



CESER
SUMMARY REPORT

U.S. DEPARTMENT OF
ENERGY

Office of
Cybersecurity, Energy Security,
and Emergency Response

Potential Benefits and Risks of Artificial Intelligence for Critical Energy Infrastructure

April 2024



U.S. DEPARTMENT OF
ENERGY

Office of
Cybersecurity, Energy Security,
and Emergency Response

SUMMARY REPORT:
Potential Benefits and Risks of Artificial
Intelligence for Critical Energy Infrastructure

THIS PAGE INTENTIONALLY LEFT BLANK.



Overview

Artificial intelligence (AI) has the potential to help build an energy sector that is safer, cleaner, more efficient, and more secure than ever before – a growing opportunity, highlighted by recent technical advances. However, as with all emerging technology, AI can cause harm if poorly implemented, insufficiently understood, or exploited by our adversaries.

While various forms of AI are already in use across the energy sector, advances in technology (including the emergence of generative AI) are driving a rapid expansion of the deployment of AI capabilities and tools. As their use expands, so does the potential that malicious actors might seek to either target energy sector AI systems directly, or use AI to enhance attempts to attack our critical energy infrastructure.

As the U.S. looks to harness the power of AI to reshape critical energy infrastructure and secure lives, it is crucial that we navigate this emerging technology with a keen, technically-grounded and risk-informed awareness of its potential and pitfalls.

As directed by Executive Order 14110, ***Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence***, the U.S. Department of Energy – the Sector Risk Management Agency for the U.S. energy sector – produced an interim assessment that identifies the potential benefits of AI use in the energy sector, as well as key sources of risk to the sector.

The assessment analyzes how risks can arise in applying AI to energy infrastructure and the potential consequences that can result. The assessment also provides key findings and key recommendations, highlighting the most salient conclusions regarding benefits and risks, as well as potential steps that could maximize the former and minimize the latter. The assessment is rooted in taxonomies of AI types and of their applications to energy infrastructure – looking at those presently in use, on the horizon, and in the far future.

The assessment was led by the Office of Cybersecurity, Energy Security, and Emergency Response (CESER), supported by machine learning researchers and energy systems engineers from Lawrence Livermore National Laboratory (LLNL), on behalf of the Department. It was developed in collaboration with energy sector owners and operators, technical subject matter experts, and interagency partners – and informed by multiple DOE National Laboratories and offices.

Given the 90-day timeframe provided to develop the interim assessment, it is not a comprehensive evaluation of the benefits and risks of AI use in the energy sector – rather, it offers an initial survey of the topic, to help orient energy sector readers and highlight areas of further interest. CESER intends to release an update to this assessment in towards the end of 2024.

This summary provides an overview of the assessment’s findings, highlights some of its recommendations, and lays out steps for further analysis and engagement that the Department is planning to undertake.



Potential Applications and Benefits

Millions of Americans use AI and machine learning, each day – often baked into aspects of our daily lives in ways we don’t even realize. Increasingly, AI is also being deployed across the energy sector, in various forms, enabling the supply of energy across the nation.

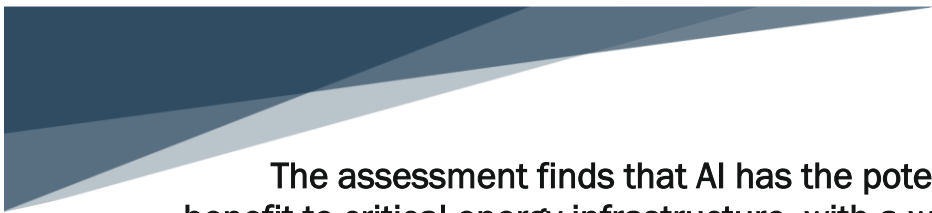
The assessment first identifies a set of ten broad categories of AI applications for critical energy infrastructure, examining each across three timeframes: current applications, applications that are on the horizon, and applications that are in the far future. This is not an exhaustive list, and new applications will almost certainly emerge as energy and AI advance.

