FECM Implementation Pathway for Responsible Artificial Intelligence

September 2023



Fossil Energy and Carbon Management Artificial intelligence (AI) holds the potential to accelerate the U.S. transition to a carbon-neutral economy and help achieve the technology research, development, demonstration, and deployment (RDD&D) goals of the U.S. Department of Energy's Office of Fossil Energy and Carbon Management (FECM) set out in its <u>Strategic Vision</u>. FECM and the National Energy Technology Laboratory (NETL) are building on the existing EDX and Science-based AI/ML Institute (SAMI) resources at NETL to create a robust AI development environment known as EDX++ (EDX + multi-cloud + compute resources). Work on this game-changing resource is now in the pilot phase, with deployment of the full EDX++ application expected in fiscal year (FY) 2026.

FECM programs currently use a range of AI tools to improve productivity and results. Once fully deployed, EDX++ stands to dramatically increase and expand these benefits. In the interim, FECM programs can actively prepare to take advantage of this transformative resource. This document describes the goals and key components of the FECM/NETL AI infrastructure and the resources and opportunities for FECM programs and staff.

Disclaimer

This material was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the United States Department of Energy, nor the Contractor, nor any of their employees, nor any jurisdiction or organization that has cooperated in the development of these materials, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness or any information, apparatus, product, software, or process disclosed, or represents that its use would not infringe privately owned rights.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, its contractors, or subcontractors.

Cover image: AdobeStock #168274790

Contents

| Dverview | 1 |
|---|------|
| Follow Official AI Guidance | 3 |
| Jnderstand the AI Infrastructure | 4 |
| EDX++ | 4 |
| NLP Sandbox at Headquarters | 5 |
| Prepare To Efficiently Utilize AI | 6 |
| Plan | 6 |
| Resources for FECM Applied RDD&D Programs | 6 |
| Business Applications at HQ | . 11 |
| Train and Educate the Workforce | . 11 |
| Leverage and Collaborate on AI | . 13 |
| Sources | . 15 |

Figures

| Figure 1. The EDX ++ framework | 4 |
|---|----|
| Figure 2. Potential AI roles identified | 7 |
| Figure 3. High-impact applications of AI in FECM | 8 |
| Figure 4. Relative value of applied AI categories by program | 9 |
| Figure 5. Grand Challenges of greatest value to FECM applied programs | 10 |
| Figure 6. Dashboard for the Natural Language Processing (NLP) Sandbox | 12 |

Overview

In concert with national and DOE guidance, the Office of Fossil Energy and Carbon Management (FECM) is developing a comprehensive artificial intelligence (AI) framework to drive technology innovation that will enable an environmentally sustainable and prosperous energy future. The incredible power of AI, while in its infancy, holds the potential to solve fundamental challenges facing our nation and planet. FECM is preparing for the coming paradigm shift as AI transforms the research, development, demonstration, and deployment (RDD&D) of energy technology and carbon management. The intent is to enable FECM programs to more effectively harness the full power of AI to address the critical global challenges of clean energy and climate change.

"The advent of artificial intelligence (AI) is rapidly revolutionizing society and the way in which researchers tackle some of science's most complex problems. Exponential increases in data volume and computing capabilities have altered the R&D landscape for both science and industry and created a unique opportunity to expedite understanding and vastly improve our world."

Artificial Intelligence Initiative, ORNL

FECM programs currently use AI to turn raw data into scientific insights or models that improve research outcomes. FECM managers and staff need to utilize the rapidly evolving power of AI to achieve critical program goals and objectives and avoid unintended impacts. Achieving the efficient, secure, equitable, and cost-effective use of AI envisioned in national AI guidance will require a full understanding of the goals and security issues, current and emerging AI capabilities, priority applications in program RDD&D (and support functions), dynamic planning horizons, and the necessity of continuous learning.

FECM conducts innovative technology RDD&D to significantly reduce, capture, convert, or transport and store the carbon dioxide (CO₂) and other greenhouse gases associated with the use of fossil energy. AI has the power to facilitate the needed research, lower costs, improve outcomes, and deliver sustainable energy benefits to the nation. As summarized in this document, FECM management is providing FECM applied technology research programs with the AI tools, training, and other capabilities required to facilitate high-impact RDD&D and operational excellence while mitigating the social, security, and ethical risks of AI.

FECM is building the acclaimed Energy Data Exchange (EDX) at the National Energy Technology Laboratory (NETL) into a robust AI framework. The pilot framework, known as EDX++, is scheduled to launch in early 2024, and the full framework is due for completion in 2026. This document addresses key steps, resources, and opportunities now available to help FECM programs and their RDD&D partners prepare to begin taking advantage of this AI framework early in 2024.

- Assure compliance with all official Al guidance: The cybersecurity of the EDX++ framework and related resources at FECM/DOE Headquarters follow the standards and protocols issued by the DOE Office of the Chief Information Officer (OCIO). NETL also aligns EDX++, related resources, and the rules governing their access and utilization with all pertinent National and Agency guidance, emphasizing equity, fairness, explainability, sharing, human-Al collaboration, and reuse. See page 3 and Appendix A for details.
- Understand the AI support infrastructure: FECM and NETL will continuously support, expand, and curate extensive scientific datasets, AI tools and models, secure workspaces, and access to computing services to enhance priority RDD&D. The hybrid, multi-cloud AI framework known as EDX++ pulls together diverse expertise, resources, and capabilities to enable the DOE/FECM research community to securely leverage a broad collection of tailored resources to expedite progress toward equitable and sustainable clean energy

solutions. FECM programs should become familiar with existing and planned resources and opportunities. Non-technical Headquarters (HQ) staff can also access secure AI workspaces (sandboxes) to try out natural language processing (NLP) models for potential use in completing mundane or repetitious tasks. See page 4 and Appendices B and C for details.

- **Prepare To Efficiently Utilize AI:** FECM applied technology research programs and business support offices can now prepare to take full advantage of EDX++ and/or the NLP Sandbox at HQ by incorporating these resources into their planning, workforce development, training, and collaborative activities.
 - Plan: FECM programs now have the following resources to guide RDD&D planning efforts and support functions (e.g., multi-year plans, research portfolios, and funding opportunity announcements).
 - