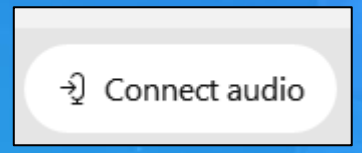
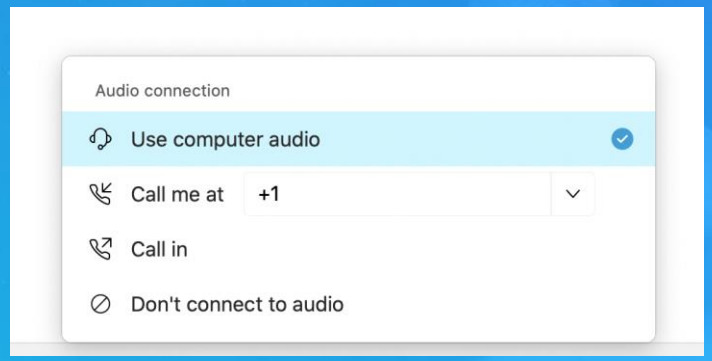
Welcome to the Open Power AI Consortium Domain Specific Model Workstream Kickoff

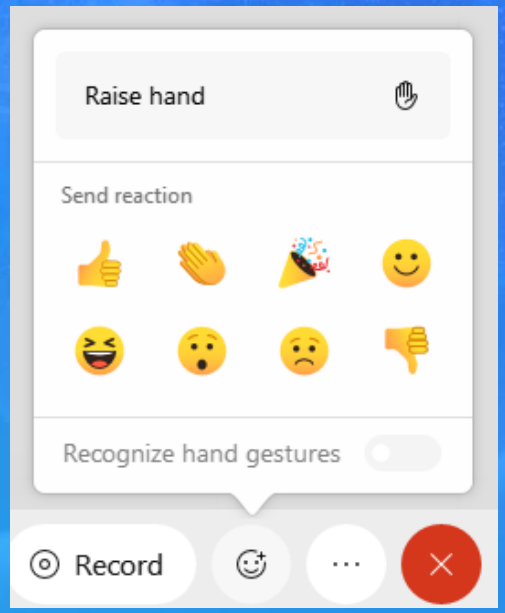
We will begin at 12:00 PM E.T.

Be sure you have an audio icon next to your name. If you do not, disconnect your audio and select one of the options shown. Your line has been muted by the host.

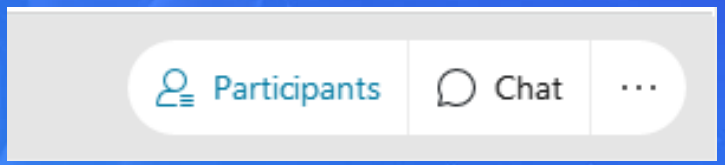




Like what you see or hear or want to raise your hand? Let the Presenters know.



You can submit a written question at any time through the presentation using the CHAT panel.
Please leave the recipient drop down to “Everyone”.



In Chat:
Name/Company

Reminder: This webcast is being recorded

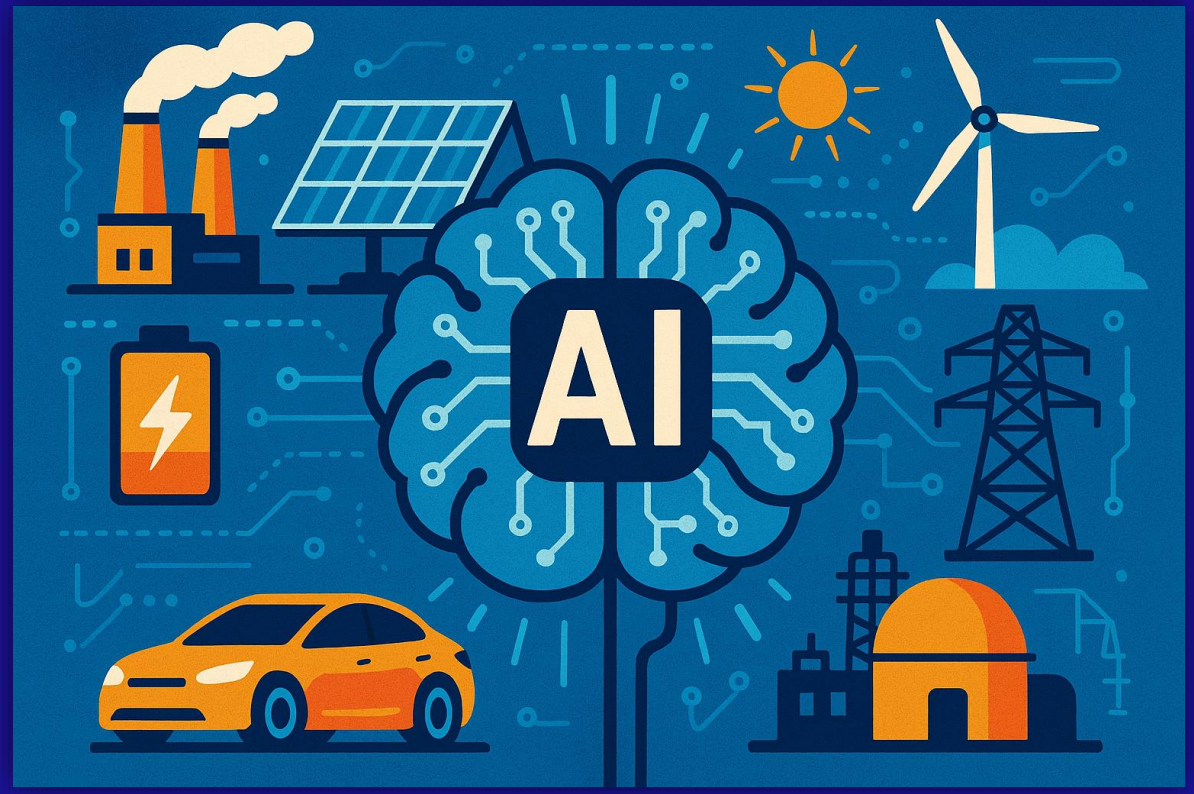
Domain Specific Model Work Stream Kickoff

Open Power AI Consortium

Ben Sooter
Program Manager
bsooter@epri.com
+1-865-218-8108 (office)

Agenda

- Welcome and EPRI Introductions
- Open Power AI
- DSM Workstream
  - Desired Outcomes/Goals
  - Working Together to Implement Solutions
  - Periodic Collaboration Virtual Meetings
  - Transferring the Lessons Learned to Industry
- Next Steps
- Adjourn



Safety Message

Halloween is coming up, be sure and drive slow and watch for children

Join at slido.com
#1619161

Do you have a favorite local model? (Llama, Gemma, DeepSeek, etc)

Open Power AI Consortium Leaders

Daniel Brooks

- OPAI Executive Lead
  - Executive Advisory Group (EAG) Lead

Jeremy Renshaw

- Overall OPAI Lead
  - Member Representative Committee (MRC) Lead

Apurba Sakti

- OPAI Operational Manager
  - Assisting with member connections, operations, and technical excellence