1. **Infrastructure Operational Awareness** – In the face of the flood of data generated by modern energy infrastructure, AI is helping system operators identify key information in real time – giving them a clearer view of their systems. At the same time, its inference capabilities can help rapidly characterize changes in the system status, even with limited or incomplete data – giving operators the awareness and context they need to respond.
2. **High-Complexity Modeling & Simulation** – AI-based models can make more sophisticated weather modeling and energy system simulation more accessible, enabling better-informed decisionmaking. AI can increase the efficiency of commonly-used physics-based models, while also enabling inference-based approaches with superior performance.
3. **Active Controls** – There is significant interest in the promise of AI’s ability to control energy system operations at machine speed, informed by human-level inference. AI could support the real-time control of energy infrastructure, whether by providing decision support to human operators (*AI-assisted*, with *human-in-the-loop*), or by directly controlling infrastructure operations (*AI-directed*) – with varying levels of human involvement (either supervised by a *human-on-the-loop*, or operating autonomously).
 - a. Based on input from energy sector entities and experts, the assessment indicates that autonomous AI systems are not likely to be implemented for active control in critical energy systems, in the near term, given the potential consequences of any errors. However, AI-assisted, human-in-the-loop decision support systems appear to be a promising alternative, and active control remains a key area for further efforts.
4. **Predictive Maintenance** – AI can provide operators with enhanced, earlier warnings of potential energy equipment degradation or failure. This could allow operators to prioritize the equipment most in need of maintenance, improving reliability and preventing costly failures before they occur. AI-based predictive maintenance has already been deployed to support everything from wind turbines to oil and natural gas compressors/pumps, and offers great potential for battery electric storage systems and distribution transformers.
5. **Anomalous Event Detection & Diagnosis** – AI can help identify non-malicious anomalies (arcs, pipe freezes, faults, oscillations, trips, etc.) in real time – potentially helping operators prevent issues from escalating, or cascading across the system. Using a wide range of AI methods, both historical information and unstructured data can be used to build a better model of what an “anomaly” looks like for a given energy system.
6. **Malicious Event Detection & Diagnosis** – AI can provide new and enhanced ways to detect, identify, and respond to both known and novel malicious activity – whether



physical or cyber – occurring in energy systems. One promising application is in enabling machine-speed analysis of operational technology (OT) and information technology (IT) data, in an effort to identify, detect, and mitigate cyber intrusions into energy infrastructure before they cause an impact.

7. **Forecasting** – AI can enhance prediction of key operating factors like weather conditions and market prices. This has the potential to unlock financial benefits, improve preparedness for extreme weather, and enable the integration of diverse resources and loads – increasing resilience and helping advance the energy transition.
8. **System Planning** – AI can offer novel capabilities to support planning for energy system operations, including changing equipment and resource mixes, as well as the long-term deployment of new infrastructure. It can help accelerate the identification of optimal system configurations – potentially helping lower costs and improve system resilience, even while integrating both new and existing infrastructure in more complex ways.
9. **Resource Exploration & Extraction** – AI can help leverage historical and unstructured physical data to identify potential deposits of fossil and mineral resources, while also enabling advanced robotics/machinery that can facilitate extraction.
10. **Scenario Generation** – AI can help generate synthetic-but-realistic system data that can be used to train system operators on emerging energy technology, enable testing of protection and mitigation measures, and can allow system operators to better leverage the benefits of more distributed, flexible energy systems by bounding uncertainties.



The assessment finds that AI has the potential to be of tremendous benefit to critical energy infrastructure, with a wide range of benefits that can dramatically improve nearly all aspects of the sector – including security, reliability, and resilience. However, the path to realizing these benefits reveals the clear need for regularly updated, risk-aware best practice guidance to facilitate the safe, secure, and beneficial deployment of AI in critical energy infrastructure.



