



March 4, 2026 | 2nd AI4Ops Symposium, Santa Fe

Blueprint: Rebuilding a Legacy

Tirthankar Ghosal, Ethan Seefried,
Rob Saethre

Oak Ridge National Laboratory



U.S. DEPARTMENT
of **ENERGY**

ORNL IS MANAGED BY UT-BATTELLE LLC
FOR THE US DEPARTMENT OF ENERGY



Current EDRM: A Deep Dive

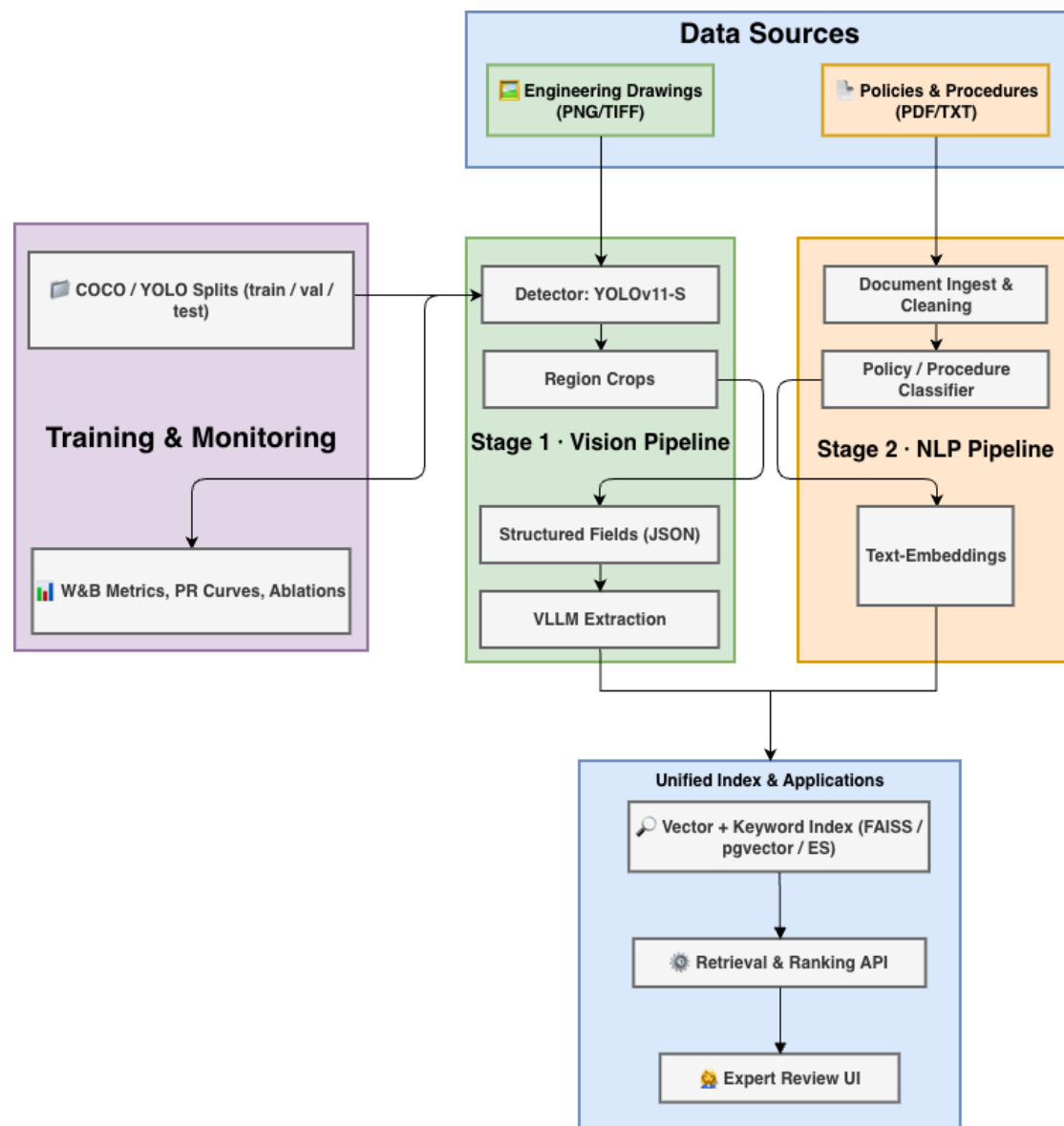
- **Missing core metadata** (extension, doc type, revision, facility/tag) → poor precision *and* recall.
- **Search fatigue**: users spend hours and still aren't confident they found the right file.
- **Firehose results**: thousands of hits with duplicates and superseded versions mixed in.
- **Version sprawl & naming chaos**: no clear “latest” and inconsistent filenames hide the source of truth.



