



FY27 NPP LDRD Type B Pre-Proposal



# Empowering ePIC with AI-Powered Collaborative Tools and Infrastructure

PIs: Thomas Ullrich (presenter), Dmitry Arkhipkin

January 8, 2025



# FY27 NPP LDRD Type B Pre-Proposal

**Primary Investigator:** Thomas Ullrich, Dmitry Arkhipkin

**Other Investigators:** Uma Ganapathy (SCDF), Brian Page, Peter Steinberg

**ECA eligibility:** No

**Proposal Term:** October 1, 2026 - September 30, 2028

**Annual funding:** FY27 \$250 , FY28 \$250

# Description of LDRD - Motivation

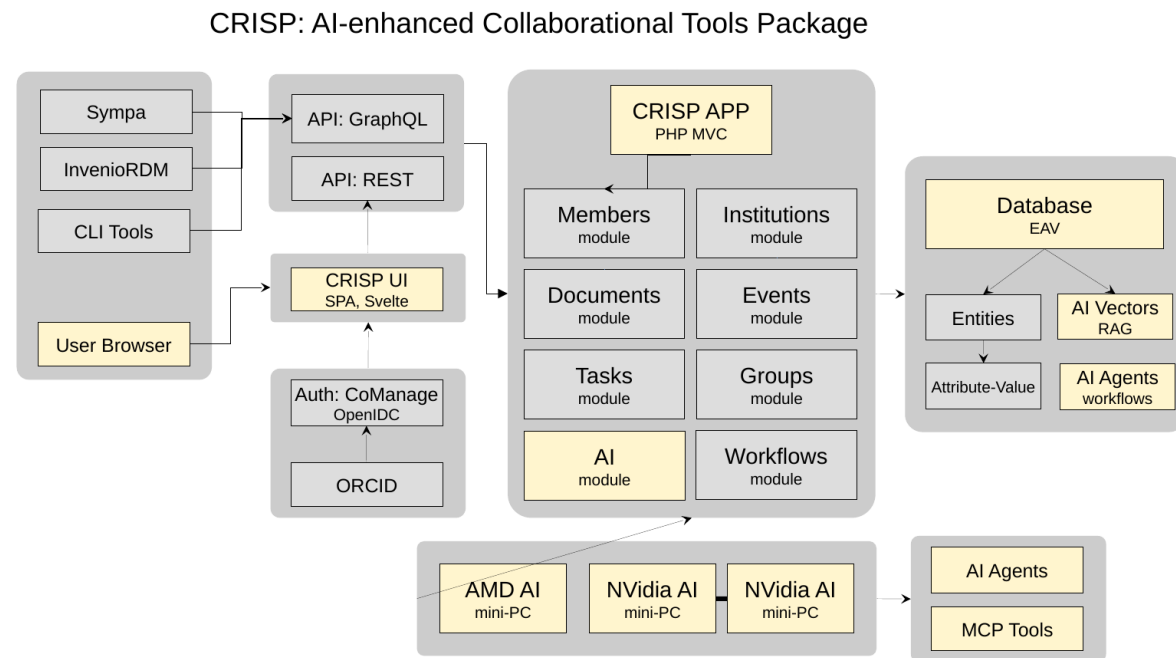
## Empowering ePIC with AI-enabled collaborative tools and infrastructure

- ePIC is a large, distributed collaboration ( $\approx 1,000+$  members; 180+ institutions; 26 countries) supporting EIC detector construction and physics.
- Coordination overhead (meetings, documentation, software workflows, design reviews) consumes significant expert time.
- Physics has a history of collaboration-driven innovation deploying the latest technologies (e.g., the World Wide Web) which today means intelligent assistance.
- Project goal: deploy privacy-preserving, locally hosted AI assistants integrated with ePIC collaboration tools to reduce friction and risk.

Key constraint: all AI inference runs on dedicated on-prem hardware (no commercial cloud) to protect sensitive internal data.

# Description of LDRD - Concept and Approach

- Build on the CRISP\* collaboration platform to add an “AI module” that connects to documents, events, tasks, and workflows.
- Run medium-sized language models locally on AI-enabled mini-PCs for secure, cost-effective deployment.
- Implement secure ingestion + indexing + retrieval (RAG) over internal ePIC documentation, code, and notes, archives.
- Provide user-facing functions: meeting/document summarization, action-item extraction, Q&A over the knowledge base, and workflow assistance.
- Two-year plan: Year 1 pilot infrastructure + prototype assistant; Year 2 expand features, fine-tune for ePIC terminology, evaluate and train users.



Design emphasis: privacy, model governance, and minimal disruption to existing collaboration workflows.