OPAI Work Stream Leads

Jason Hollern

- Implementation Work Stream Lead

Adrian Kelly

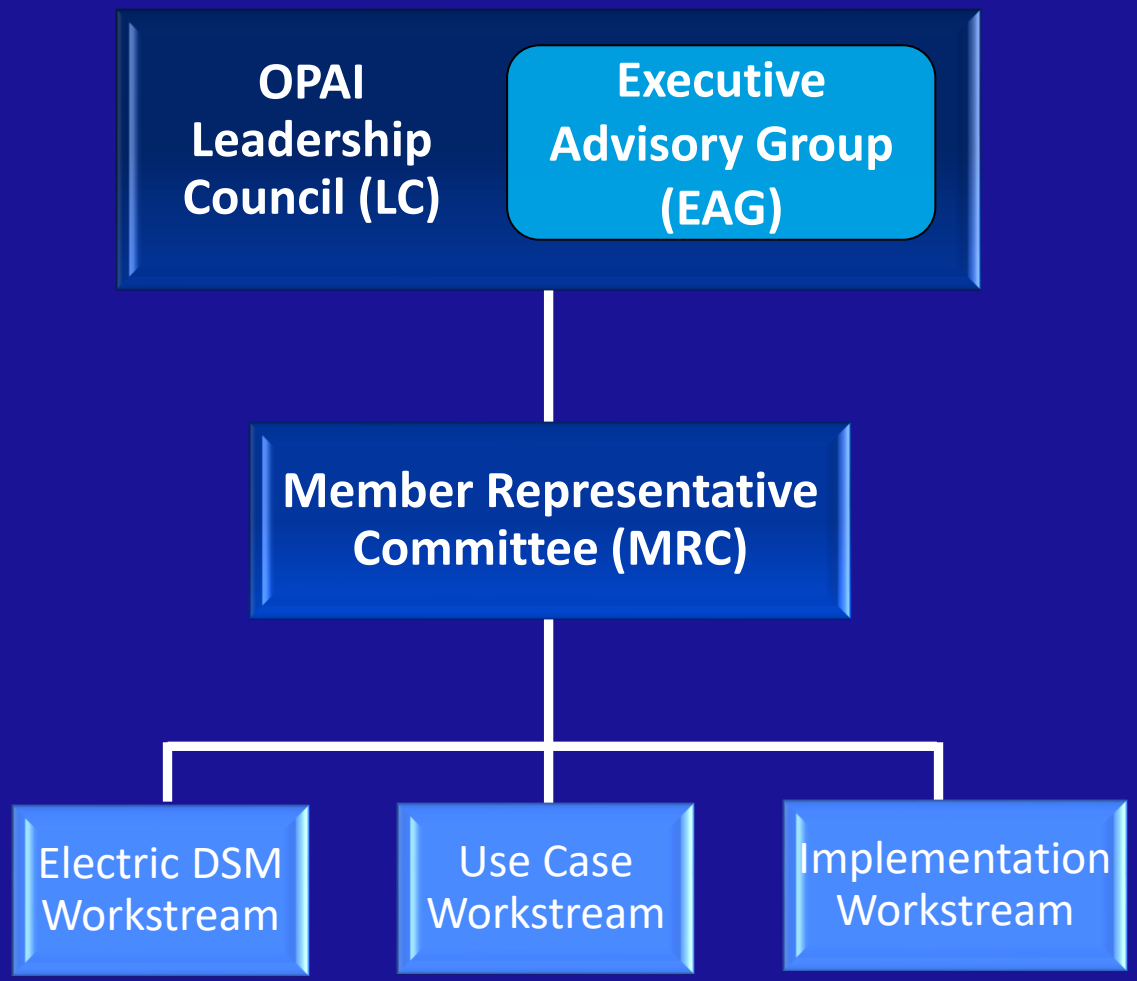
- AI Use-Cases Work Stream Lead

Ben Sooter

- Domain-Specific Model Work Stream Lead

Open Power AI Status Update

OPAI Consortium Member Engagement



https://msites.epri.com/opai

Consortium Engagement

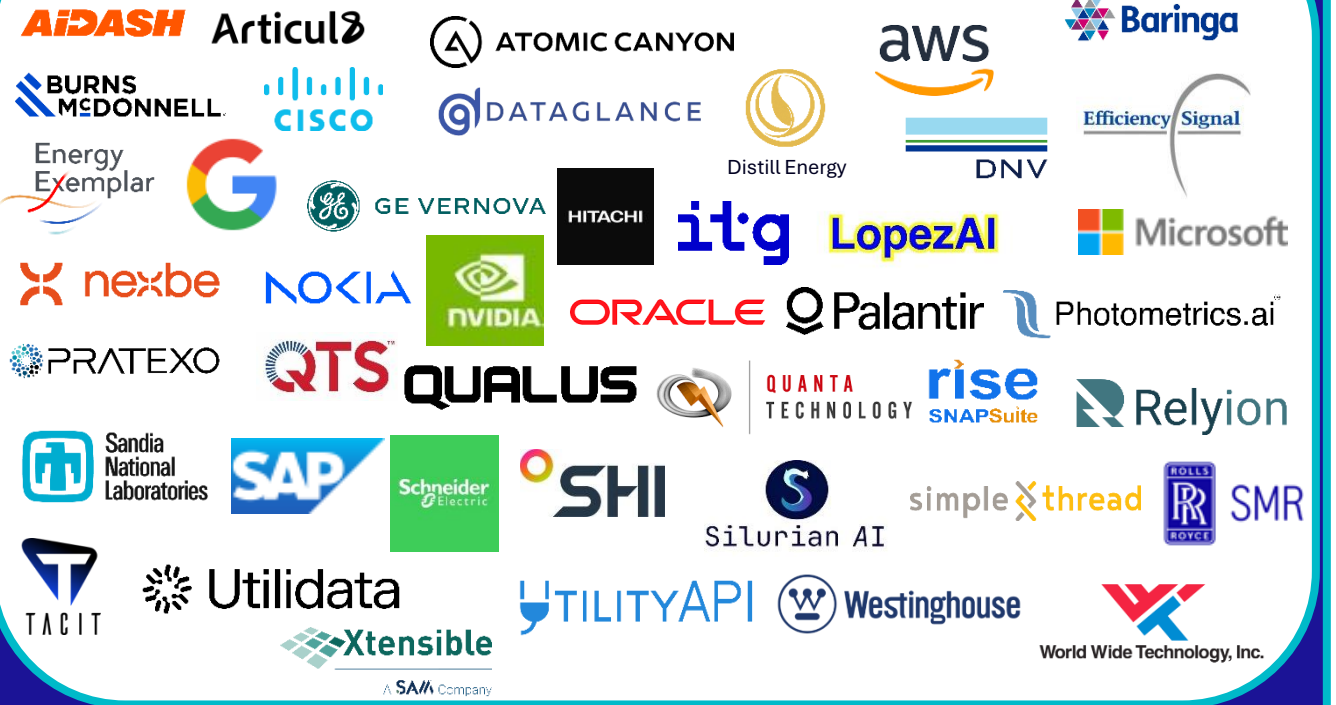
Many companies continue to join the consortium

>175 Confirmed Members

>125 MOU/Requests In-Process

OPAI Participants

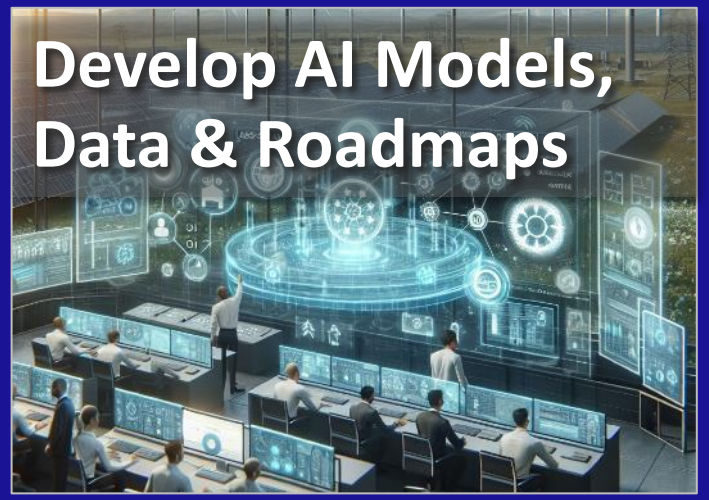
AI and Energy Technology Partners



Academia & Other Strategic Partners

Energy Partners

Open Power AI Consortium – Key Objectives



Create an ecosystem for stakeholders to identify, develop, validate and deploy AI solutions to transform the electric sector

Images created with Copilot

Open Power AI Consortium – Key Objectives

Implementation and Lessons Learned

- Collaborate with leading AI developers to deploy AI models.
- Develop a feedback loop from deployments for refinement and optimization.
- Maximize impact via knowledge sharing among consortium members.

