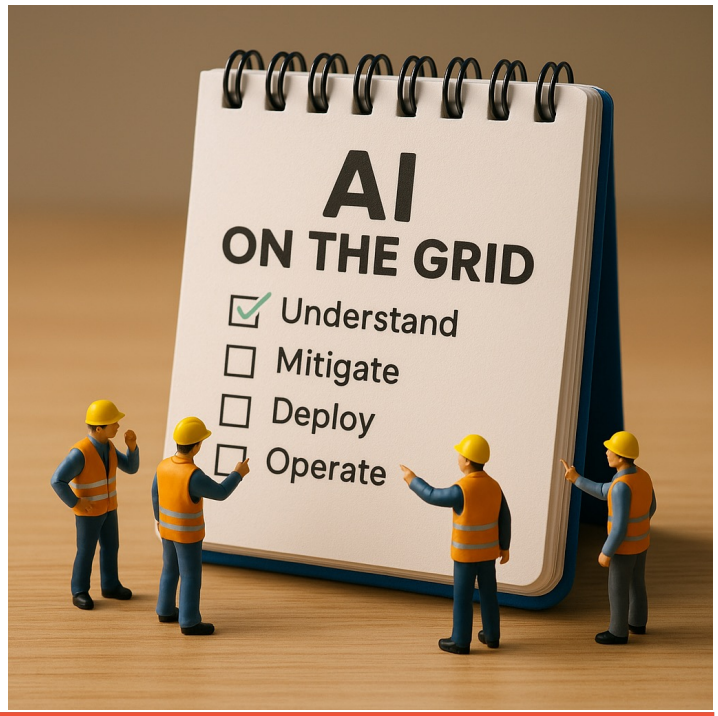
AI on the Grid: What's Here, What's Hidden, and What's Next

Matthew E. Luallen
Information Trust Institute
University of Illinois, Urbana-Champaign
mluallen@illinois.edu

AIEC Cooperative Technology Conference 2025

Agenda

- Lightning strikes then and now.
- What is here and why does this matter now?
- What's hidden among the AI landscape
- How can we use AI for real world tasks
- An invitation



The Flicker Heard Round the Network

⚡ Storm → 6 power flickers in 60 seconds

📡 Starlink Gen 2 reset to factory defaults (no password!)

🔥 Firewalla required firmware reflash

🎯 Key lessons:

- ~~UPS~~, physical events trigger cyber vulnerabilities (and vice-versa)
- AI missed this dependency

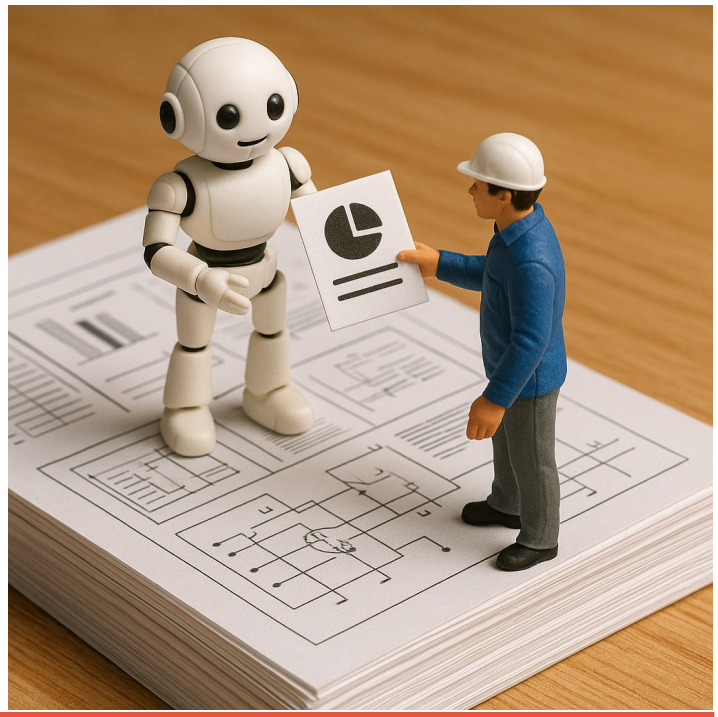
Why This Matters to Co-ops

🏭 Co-ops face same risks as major utilities and current innovations

💰 But with fewer staff and tighter budgets

🤖 AI promises help if designed intentionally

📊 While needing systematic approaches, not ad-hoc fixes



What's Here Today

🔧 Predictive maintenance (sensors + AI analytics)