\* Collaborative Research Information Sharing Platform

# Deliverables, Team, and Resources

<b>Deliverables</b>	<b>Personnel</b>	<b>Procurements &amp; Milestones</b>
<ul style="list-style-type: none"><li>• Operational AI-enhanced prototype integrated with ePIC tools</li><li>• Deployable software stack for on-prem AI assistants</li><li>• Documentation + training materials for users and maintainers</li><li>• Evaluation report with metrics and recommendations</li></ul>	<ul style="list-style-type: none"><li>• New postdoc (75%): integration, model adaptation, pilot rollout</li><li>• D. Arkhipkin (5%): infrastructure + AI deployment</li><li>• U. Ganapathy (10%): systems integration</li><li>• T. Ullrich / P. Steinberg / B. Page (0-5% each): ePIC coordination + user engagement</li></ul>	<ul style="list-style-type: none"><li>• Equipment: ~\$20k in Year 1 (AI-enabled mini-PCs for pilot users)</li><li>• M6: hardware + base model operational</li><li>• M12: pilot assistant tested with user feedback</li><li>• M18: expanded features integrated</li><li>• M24: evaluation (targeted time savings and user satisfaction)</li></ul>

Success metrics (targets): ~30% reduction in time spent on documentation/coordination (survey-based) and ≥80% positive feedback from pilot users.

# Intellectual Merit Summary

- **Core idea:** Bring on-prem AI assistants into ePIC's collaboration environment to automate recurring friction: meeting capture → action items, documentation drafting, and fast retrieval of internal knowledge—while keeping sensitive material inside the firewall.
- **What's novel:** Focuses on AI for collaboration operations (how the experiment works day-to-day, not physics analysis), complementing AI-for-physics analysis efforts and avoiding cloud/vendor dependence for unpublished or internal content.
- **Approach:** Deploy “personal AI” mini-PC nodes plus a secure ingest/index pipeline over ePIC artifacts (docs/repos/notes); tune models to ePIC terminology and engineering context; evaluate with pilot users using measurable productivity metrics.
- **Deliverables:** An integrated prototype and deployable stack, ~5 configured AI nodes, user-facing docs/training, and a final evaluation report with a plan for scaling.

# Return on Investment & Potential Future Funding

- **ROI target:** Demonstrate  $\geq 30\%$  reduction in time spent on documentation/coordination (survey-based), plus faster “find-the-right-info” loops and more consistent technical records.  $\Rightarrow$  **Savings in staff time.**
- **Cost-effective:** A one-time  $\sim \$20k$  hardware procurement (AI mini-PCs + DGX Spark units) enables secure local inference and rapid iteration without ongoing cloud costs or data exposure risk.  $\Rightarrow$  **Cybersecurity, protected information**
- **De-risking strategy:** Start with a limited pilot (postdocs/engineers), instrument usage and outcomes, then scale in Year 2 based on evidence and user feedback.  $\Rightarrow$  **Reduce technical risk.**
- **Future funding path:** Use validated results and metrics to pursue DOE-relevant opportunities (NP/ASCR/SciDAC/crosscutting AI initiatives) and commit to at least one follow-on proposal.  $\Rightarrow$  **Competitiveness for follow-on funding**

# Broader Impact on the Laboratory

- **BNL-wide value:** Improves the practice of large-scale science—coordination, communication, and decision throughput—so **technical work moves faster with less overhead**.
- **Reusable capability:** Establishes a privacy-preserving, on-prem “reference architecture” for AI-enabled collaboration that **can be replicated across other BNL/DOE projects** handling sensitive pre-publication content.
- **Leadership signal:** **Positions BNL as a leader** in responsible, deployable AI for real research workflows (not just demos), grounded in security and operational needs.
- **People impact:** **Develops staff expertise in practical AI deployment** and lowers barriers for distributed collaborators (career stage/time zone), strengthening collaboration health.

# Names of Suggested BNL Reviewers

- Jin Huang
- Jeff Landgraf
- Lijuan Ruan
- Eric Lancon
- Vincent Garrone (SCDF)
- Alexei Klimentov (SCDF)

