Power Delivery & Utilization

TECHNICAL BRIEF

Cyber Security for Power Delivery and Utilization, P183

UTILITY CYBER SECURITY AND ARTIFICIAL INTELLIGENCE CHALLENGES AND OPPORTUNITIES

Data Management Will Be Key to Successful and Secure AI Adoption

THE CHALLENGE: SECURELY LEVERAGING AI FOR UTILITY CYBER SECURITY OPERATIONS

Artificial intelligence (AI) has seized people's attention with recent introductions of accessible generative AI tools such as ChatGPT, Bing AI, and Dall-E2. Reactions have been mixed—amazement, excitement, trepidation, and concern. EPRI's Energy Delivery and Customer Solutions (ED&CS) Program 183 cyber security team asked two important questions:

- Can AI effectively aid operational technology (OT) security programs and operations for utilities?
- What security impacts should utilities anticipate from AI?

Full answers will take time to develop, but certain conclusions can be drawn from ongoing EPRI research that may surprise you. There is no doubt that this topic is worthy of continued research to identify opportunities and challenges in deploying AI-enabled applications in security operations and securing AI solutions and services for utilities.

AI is both an opportunity and a risk for utility cyber security programs and operations. Across all industries, global investments in AI for cyber security are growing. A recent Stanford University study noted that the investment reached almost $4B USD in 2022,[4] more than double the 2021 total. The same report noted that cyber security is considered the greatest risk to mitigate in AI adoption.

AI automates the ingestion of data at a large scale—making it "big automation." AI tools will help manage the growing volumes, velocities, and varieties of data also known as "big data" found in security operations centers. AI holds promise to improve defensive capabilities of cyber security operations that rely on sifting through multi-system data to detect anomalous patterns of activity. It also presents possibilities of automating routine and lower value tasks to improve scarce cyber security resource productivity and fill workforce gaps in security operations. As solution vendors integrate more AI-enabled capabilities into their products, expect to see more autonomous capabilities emerge with a caveat: Utilities must trust the systems to make unsupervised decisions impacting cyber security operations. At the same time, AI will require intensive security controls of data and comprehensive data governance to ensure that it earns the confidence of utilities to evolve from AI-enabled support of human decisions to fully autonomous AI-driven decisions.

> "The challenge with AI is fundamentally trust. How do we know it works to focus human capacity elsewhere? **Fundamentally it is AI until we trust it, then it becomes automation.**"
>
> **DAVID REBER ,** *Nvidia Chief Security Officer*

4 2022 AI Index Report can be downloaded at https://aiindex.stanford.edu/report/

Table 1. *AI creates new data risks and needs for mitigations*

RISK	MITIGATION
Compromised external data for training models	• Supply chain controls • Data lineage controls
Compromised internal data for security systems	• Data integrity controls • Data lineage controls
Compromised hardware and software	• Supply chain controls • Secure remote access controls

AI AND DIGITAL TRANSFORMATIONS IN CYBER SECURITY OPERATIONS

Digital transformation is one of four metatrends in EPRI's *Cyber Security Roadmap for 2030*[5] that can help build intrinsic cyber security for utilities. Digital transformation emphasizes the importance of data to automate the entire security life cycle, whether for risk management, incident response, or program performance. Digital transformation means that more data about asset conditions, functional domains, and adversaries can inform your security operations performance and investment decisions. However, that data must be well-curated and of highest quality as well as accessible to authorized stakeholders to ensure that security decisions can be made with confidence. These conditions may be addressed through comprehensive and domain-sensitive data governance policies and data management practices.

AI, like all advanced analytics capabilities, has an extremely strong reliance on trustworthy data to deliver decision-making confidence. The National Institute of Standards and Technology (NIST) recently published a document about the establishment of trustworthy AI[6] that is broadly applicable to all business sectors. AI adheres to the old principle of "garbage in, garbage out." An AI-enabled solution or service bases its analytical capabilities on the data it has. That means AI requires active and ongoing data management for cyber security operations to obtain the most value from data.

Here are two examples about trustworthy data to illustrate the point about managing data to optimize value. The first scenario explores the training data for an AI-enabled system. AI algorithms are combinations of different statistical analysis formulas and are trained on data sets to help improve their probabilities calculations. A training data set that has too limited an amount of data to effectively enrich statistical analysis will not deliver excellent data value. Data sets that are populated with outdated or inaccurate data also impair the outcomes of AI systems—just as would occur with traditional security solutions. If the training data set is deliberately compromised—a condition known as *data poisoning*—the consequences can be severe. The integrity of the data supply chain and provenance of the data, whether externally or internally sourced, must be managed from solution vendor to end user to help mitigate this significant AI vulnerability.