📈 Demand forecasting and load balancing

☀️ Distributed Energy Resource (DER) optimization

🛡️ Threat detection (log analysis, anomaly detection)

📦 Already embedded in many vendor solutions

AI may be a team member that can do the mundane, repetitive jobs so that you can focus on what matters most

What's Hidden

🔍 AI embedded in vendor products (often undisclosed)

☁️ Cloud-based tools quietly running AI models

⚠️ Insecure default settings (training data, unknown model cards)

🔒 Offline vs. online AI trade-offs

🌐 Data sovereignty and privacy concerns

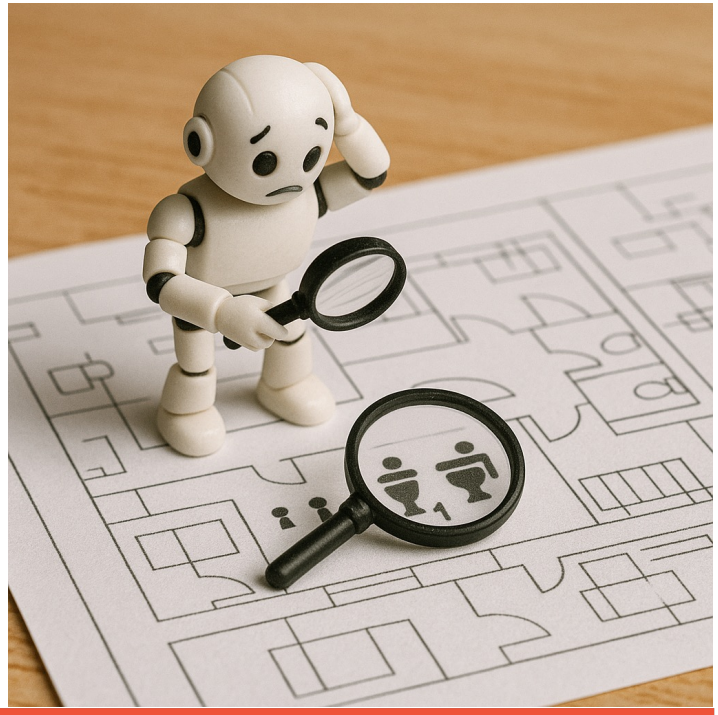
Multimodal Challenges in the Real World

🤔 IMECE Case: LLMs struggle with diagrams, schematics

🔋 BMS: DERs need continuous awareness and wide contextual analysis, ~~hazards~~

📊 OpenAI GDPval: Benchmarking AI on real tasks

🎯 Lesson: AI must be evaluated in its operating environment



AI Democratizes Genius, Not Context (at this time)

🧠 AI provides unprecedented problem-solving power

❓ But currently lacks situational and operational awareness

👨‍🔧 Cannot yet replace domain expertise and judgment

🤝 Humans remain the context, hands, arms, and legs

⚡ Partnership model: AI as force multiplier, not replacement

Addressing Weaknesses at Scale

🐛 Thousands of vulnerabilities across systems

🔨 Cannot fix bugs one-by-one (whack-a-mole fails)

📐 Need systematic, scalable approaches

🎯 Focus on high-impact, common weaknesses

🤖 Leverage automation and AI for multimodal operations and defense

- Linguistic and non-linguistic AI

MITRE CWE Top 25 (2024)

🥇 #1: Cross-Site Scripting (XSS) - CWE-79
🥈 #2: Out-of-Bounds Write - CWE-787
🥉 #3: SQL Injection - CWE-89

📊 These represent the 'greatest hits' of software weaknesses
💰 Focusing here = biggest defense ROI

OT/ICS Weaknesses (CWE View 1358)

(based on SEI-ETF whitepaper)

- 🔓 CWE-1366: Frail Security in Protocols
- ➖ CWE-1364: Zone Boundary Failures
- 📦 CWE-1372: Supply Chain Corruption
- 🔀 CWE-1369: IT/OT Convergence Frailties

⚠️ Higher stakes: OT failures → physical damage

People & Process Weaknesses

- 🤝 CWE-1373: Trust Model Problems
- 👥 CWE-1374: Maker-Breaker Blindness
- 🚧 CWE-1376: Security Gaps in Commissioning
- 👤 CWE-1379: Human Factors

- 🎓 Culture & training as critical as technical controls

Fail-Safe Design Principles (AI here?)