Data Readiness White Paper

How can we properly leverage data to support AI initiatives?

The first in a series of guidance documents to accelerate AI implementation

Domain Specific Model Workstream

DSM Workstream Purpose and Goals

- **Workstream Purpose**
  - Develop and validate **Language Models purpose-built for the electric power sector**, capturing the terminology, data, and safety requirements unique to grid operations, planning, and compliance.
- **Goals**
  - **Advance Domain-Specific LLMs** that outperform general models on utility tasks:
    - Dispatch, forecasting, compliance, and operator assistance, etc
  - **Integrate AI safely into operational environments**, aligning with NERC CIP obligations.
  - **Leverage Open Power AI infrastructure** to accelerate evaluation and deployment.
  - **Collaborate across utilities, vendors, and labs** to build shared datasets, benchmarks, and evaluation frameworks.
  - **Demonstrate measurable value** in reliability, operator efficiency, and compliance automation.

What Is a Domain-Specific Model (DSM)?

- **Definition**
  - A Domain-Specific Model (DSM) is a large language model adapted for a focused industry or technical domain - in our case, the electric power sector.
  - It combines general-purpose language reasoning with specialized training data, tools, and constraints that reflect the grid's physics, terminology, and safety environment.
- **Why Domain Specificity Matters**
  - General LLMs lack grounding in grid physics, procedures, and codes.
  - Utilities require models that can reason deterministically, cite standards, and align with NERC CIP and operational safety policies.
  - DSMs provide expert reasoning + provenance, reducing hallucinations and regulatory risk.

How DSMs Are Built (Common Approaches)

- **Fine-Tuning on Domain Knowledge**
  - Train on manuals, standards, operational logs, and simulation transcripts (as in GAIA and A8-Energy).
  - Uses Parameter-Efficient Fine-Tuning (PEFT) methods like LoRA and QLoRA to minimize compute costs
- **Retrieval-Augmented Generation (RAG)**
  - Keeps a curated knowledge base of standards, procedures, or grid codes; retrieves relevant passages at query time (used in GridCodex, Grid Code Copilot).
- **Tool-Calling + Simulation Integration**
  - LLM outputs validated through deterministic solvers or digital twins (eGridGPT, GridMind) for safety-checked answers.
- **Multi-Agent and Chain-of-Reasoning Architectures**
  - Separate planning, validation, and explanation agents improve auditability (Grid-Agent, PowerAgent MCP).
- **Foundation-Model Layering**
  - Builds on shared grid encoders such as GridFM for physics-aware embeddings and time-series understanding

Why Now: The Need for Domain-Specific Models in Power Systems

- AI demand is reshaping the grid: rapid load growth from data-center and AI infrastructure is accelerating the need for intelligent, adaptive grid operations.
  - AI is now both a load and a solution
- Utilities face higher complexity: integration of renewables, distributed energy resources, and bidirectional flows is outpacing the capabilities of rule-based systems.
- Reliability & safety remain paramount.
- **Why General LLMs Aren't Enough**
  - Trained on internet data → lack domain grounding, technical vocabulary, and safety constraints.
  - Unable to reason with physics-based models or verify results against digital-twin simulations (as done in NREL's eGridGPT).
  - Risk of hallucinated recommendations and non-deterministic outputs → unacceptable for control-room or compliance use cases.
- **Opportunity for Domain-Specific Models**
  - Early testing shows fine-tuned LLMs can show 2–5× accuracy improvements on domain specific tasks and achieve verifiable reasoning