# LDRD Funding Table

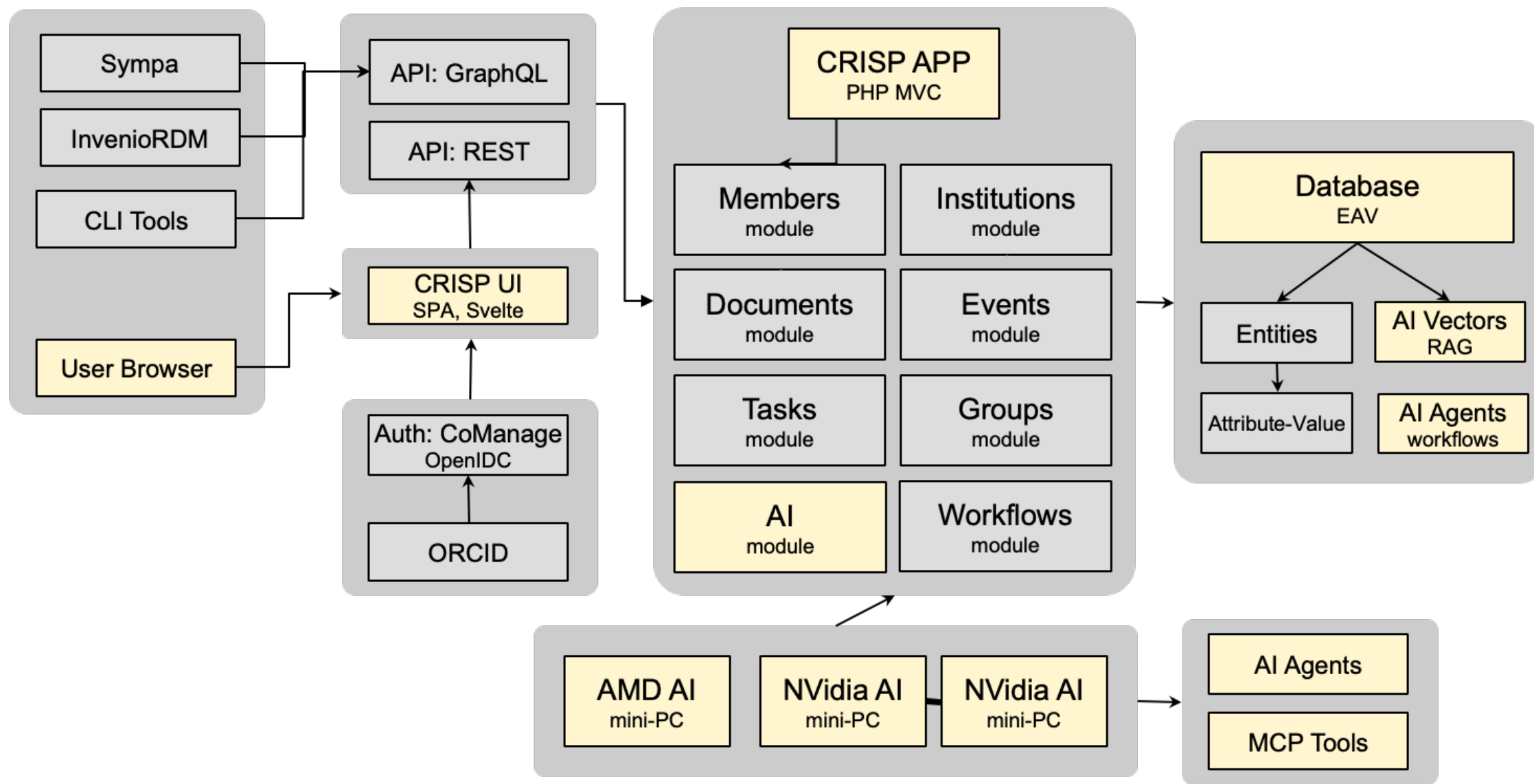
Title: Empowering ePIC with AI-Enabled Collaborative Tools  
and Infrastructure  
LDRD Type B  
PI: Thomas Ullrich

Resource Category	DESCRIPTION	FY27	FY28
050	Salary - Scientific	18,442	19,090
051	Salary - Research Assoc	81,743	85,013
050	Salary - Professional	32,451	33,749
050	Salary - Technical	0	0
050	Salary - Management & Admin.	0	0
	<b>Total FTEs</b>	<b>0.97</b>	<b>0.97</b>
<b>TOTAL SALARY/WAGE &amp; FRINGE</b>		<b>132,635</b>	<b>137,851</b>
	various Contracts - Low Value	0	0
	280 Foreign Travel	0	0
	290 Domestic Travel	3,600	5,100
	various Purchases	18,000	7,000
<b>TOTAL MSTC</b>		<b>21,600</b>	<b>12,100</b>
<b>TOTAL DIRECT COSTS</b>		<b>154,235</b>	<b>149,951</b>
	251 Electric Distributed (Electric Power Burden)	1,326	1,379
	700/701/481 Organizational Burden	18,907	20,291
<b>TOTAL ORGANIZATIONAL BURDEN</b>		<b>20,234</b>	<b>21,670</b>
	745 Procurement (Material Handling)	1,512	847
	735 G&A Burden	0	0
	730 Common Institutional Support	74,019	77,532
	722 Safeguards & Security Assess	0	0
<b>TOTAL LABORATORY BURDEN</b>		<b>75,531</b>	<b>78,379</b>
	705 LDRD Burden	0	0
<b>TOTAL PROGRAM COSTS</b>		<b>250,000</b>	<b>250,000</b>
	740 Full Cost Recovery	0	0
<b>TOTAL PROGRAM COSTS</b>		<b>250,000</b>	<b>250,000</b>

Labor Band	Name	FY27		FY28	
		FTE	Amount	FTE	Amount
RA1	TBD	0.75	81,743	0.75	85,013
SCI4	T. Ullrich	0.02	7,714	0.02	7,933
SCI1	B. Page	0.05	10,728	0.05	11,157
PROF4	D. Arkhipkin	0.05	13,857	0.05	14,411
PROF2	U. Ganapathy	0.10	18,594	0.10	19,337
Total		0.97	132,635	0.97	137,851

# Backup Slides

# CRISP: AI-enhanced Collaborational Tools Package



# CRISP: AI topologies to be implemented and tested