Problem Statement

- **Unlocking legacy engineering knowledge:** Thousands of archived drawings and procedural documents remain trapped in unstructured formats, limiting access and reuse.
- **Dual-pipeline AI system:** A **vision** and **language** pipeline work together to extract, classify, and link visual and textual information.
- **Concrete example:** "Engineer asks 'Where are target rupture disc replacement procedures?' → system returns procedure doc + equipment drawings"
- **Time/cost saved:** "Manual search: May never find the doc → Automated: 10 seconds"
- Protection against loss of institutional knowledge
- This project used ProjectWise data from 2020 not current data in EDRM
- We need to have parent/child relationships extracted from engineering drawings to build data structures in PLM

Stage 1: Holistic Overview



Vision Pipeline

- Three core stages

- Detect

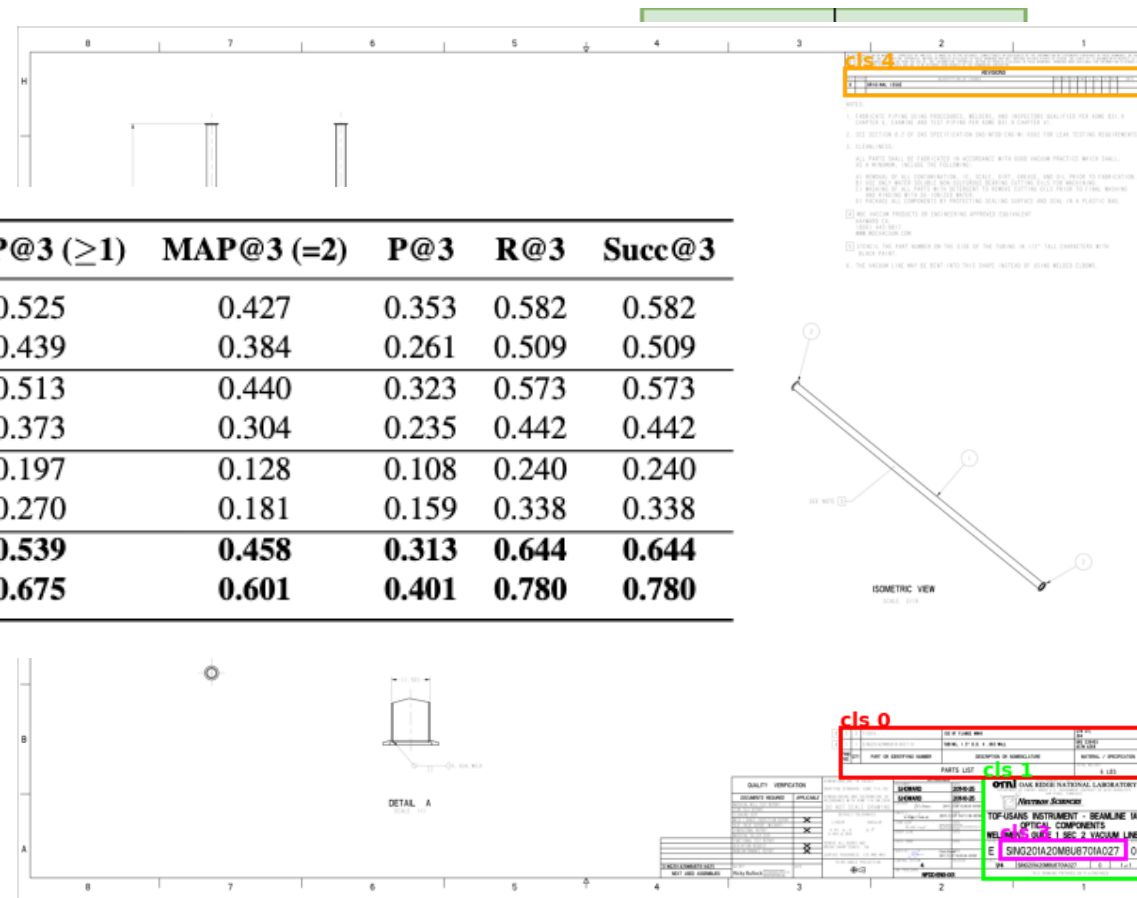
- Core info

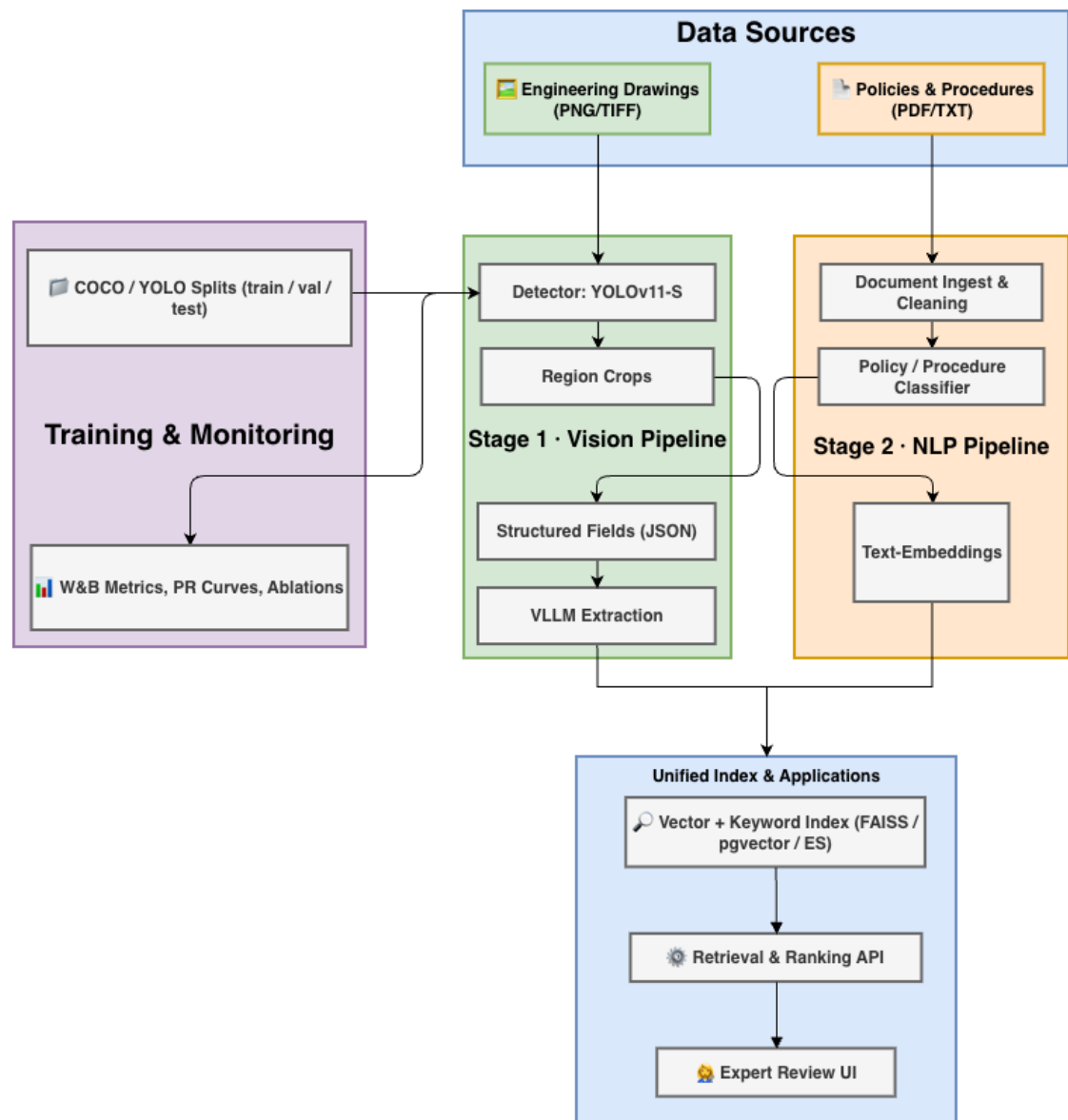
- Parts

- Utilizes

- Yolo,

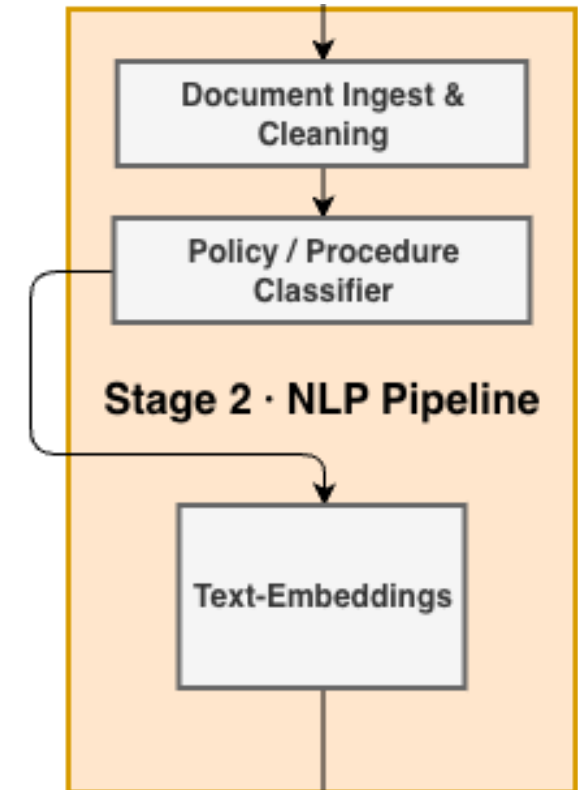
System Variant	nDCG@3	MAP@3 (≥ 1)	MAP@3 (=2)	P@3	R@3	Succ@3
Tesseract (Pred boxes)	0.532	0.525	0.427	0.353	0.582	0.582
Tesseract (Oracle boxes)	0.457	0.439	0.384	0.261	0.509	0.509
EasyOCR (Pred boxes)	0.525	0.513	0.440	0.323	0.573	0.573
EasyOCR (Oracle boxes)	0.390	0.373	0.304	0.235	0.442	0.442
Tesseract (Fullpage) <i>diag.</i>	0.208	0.197	0.128	0.108	0.240	0.240
EasyOCR (Fullpage) <i>diag.</i>	0.287	0.270	0.181	0.159	0.338	0.338
BLUEPRINT (Pred boxes, VLM-OCR)	0.563	0.539	0.458	0.313	0.644	0.644
BLUEPRINT (Oracle boxes, VLM-OCR)	0.699	0.675	0.601	0.401	0.780	0.780