Potential Risks

The assessment finds that while a number of significant risks exist if AI is used or deployed naïvely, most risks can be mitigated through best practices, putting appropriate protections around important data and models, and in some cases, funding further research on mitigation techniques.

Application areas in which an AI might directly cause changes to physical processes can lead to higher risk, but human supervision of AI can often mitigate the most significant risks in that case as well. Regardless of whether AI technologies are deployed in service of energy infrastructure, the potential remains that hostile actors who wish harm to the sector may find ways to leverage AI to their own ends.

The assessment also considered areas where not leveraging AI in service of energy systems might leave potential benefits unrealized – or gaps in capabilities unfilled. It also considered the impacts that various risks could have on the energy system applications described above. These range widely: from loss of operational efficiency, to mistakes driven by overreliance on AI decision support tools, to (in certain cases) the potential misoperation or failure of energy infrastructure or its components.

While DOE's analysis identifies a set of four risk categories, as AI and its energy sector applications continue to evolve, it is expected that these categories will change.

Risk Category 1: Unintentional Failure Modes of AI

This category refers to AI created for beneficial purposes, but which is unintentionally misused or has unintentional failures, leading to negative outcomes. This report further breaks this category down into four failure modes:

- **Bias** in AI is the systematic shift of a decision-making process away from its goal, usually due to a mismatch between training data and real-world use. When applied to energy systems, training data that presents an incomplete picture of energy infrastructure based on limited sensor data could skew an AI model's replication of system behavior.
- **Extrapolation** is the use of a model to make predictions about “unexpected” events – situations outside that model's experience – which can lead to unpredictable or inaccurate behavior. For example, when confronted with an extreme weather event beyond its training, an AI model that informs the use of energy storage resources might over/under commit capacity or attempt to rely on resources impacted by the event.
- **Misalignment** is when an AI model's behavior deviates from the goals of its designers, typically due to poorly aligned training data or poorly defined objectives. Without robust and scalable human oversight, AI decision-support tools that are misaligned may end up prioritizing economic gain over, for example, grid reliability. This risk is especially pronounced when energy systems are under stress, as during extreme weather events.
- **Energy Use of AI** is a slightly different risk than the others presented here – not AI errors, but the implications for the energy system from the energy consumption of training and using large AI models. This is an area where significant further research can help characterize the shape and scale of expected AI load, identify opportunities for efficiency gains, and explore how AI deployment trends can help drive resilience and environmental benefits.



Regardless of the energy system application, it is difficult to eliminate these risks entirely – but potential challenges may be addressed through the applications of best practices and/or further research into mitigations.

Risk Category 2: Adversarial Attacks Against AI

Machine learning-based AI systems are susceptible to a variety of novel vulnerabilities, in addition to traditional cybersecurity vulnerabilities. These vulnerabilities can be exploited by an adversarial attack – which occurs when AI designed and deployed for beneficial purposes is intentionally manipulated by adversaries, to create negative outcomes. Adversarial attacks are distinct from traditional energy system cybersecurity risks – often exploiting the data-driven nature of AI methods – and vary by objective, access requirements, and knowledge requirements. Common types of attacks include the following:

- **Poisoning attacks** add, modify, or alter the data used to train an artificial intelligence model, in order to force the model to learn the wrong behavior. This can include modifying data on energy system operations, so that a model develops an incorrect conception of what “normal operations” look like. It can also include more sophisticated efforts to create a “backdoor,” which yields specific results when triggered – for example, poisoning the training data so that a model meant to detect physical wear in oil and gas equipment *never* declares an equipment to need maintenance, when presented with a very specific image.
- **Evasion attacks** use adversarial *input* data, which may look indistinguishable from regular data to a human, to produce a desired model output – typically counter to the wishes of the model creator. An evasion attack might slightly alert the data presented to a model trained to predict energy market prices, in a carefully engineered manner that causes the model to incorrectly overestimate or underestimate prices.
- **Data extraction attacks** seek to learn sensitive information about an artificial intelligence model, or the data it has been trained on. For AI tools that are customized for energy applications, or even tuned to specific energy infrastructure, a data extraction attack could allow an adversary to access the closely held information about an energy system of interest that is embedded in the AI tool.