The second scenario focuses on the utility data required for OT cyber security operations. EPRI has direct experience working with utilities to ingest data into statistical formulas that deliver quantitative security performance metrics. Research is underway to learn how utilities define data governance and structure data management for OT cyber security data. Initial results lead EPRI to conclude that there are knowledge gaps that can hinder the optimal management of data for advanced analytics applications in OT cyber security programs. This work can also build foundational knowledge about data classifications for authorized access and use as well as how to structure processes for data lineage—an important prerequisite for AI trustworthiness. This research can help identify the appropriate mitigations for the risks that AI technologies and applications may introduce into utility operations. Table 1 above identifies several AI risks and potential mitigations that could be enacted by utility cyber security and data governance teams.

Current data governance and management practices have served utilities well until now, but digital transformations—particularly those that deploy AI—will require a new approach to data management to address new security risks and the increased complexities of big data.

DATA AS AN ASSET IN CYBER SECURITY OPERATIONS

Data is an asset that requires comprehensive asset management, governed by data policies and practices that deliver the appropriate data classifications and security controls to meet today's objectives and tomorrow's digital transformations. It is also an asset that must be managed to ensure trustworthy data for advanced analytics including AI. Excellent data management helps utilities address the growing volumes, varieties, and velocities of data and the complexities of meeting business objectives, stakeholder needs, and regulatory requirements. Comprehensive and holistic data governance can help utilities realize the full value of data within their operations, including cyber security operations. Data governance that reflects the unique needs of cyber security

5 *Cyber Security Roadmap for 2030* can be downloaded at https://www.epri.com/research/products/000000003002017753

6 NIST's *Artificial Intelligence Risk Management Framework* can be downloaded at https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

operations will properly classify data and engineer controlled access to data that is important to cyber security operations or specific functions such as forensics or data lineage.

EPRI experience and literature scans reveal some common themes about AI deployment across all business sectors that may surprise you. The top success indicator is the right data governance. Companies that successfully deployed AI applications enacted data governance policies and managed data as an asset to maintain data security, trustworthiness, and compliance. The importance of the right data governance policies cannot be overstated. Gartner observed that 80% of organizations that seek to grow digital business will fail because of misguided data governance strategy. A second success factor is data collection. Determining what is in the cyber security data domain is important, along with identifying who owns data and how access is defined and managed. EPRI's metrics research confirms the importance of this success factor. The third success factor is consistent data quality. Organizations with trusted data quality conduct regular monitoring as part of their overall data management practices to ensure consistent quality.

EPRI's research will continue to investigate the most effective data governance policies and data management practices to support digital transformations of OT cyber security operations. Figure 1 illustrates a progression of decision capacities in advanced analytics applications that can help utilities begin digital transformation within their cyber security operations.

EPRI metrics data[7] has been successfully gathered and used to support data-driven decisions by utilities. EPRI metrics calculations require a fraction of the data volume that is typically used by AI-enabled applications but have similar requirements for data availability, quality, and consistency. Deployment of EPRI's cyber security metrics may be a useful indicator of data readiness (as well as data literacy and culture) for utilities contemplating AI adoption in their cyber security operations.

User entity and behavior analytics (UEBA) is an application that can aid in the detection of insider threats based on changes in normal patterns of use. It analyzes large quantities of data to detect anomalies or outliers in behaviors and may include risk assessments to generate security alerts if certain thresholds are exceeded. It is a semi-autonomous AI application because it presents information for human action. UEBA requires training data, and that data supply chain must be managed to ensure trustworthiness. Semi-autonomous applications such as UEBA are a stepping stone to more autonomous cyber security systems.

A final stage in digital transformation is the adoption of autonomous applications that harness trusted data sets to train AI systems to recognize threats and take actions without human intervention. Trustworthiness is an absolute requirement—and utilities will be tasked to ensure that data governance and management integrate data lineage and other supply chain controls into training data sets and other vectors for data poisoning. Automated threat mitigation (ATM) is a longer term goal for defensive cyber security and will likely be an essential future tool to help utilities manage dynamic threat landscapes with limited skilled workforce resources.

There are other digital transformation options as well—see the sidebar "Exploring Natural Language Processing (NLP) in Vendor Documentation Management." This EPRI research activity examines the possibilities of melding unstructured text and machine learning to improve secure intelligent electronic device (IED) management.

7 Data collection for cyber security metrics is challenging because of legacy OT environments. Legacy equipment constraints are well known to utility cyber security professionals and generally require equipment replacement or upgrade to resolve.