Current Landscape: Domain-Specific Models

- Established Domain Models
  - **GAIA (Grid Artificial Intelligence Assistant):**
    - Fine-tuned from LLaMA2 for dispatch, monitoring, and black-start operations.
      - Benchmarked on ElecBench, scoring highest on factuality & safety
    - https://www.nature.com/articles/s41598-025-91940-x
  - **NREL eGridGPT:**
    - Control-room copilot using digital-twin validation and human-in-the-loop review; CIP-aligned for secure on-prem deployment.
    - https://docs.nrel.gov/docs/fy24osti/87740.pdf
  - **A8-Energy (Articul8 + EPRI):**
    - Trained on 10k+ domain documents (≈ 400 k images, 230 k tables)
    - https://www.articul8.ai/blog/building-energy-domain-specific-gen-ai-models-that-reason-like-experts
  - **LFLLM (Load Forecasting LLM):**
    - Uses parameter-efficient fine-tuning (LoRA/QLoRA) for short-term load forecasting at multiple voltage levels.
    - https://www.techrxiv.org/users/712819/articles/697227/master/file/data/LFLLM__A_Large_Language_Model_for_Load_Forecasting20231231/LFLLM__A_Large_Language_Model_for_Load_Forecasting20231231.pdf

Current Landscape: DSM Frameworks

- **Open Frameworks & Agentic Ecosystems**
  - **PowerAgent Community (Harvard / Power & AI Initiative):**
    - Open toolkit for LLM-powered grid agents – includes PowerFM (models), PowerMCP (protocol), and PowerWF (workflows).
    - https://power-agent.github.io/
  - **ChatGrid (PNNL):**
    - LLM query interface that turns operator questions into grid visualizations.
    - https://www.pnnl.gov/news-media/chatgridtm-new-generative-ai-tool-power-grid-visualization
  - **GridCodex:**
    - RAG framework for grid-code and compliance reasoning
    - https://arxiv.org/abs/2508.12682
  - **GridFM (Linux Foundation Energy):**
    - Open-source foundation-model framework for the grid, led by IBM Research & Hydro-Québec; defines shared architectures and datasets for physics-informed, interoperable AI supporting dispatch, planning, and forecasting models
    - https://lfenergy.org/projects/gridfm/

Are there other DSM models or frameworks you're aware of or using?

Benchmarking Strategy for Domain Specific Models

- **Purpose**
  - Establish a standardized, transparent framework for evaluating power-sector Large Language Models across accuracy, safety, and operational reliability dimensions — ensuring all DSMs can be compared on equal footing.
- **Why Benchmarking Matters**
  - Creates quantifiable proof that DSMs outperform general LLMs on grid tasks.
  - Enables cross-utility trust through reproducible, open-standard metrics.
  - Anchors all pilots and MVPs to a common evaluation language
- **Existing Benchmarks**
  - **ElecBench**
    - First comprehensive benchmark tailored for power dispatch and operations.
    - Evaluates six primary metrics: Factuality, Logicality, Stability, Security, Fairness, Expressiveness, with 24 sub-metrics
    - https://github.com/xiyuan-zhou/ElecBench-a-Power-Dispatch-Evaluation-Benchmark-for-Large-Language-Models
  - **EPRI Benchmarking Question Bank**
    - Over 2,000 questions tailored to measure knowledge across the power industry.

Domain Expanded Benchmarks Under Development

Category	Model / Framework Source	Benchmark Goal
Forecasting	LFLLM	Extend ElecBench with load & renewable forecasting tasks.
Compliance Reasoning	GridCodex	Evaluate accuracy of regulatory and grid-code Q&A (citation coverage).
Operational Decision Support	eGridGPT / GridMind	Measure reasoning with digital-twin verification and safe action recommendations.
Time-Series Foundation Models	GridFM / PowerGPT	Benchmark physics-informed encoders for anomaly detection and scenario modeling.

Other benchmarking considerations we missed?