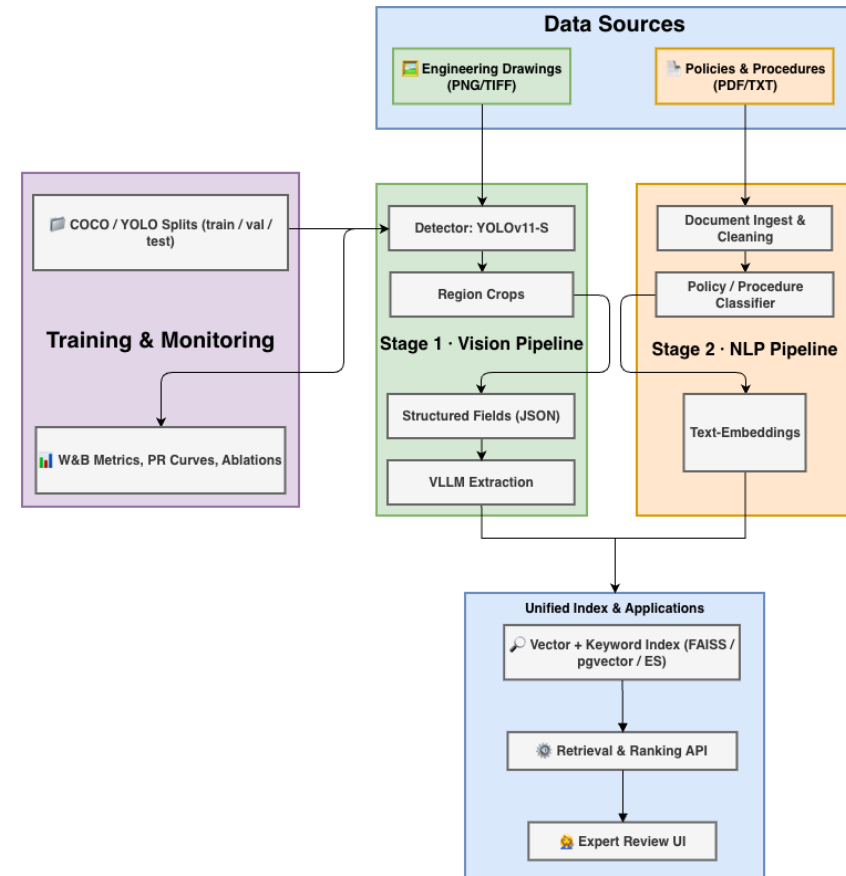




NLP Pipeline: LLM-Based Document Analysis

- Each document is processed through a Large Language Model to interpret and summarize content.
- The model identifies document type, topic, and purpose (e.g., policy, procedure, report).
- Text is transformed into structured embeddings for semantic search and downstream tasks.
- Provides the foundation for intelligent retrieval and cross-domain linking.





Unified Retrieval System

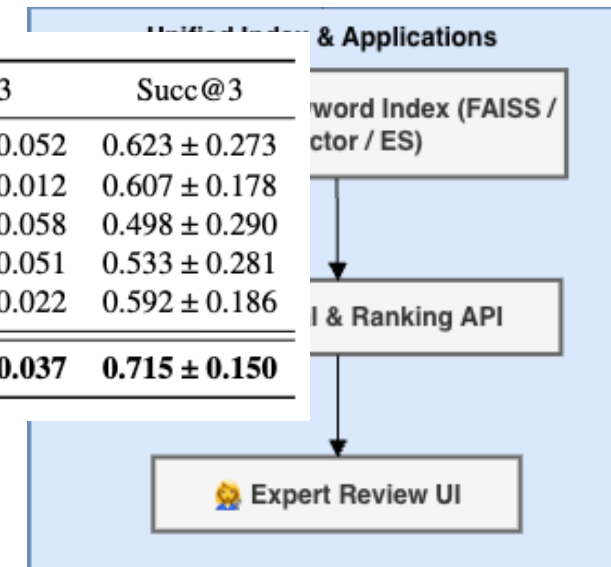
- Combines vision and language embeddings into a unified searchable index