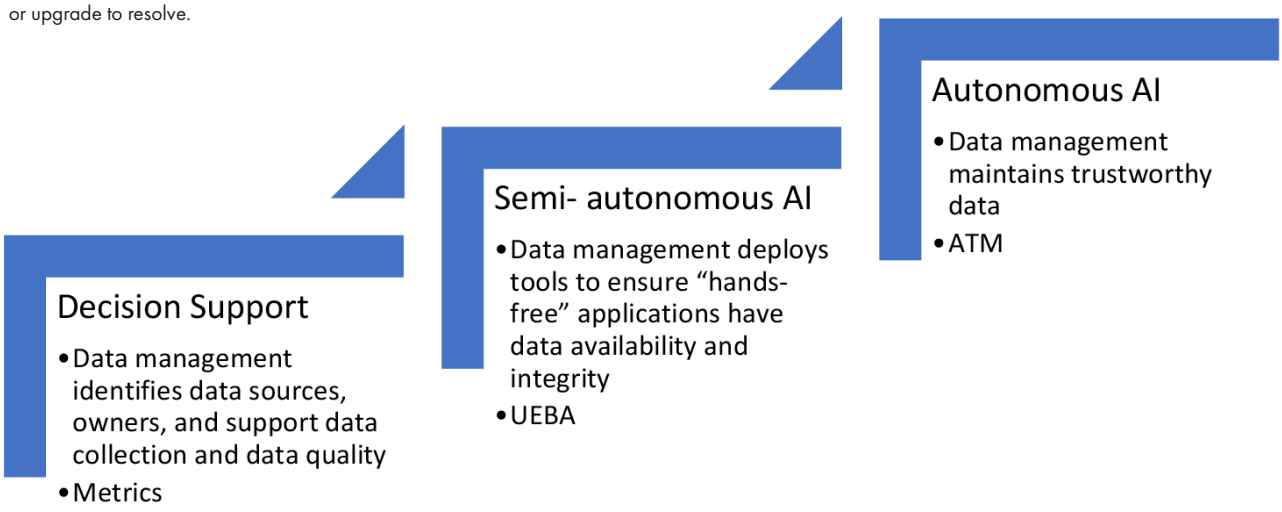

Figure 1. *The evolution of advanced analytics for utility cyber security operations*

CONCLUSIONS AND ACTIONS FOR CYBER SECURITY OPERATIONS

AI is certainly overhyped at the moment, but we see legitimate use cases for utility OT environments and continued growth in information technology (IT) environments. OT cyber security use cases may be a couple years down the road, but even some of the "mothers and fathers" of modern AI are astonished at the speed of innovation and adoption of advanced analytics such as generative AI and other structured and unstructured deep learning systems.

There is a caveat, however. Data readiness is a significant factor in the successful adoption of any advanced analytics application for cyber security. Utilities have an opportunity to determine their data readiness and identify policy, process, technology, and workforce gaps by deploying EPRI's metrics capabilities in our Cyberjoule™ platform. This action delivers twofold benefits: quantified and consistent performance measurements for security operations and identified next steps in data governance and data management updates to be ready for digital transformations accelerated by AI.

CONTACT INFORMATION

For more information, contact the EPRI Customer Assistance Center at 800.313.3774 (askepri@epri.com).

EPRI RESOURCES

EPRI members interested in engaging in and supporting this effort should contact EPRI for further discussion.

Christine Hertzog, ***Principal Project Manager***
650.314.8111, chertzog@epri.com

Cyber Security for Power Delivery and Utilization, P183

EXPLORING NATURAL LANGUAGE PROCESSING (NLP) IN VENDOR DOCUMENTATION MANAGEMENT

Intelligent electronic devices (IEDs) are deployed for protection and control functions in substation environments. The current methods of managing IEDs require in-depth knowledge of these devices by utility resources and, given cyber security workforce scarcity issues, are unsustainable and a growing security risk. Disparate IED vendor documentation and data formats complicate security professionals' efforts to establish a consistent security approach across all devices. Natural language processing (NLP) is a machine learning technology that can ingest unstructured text such as vendor documentation and reorganize important information into a consistent structure to enhance its accessibility and enable quick comprehension. EPRI's research—conducted through the Secure IED Management Strategies project [4] — is investigating if NLP processing of different vendor documentation may enhance the efficacy and efficiency of security efforts, improve workforce productivity, and help ensure that systems and devices are secure and compliant with established standards. Our research will learn if NLP-structured information may help:

- Assess security management needs and risks associated with individual devices
- Validate compliance artifacts
- Execute queries for equipment troubleshooting
- Create more granular models of risk from the device level up to the systems and ultimately to the grid

4 A project description can be downloaded at https://www.epri.com/research/products/000000003002022701

About EPRI

Founded in 1972, EPRI is the world's preeminent independent, non-profit energy research and development organization, with offices around the world. EPRI's trusted experts collaborate with more than 450 companies in 45 countries, driving innovation to ensure the public has clean, safe, reliable, affordable, and equitable access to electricity across the globe. Together, we are shaping the future of energy.

3002026339 March 2023

EPRI
3420 Hillview Avenue, Palo Alto, California 94304-1338 USA • 800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com

© 2023 Electric Power Research Institute (EPRI), Inc. All rights reserved. Electric Power Research Institute, EPRI, and TOGETHER...SHAPING THE FUTURE OF ENERGY are registered marks of the Electric Power Research Institute, Inc. in the U.S. and worldwide.