Adversarial attacks against AI - and defense against such attacks - is an active area of research, distinct from traditional cybersecurity. While current defense techniques are not mature enough to guarantee security against sophisticated attacks, potential risks can be mitigated through a range of best practices including training data curation, access controls, and human supervision.

Risk Category 3: Hostile Applications of AI

AI can be created and used by adversaries to plan or execute cyber or physical attacks on energy infrastructure. In some cases, AI may lower the difficulty of an attack, enabling less-sophisticated adversaries to carry it out. In others, the use of AI may still require sophistication, but could enable more effective attacks than were previously possible. The report identifies a non-exhaustive list of ways in which adversaries might use AI for hostile means:

- **Automatic parsing of text for vulnerabilities** can reduce the amount of time and manpower an adversary needs, in order to conduct reconnaissance on energy



infrastructure being targeted for a potential attack. These techniques can allow adversaries to identify key pieces of information, vulnerabilities, or system architecture detail based on data and documents from a broad variety of sources.

- In the near future, **model inference or model completion based on available data** may allow adversaries to leverage AI to help fill in missing details or information needed to design an attack on energy infrastructure. Adversaries with little to no information about a targeted energy system might be able to rely on the inference capabilities of AI tools to provide insights (whether factual or synthetic) that can inform attack design.
- AI tools can also be used to facilitate **model-based design of attacks**, allowing adversaries to combine AI inference with information about specific energy system infrastructure (aided by traditional modeling), to design and simulate more complex attacks that are optimized to create the most significant or disproportionate impact.
- AI can also enable **autonomous control of devices for physical attacks**. Adversaries could combine AI-enhanced capabilities with other technologies, such as unmanned drone systems, in order to execute remote physical attacks on energy infrastructure.
- AI-driven **autonomous malware** may be able to flexibly adapt to the system it is attacking, to more effectively seek out and target high-value systems within a network, or autonomously make decisions to update its objectives over time.
- AI techniques can help **cyber attacks evade detection**, by allowing attack tools to learn how to bypass firewalls, mask themselves from detection, or even masquerade as beneficial software – all of which would increase the difficulty of detection by defenders.

An analysis of publicly available energy information should be performed to assess what sensitive information may be inferable with AI methods. If that analysis identifies sensitive information is potentially inferable, additional steps may be needed to mitigate risks.

Risk Category 4: Compromise of the AI Software Supply Chain

Unlike Risk Category 2, this category is not focused on the manipulation of AI models themselves, but rather analyzes the ways in which AI software supply chains might face traditional cybersecurity risks — such as those common to many digital systems currently used in energy system operations. As AI is software, it is subject to the same cybersecurity risks of other software – as the assessment explores through examination of recent cyber supply chain attacks. An adversary may exploit AI software not only to attack the AI system, but also as an intrusion vector to a victim’s broader energy infrastructure systems. This can occur through both proprietary and open-source software, which AI systems are often heavily reliant on – and is potentially a particular concern as AI tools, particularly those relying on generative AI techniques, shift from bespoke design to relying on common tools, libraries, and in some cases, foundation models. As such, cybersecurity and energy system supply chain security best practices are critical to securing the AI software supply chain.



Next Steps

The Department of Energy and CESER view the role of artificial intelligence in the U.S. energy system as a critical and dynamic topic – requiring an enduring program of effort and sustained engagement as this technology and its uses within the energy sector continue to evolve.

Over the course of calendar year 2024, CESER will be expanding its engagement with sector partners on artificial intelligence, from the security and resilience perspective, while also working to identify areas of existing and upcoming programmatic focus, where artificial intelligence is a key consideration.