- Allows unified procedure

System	nDCG@3	MAP@3 (≥ 1)	MAP@3 (=2)	P@3	R@3	Succ@3
LLaMA 3.2 Vision	0.521 ± 0.248	0.497 ± 0.244	0.330 ± 0.256	0.342 ± 0.145	0.151 ± 0.052	0.623 ± 0.273
Llama 4 Scout 17B	0.519 ± 0.158	0.503 ± 0.156	0.327 ± 0.200	0.364 ± 0.096	0.137 ± 0.012	0.607 ± 0.178
LLaVA 1.6 Mistral 7B	0.400 ± 0.246	0.378 ± 0.236	0.193 ± 0.138	0.275 ± 0.154	0.109 ± 0.058	0.498 ± 0.290
PaliGemma 2	0.422 ± 0.240	0.395 ± 0.232	0.195 ± 0.169	0.307 ± 0.166	0.122 ± 0.051	0.533 ± 0.281
Pixtral 12B (2409)	0.502 ± 0.159	0.486 ± 0.153	0.314 ± 0.233	0.354 ± 0.114	0.131 ± 0.022	0.592 ± 0.186
BLUEPRINT (Ours)	0.626 ± 0.146	0.608 ± 0.147	0.407 ± 0.241	0.435 ± 0.091	0.222 ± 0.037	0.715 ± 0.150

- Bridges visual and written policies.

- Forms the backbone of a retrieval-augmented decision-support platform.



Stage 2: Parent/Child Relationships

Automatically extract structured parts-list items from detected regions, normalize them, and link entries to parent drawings and related documents to enable precise retrieval and expert ranking.

Combine OCR output, lightweight parsing heuristics, and weak supervision (rules + LLM-assisted normalization) to scale extraction across ~750 drawings with human-in-the-loop verification.

Linking & Displaying Connected Files

- Each parsed row becomes a child entity linked to its parent drawing.
- Build a unified index (text embeddings + filters) for parts + documents.
- Enable queries like “find all drawings with Part A100” or “valves near Facility X.”
- Rank results by confidence + expert feedback; preview where the part appears.
- Surface related docs (procedures, policies) via shared parts/facility context.

	drawing_number	item_no	part_number
0	155Y502672	25	95045A036
1	155Y502672	24	SSN-001
2	155Y502672	23	92676A029
3	155Y502672	22	94812A415
4	155Y502672	20	91104A047
5	155Y502672	19	98023A036
6	155Y502672	18	90295A170
7	155Y502672	17	92165A029
8	155Y502672	16	90295A122
9	155Y502672	14	92941A537
10	155Y502672	12	SST-001
11	155Y502672	11	93025A964
12	155Y502672	10	98831A411
13	155Y502672	9	BAL154/TB
14	155Y502672	6	155Y502682
15	155Y502672	5	155Y502674
16	155Y502672	4	155Y502673
17	155Y502672	3	155Y502601
18	155Y502672	2	155Y502594
19	155Y502672	1	155Y502396
20	155Y502672	?	155Y502672

Impact of the Project

1. The Critical Knowledge Transfer Challenge; Accelerated Onboarding and Training
2. The Complexity Barrier – Digitizing the knowledge (cross-referencing the multitude of diagrams)
3. Enhanced Operational Safety and Efficiency (reduce misinterpretation risk, informed upgrades, faster troubleshooting, etc.)
4. Preservation of Institutional Knowledge (digital institutional memory that never retires)
5. Accelerated Innovation and Upgrade – Second Target Station (faster design decisions and planning)
6. Competitive Advantage for SNS vs others (position as an AI-augmented facility)
7. Scalability Beyond SNS – HFIR
8. Frontier AI Challenge with Practical Application (long-term research benefits and reputation)
9. Enabling Research Productivity (only focus on the science, let AI decipher the complex drawings)