🛑 Systems will fail - design for safe failure modes

🔒 Default to secure state when errors occur

🚗 Mechanical analogies: emergency brakes, safety valves

💻 Cyber fail-safes: default deny, lockdown on anomaly; or fire detected opposite example

⚡ Example: <insert situation> caused (open/closed) **default (fail-safe)** condition.

AI as Defensive Sidekick

🔍 Automated vulnerability detection at scale

⚡ Accelerated code analysis and patching

📊 Anomaly detection in logs and network traffic

⚡ Anomaly detection in voltages and currents

🎯 DARPA programs: INGOTS for automated vuln discovery

🤖 AI as tireless analyst, not replacement (at this time). All hazards analysis is paramount

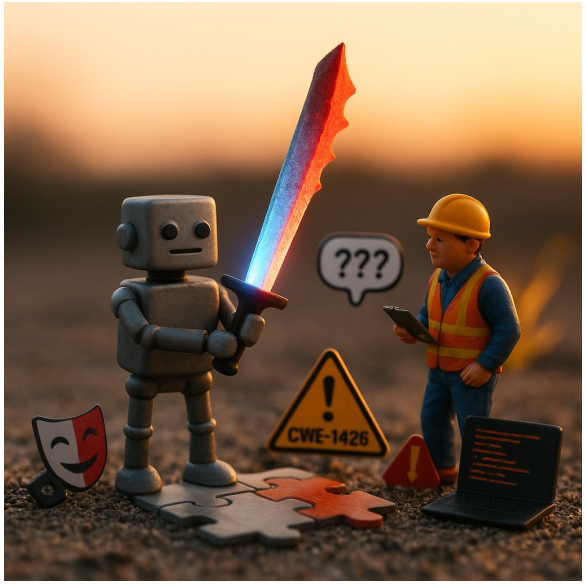
The Double-Edged Sword of AI

🎭 Attackers leverage AI for automated reconnaissance

🤔 Model hallucinations and false positives

⚠️ CWE-1426: Improper Validation of Generative AI Output

💻 AI-generated code may introduce new vulnerabilities



🎯 At this time, always validate AI outputs, never trust blindly

MITRE Caldera: Adversary Emulation

- 🎭 Automated adversary emulation platform
- 🎯 Simulates real attack scenarios safely
- ♻️ Scalable, repeatable security testing
- 🔧 Open-source and extensible
  - OT plugins available
  - Student assignment this semester

- 🚨 Think: Fire drill for cyber attacks
  - (RECIPE Coalition)

MITRE ATLAS: AI Threat Landscape

📚 Knowledge base of AI/ML system attacks

🎯 Tactics: Data poisoning, model theft, prompt injection

👻 Evasion attacks (adversarial examples)

🌐 Community-driven, continuously updated

🛡️ The ATT&CK framework for AI security

CWE AI Working Group

📖 Expanding CWE catalog for AI-related risks

🆕 Recent additions: CWE-1426 (Generative AI Output)

🤝 Community-driven standardization effort

💬 Enables clear communication about AI vulnerabilities

🎯 Critical for tool integration and vendor accountability

Metrics That Matter (not AI, Human response)

📊 Patch Management Compliance Rate (30-day window)

👥 Security Awareness Rate (employees, contractors, suppliers)

🚨 Incident Response Preparedness (drill frequency)

💾 Critical Function Backup/Restore Success Rate

📣 Active Event Notification Coverage

✅ These are scorecards you can actually use

Call to Action: Tomorrow's Priorities

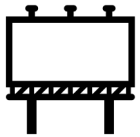
✓ Ask vendors: Are you using memory-safe languages, and AI to address your weaknesses?

✓ Demand transparency in AI tool usage

✓ Implement fail-safe design principles

✓ Apply Cyber-Informed Engineering (CIE) principles

✓ Confirm at least the “Metrics That Matter” as accountability framework



Mini-demo (if time permits, and room allows it)

Building the Future Together

- AI democratizes problem-solving genius
- But humans provide the context and judgment
- **Intentional design beats bolt-on security**
- Co-ops can lead in secure AI integration

Let's collaborate: ITI, ReCIPE, or …

- Matthew E. Luallen
  mluallen@illinois.edu