Energy Sector Engagement

- CESER plans to brief the highlights of the assessment to energy sector leadership, including the Oil and Natural Gas Subsector Coordinating, the Electricity Subsector Coordinating Council, and entities identified under Section 9 of Executive Order 13636 to inform future engagement on policies, tools and technologies, and risk management efforts.
- CESER will host virtual listening sessions with energy sector partners and SMEs on artificial intelligence in Summer 2024. The goal of these sessions will be to engage with the sector on knowledge gaps, topic areas requiring further analysis or examination, and opportunities to use AI to help improve sector resilience and security.

Artificial Intelligence and the Energy Sector Risk Management Agency

- CESER will continue to serve as a trusted partner for the sector, and within the government, in offering candid, science- and engineering-driven insights on the benefits and risks that artificial intelligence presents for the sector.
- CESER will examine, in coordination with energy sector stakeholders, the public availability of energy sector data and its potential to impact the security posture of owners or operators of energy infrastructure. This will include evaluating the potential for such data to be leveraged to enable cyber or physical attacks against energy infrastructure, including using AI tools.
- CESER will develop, and update as needed, an internal plan to scope and guide its efforts to leverage the benefits and manage the risks of AI use in the energy sector.
- CESER will explore the ways in which AI can be leveraged by existing CESER programs, such as Energy Threat Analysis Center (ETAC), to ensure that they continue to operate at the speed and scale required by a changing energy system and dynamic risk landscape. Potential areas of application include utilizing AI tools to enable machine-scale analysis of operational and IT data from energy systems, for potential risks.
- CESER will continue to support the research and development of innovative tools and technologies that leverage artificial intelligence in ways that seek to strengthen the security and resilience of the U.S. energy system.
- CESER will engage in these activities as part of the Department-wide efforts on AI, being led by the newly established Office of Critical and Emerging Technologies.

As we navigate the continued deployment of AI across our energy system, a risk-informed approach will allow us to realize the security, resilience, and systemic benefits that AI promises. CESER appreciates the support of energy sector stakeholders and subject matter



experts in reviewing and providing feedback and recommendations for improvement on this assessment and is planning to issue an update to this assessment by the end of 2024.

Background: Executive Order 14110

Executive Order 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, issued on October 31, 2023, identified the Administration's eight core principles for advancing and governing the development and use of AI:

- Artificial Intelligence must be safe and secure.
- Promoting responsible innovation, competition, and collaboration will allow the United States to lead in AI and unlock the technology's potential to solve some of society's most difficult challenges.
- The responsible development and use of AI require a commitment to supporting American workers.
- Artificial Intelligence policies must be consistent with the Administration's dedication to advancing equity and civil rights.
- The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected.
- Americans' privacy and civil liberties must be protected as AI continues advancing.
- It is important to manage the risks from the Federal Government's own use of AI and increase its internal capacity to regulate, govern, and support responsible use of AI to deliver better results for Americans.
- The Federal Government should lead the way to global societal, economic, and technological progress, as the United States has in previous eras of disruptive innovation and change.

CESER, with support from LLNL, developed this assessment in response to Section 4.3 of the Executive Order (*Managing AI in Critical Infrastructure and in Cybersecurity*). To ensure the continued protection of U.S. critical infrastructure, the section requires that:

Within 90 days of the date of this order, and at least annually thereafter, the head of each agency with relevant regulatory authority over critical infrastructure and the heads of relevant SRMAs, in coordination with the Director of the Cybersecurity and Infrastructure Security Agency within the Department of Homeland Security for consideration of cross-sector risks, shall evaluate and provide to the Secretary of Homeland Security an assessment of potential risks related to the use of AI in critical infrastructure sectors involved, including ways in which deploying AI may make critical infrastructure systems more vulnerable to critical failures, physical attacks, and cyber attacks, and shall consider ways to mitigate these vulnerabilities. Independent regulatory agencies are encouraged, as they deem appropriate, to contribute to sector-specific risk assessments.