Regulatory & Assurance Guardrails

- **Purpose**
  - To ensure all Domain-Specific Models (DSMs) are safe, auditable, and compliant with existing electric-sector reliability and cybersecurity regulations.
- **Key Regulatory References**
  - NERC CIP-015-1 (Internal Network Security Monitoring)
    - Approved by FERC June 26, 2025; effective Sept 2 2025.
    - Requires internal network monitoring within ESPs and directs NERC to extend coverage to EACMS and PACS assets.
    - Implication: any LLM or AI component connected to EMS/DMS networks must provide telemetry, logging, and segmentation evidence.
  - NERC CIP-005 & CIP-007 – Electronic Security Perimeter & System Security Management
    - Define network segmentation, patching, and malware detection—critical for LLM hosting environments.
  - DOE AI Risk Management & INL Assurance Framework
    - Establishes expectations for AI safety cases, documentation, and independent review before operational use.
  - EPRI Open Power AI Safety Charter
    - Commits all consortium models to be: Transparent, Testable, Traceable, and Tamper-resistant.
- **Assurance & Safety Controls**
  - Human-in-the-Loop Review – no autonomous actions; operators confirm all LLM recommendations.
  - Read-Only Interfaces – initial MVP deployments restricted to non-control data paths.
  - Sim-in-the-Loop Validation – all action recommendations checked via digital twin or physics solver before surfacing.
  - Provenance & Citation Logging – every model output must link back to a source or dataset version.
  - Red-Team & Regression Testing – unbiased evaluations expanded to include security, compliance, and fairness dimensions
  - INSM Telemetry Integration – real-time visibility of model queries and responses within monitored network zones.

Are there regulatory or assurance guardrails missing or other considerations this group should consider?

Uses Cases and MVPs

- What are the highest-value, lowest-risk pilots to prove DSM capability in real utility environments?
  - Examples:
    - Control-Room Copilot
    - Outage & Work-Order Triage
    - Forecasting Assistant
    - Cyber Event Summarizer

What are the highest-value, lowest-risk pilots to prove DSM capability in real utility environments?

Proposed Workstream Research Output

- Possible papers, deliverables, and outputs from the work stream could include the following:
  - Papers or guidance
    - Deployment playbooks
    - Implementation best practices
  - Websites or wiki pages
    - Success stories
    - FAQs
  - Domain Specific Models
    - Available to download from sites like Hugging Face or NVIDIA

Outreach and Knowledge Transfer

- Periodic Web Meetings
  - Target monthly
  - 1 – 1.5 hours depending on content
  - DSM presentations – commercial and internally developed
  - Implementation lessons learned
- OPAI Workshops
  - Implementation sessions and demonstrations

Workstream Next Steps

Building a Web Meeting Schedule

OCTOBER

Su	Mo	Tu	We	Th	Fr	Sa
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Work Stream Kickoff

NOVEMBER

Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

Commercial Tool Intro
Utility Implementation Story
1.5 hr Meeting

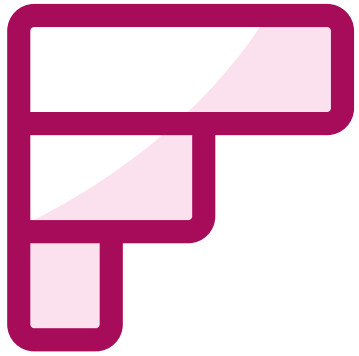
DECEMBER

Su	Mo	Tu	We	Th	Fr	Sa
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

Commercial Tool Intro
1 hr Meeting

- Implementation Work Stream Virtual Meeting
- Use-Case Work Stream Virtual Meeting
- Domain-Specific Model Virtual Meeting
- OPAI Executive Advisory Group
- OPAI Member Representative Committee
- EPRI Holiday

Call for volunteers to present.
Send invitations for remaining 2025 meetings.
Develop 2026 schedule over the next months.



Favorite Day of the Week for DSM meetings?

Questions?

TOGETHER…SHAPING THE FUTURE OF ENERGY®

www.epri.com