Guided conversation to governed knowledge

1 GUIDE	2 CAPTURE	3 TRANSCRIBE	4 ANONYMIZE	5 PROCESS	6 STORE + STRUCTURE
Dedicated interview guide Tiered prompts Consistent knowledge capture	**Interview audio** SME experiences Secure recording input	**Whisper** On-premises speech-to-text Transcript draft	**Microsoft Presidio** Custom utility-name list Custom location list	**Python processing** Timestamp boundaries Token-window segments	**PostgreSQL + Neo4j** Redacted transcript + metadata Entities, relationships Embeddings + vector index

INTERVIEW-GUIDED CAPTURE

REDACTED TEXT DOWNSTREAM

RELATIONSHIP-FIRST GRAPH

2 KNOWLEDGE DISCOVERY PIPELINE

KAG • KNOWLEDGE-AUGMENTED

1 ASK	2 UNDERSTAND	3 RETRIEVE	4 RANK	5 GROUND	6 PRESENT
Natural-language query User searches WALZ	**Gemma-3-27B** Query intent Query embedding	**Neo4j vector search** Top matching chunks Similarity + graph context	**Transcript aggregation** Group by transcript ID Rank matching interviews	**KAG** Graph relationships Transcript evidence + metadata	**WALZ interface** Relevant interviews Matched excerpts + connections

SEMANTIC + GRAPH RETRIEVAL

GROUNDED IN REDACTED EVIDENCE

DATA PLACEMENT & PLATFORM

- **OBJECT STORAGE**
  - Encrypted interview audio
  - Enterprise object storage
- **POSTGRESQL**
  - Authoritative redacted text
  - Versions, metadata, processing status
- **NEO4J**
  - Entities, relationships, provenance
  - Chunk embeddings + vector index
- **APPLICATION**
  - Streamlit user interface
  - Python / API services
- **LLM**
  - Gemma-3-27B
  - Keyword extraction + querying

EPRI ENTERPRISE INFRASTRUCTURE

- On-premises operation
- Internal data governance
- Privacy-first architecture
- No public-cloud dependency

Supplemental Project Notice: Walz Portal

- **Deliverables**
  - Oral history app and web platform; transcription and anonymization workflow
  - Initial anonymized transcript database and hybrid search framework
  - Participant-specific knowledge graphs and interview guides
- **Price and Schedule**
  - $40,000 total; minimum of five funders required for full scope
  - Two-year project; first website rollout targeted for late 2027
- **Contacts**
  - Lea Boche, lboche@epri.com
  - Avery Hodges, ahodges@epri.com

Objective

- Reduce the risk of knowledge loss from workforce transitions and retirements
- Pilot a mobile and web platform that captures expert insights and lessons learned not documented elsewhere

Approach

- Guided interviews, transcription, and anonymization into a secure, searchable repository
- LLM-enabled summaries paired with rules-based search for traceable answers

https://www.epri.com/research/products/000000003002036852

© 2026 Electric Power Research Institute, Inc. All rights reserved.

TOGETHER…SHAPING THE FUTURE OF ENERGY®

www.epri.com

© 2026 Electric Power Research Institute, Inc. All rights reserved.

Ongoing Progress



Report on First Results



Internal Development & Testing of the app



Integration of Feedback



First Member Beta Test

© 2026 Electric Power Research Institute, Inc. All rights reserved.

Core Objectives

Capturing Undocumented Expertise	The app records informal knowledge and experiences to fill gaps in institutional memory effectively.
Supporting Knowledge Retention	A robust research database stores valuable insights, enhancing operational practices and fostering innovation.
Enabling Industry-wide Sharing	The app offers a scalable framework for sharing lessons and legacy information across the energy sector.
Ensuring Privacy and Security	Anonymization and secure data handling ensure compliance with privacy standards and maintain data integrity.

© 2026 Electric Power Research Institute, Inc. All rights reserved.

AI value is real — but the path to scale changes with consequence

BPA's experience points to a few practical lessons that now shape how we adopt, govern, and scale AI.

LESSON 1

Value can move fast

Enterprise productivity can reach broad adoption and measurable savings in months — not years.

LESSON 2

Maturity should be uneven

Low-friction uses can scale while operationally sensitive capabilities remain intentionally gated.

LESSON 3

Assurance becomes part of the product

As consequence rises, data, testing, controls and accountable humans become part of the capability itself.

REAL-WORLD SIGNALS

BPA is already seeing adoption, value, and a growing pipeline at the same time.

~2,000	$4.1M	20+
active Joulix users	estimated annual expense savings	proposed use cases

TRANSFERABLE PATTERN



The goal is not to push every use case through the same path at the same speed.

TAKEAWAY **Scale what can move quickly; add assurance as the business consequence rises.**

One portfolio — four maturity lanes

The important signal is not the vendor mix; it is the maturity mix.

SCALED

Enterprise productivity

Approved enterprise GenAI is already supporting broad knowledge and productivity work.

Joulix
ServiceNow Virtual Agent*

Broad adoption

MATURITY



Scaled

PILOT / VALIDATE

Functional workflows

Targeted workflows can prove value quickly in sandboxes before broader integration.

Joulix TUI
Defensive Cybersecurity

Evidence building

MATURITY



Proving

BUILD / ASSURE

Enterprise decision support

Cross-functional use raises the bar for common data, controls, and repeatability.

ChatGPT Enterprise
Databricks
ServiceNow Virtual Agent*

Integration matters

MATURITY



Assuring

GATED

Grid-facing capability

Operationally sensitive uses remain gated until evidence and accountability are strong enough.

GridUnity
EPRI Quest/Raphson
GridCare

Assurance first

MATURITY



Intentionally gated

INTENTIONAL ASYMMETRY **Different lanes should mature at different speeds because the cost of being wrong is different.**

TAKEAWAY **An uneven portfolio is a feature: maturity should match business consequence and evidence.**

As impact increases, the meaning of “ready” changes

Higher consequence does not mean “no AI” — it means stronger evidence before dependence.

INCREASING BUSINESS CONSEQUENCE → INCREASING ASSURANCE

1× General support

BUSINESS CHANGE

Broad productivity / knowledge support

WHAT “READY” MEANS

- Approved platform
- Basic guidelines + prohibitions
- Light governance
- User training

2× Local / functional

BUSINESS CHANGE

Defined workflow or functional decision support

WHAT “READY” MEANS

- Named sponsor
- Clear scope
- Clear success criteria
- Sandbox evidence

3× Enterprise

BUSINESS CHANGE

Cross-functional process or enterprise decision support

WHAT “READY” MEANS

- Common data / pipelines
- Shared validation
- Increased governance
- Robust cyber / privacy controls

4× Mission critical

BUSINESS CHANGE

Operational capability with material consequences

WHAT “READY” MEANS

- TEVV + benchmarks
- Monitoring
- Clear human accountability

TAKEAWAY **Higher consequence does not mean “no AI”; it means a higher bar of evidence before dependence.**

A metric to balance exploration with operation

The metric is a management tool — not the story itself.

WORKFLOW IMPORTANCE

1× → 2× → 3× → 4×

ADOPTION STAGE

0% → 20% → 40% → 70% → 100%

WEIGHTED ADOPTION-TO-SCALE INDEX

A common portfolio signal

Weighted average of stage score by workflow importance

STAGE SCORES

0%	20%	40%	70%	100%
No active use	Active pilot	Validated / approved	Early production	Scaled production

USE IT TO MAKE BETTER PORTFOLIO DECISIONS

- Pilot depth
- Funding / sponsorship
- Evidence required
- Control requirements
- Scale timing
- Peer benchmarking

TAKEAWAY Count pilots, but reward validated production and scale — especially in important workflows.

Dependable AI requires an operating stack around the model

The scalable unit is not the model alone — it is the surrounding data, test, policy, and operating system.

01 **Approved enterprise GenAI platform**
A controlled place for broad adoption and common capabilities

02 **Common data layer + access controls**
Reusable, governed access to the information workflows actually depend on

03 **Sandboxes + national-lab test environments**
A place to experiment, isolate risk, and collect evidence before production

04 **TEVV + benchmark framework**
Repeatable testing, evaluation, verification, and validation for consequential uses

05 **Policy, cyber, privacy + sensitive-data boundaries**
Explicit limits on where data and decisions may flow

WHAT THIS STACK ENABLES

Faster learning

without putting production at risk

Reusable evidence

that can support procurement, approval, and peer comparison

Local operating authority

shared assurance where useful; local decisions where necessary

TAKEAWAY **The scalable unit is model + data + test environments + assurance + operating controls.**

Governance is the path to scale — not a checkpoint at the end

Define the evidence and accountability before the pilot so progression can be repeatable.

ADOPT

Is this a legitimate use?

- Named sponsor
- Target workflow
- Approved environment

BUILD

Can it produce useful evidence?

- Sandbox access
- Data path
- Success criteria

ASSURE

Is it ready for dependence?

- TEVV / benchmark
- Cyber + privacy
- Human decision rights

OPERATE

Can we sustain it safely?

- Production approval
- Monitoring
- Feedback / stop conditions

OPERATING DISCIPLINE

Close data and testing gaps before procurement and production.

Expand only what clears the agreed governance gates.

Program manager + governance subcommittee turn policy into execution.

TAKEAWAY **Define the gate before the pilot: evidence and accountability make progression repeatable.**

Build assurance once; reuse it many times

The shared work is the work that does not need to be reinvented utility by utility.

01

Common risk taxonomy + minimum scorecard (SAFER.ai)

Shared definitions of consequence, evidence, controls, and readiness

02

Benchmark suites + TEVV patterns (OPAI)

Reusable tests and acceptance evidence for reference use cases

03

Data schemas + sandbox patterns (Data Mgmt)

Portable structures for testing without forcing shared operations

04

Governance + change playbooks (SAFER.ai)

Training, decision rights, communications, and lessons that travel

A PRACTICAL 90-DAY START

1. Pick 2 reference use cases
2. Agree a minimum assurance scorecard
3. Run a shared sandbox cohort
4. Publish the reusable artifact pack

120+

utilities = enough scale to reduce duplicated testing and raise the assurance floor

TAKEAWAY **Standardize assurance assets centrally; preserve local operating decisions locally.**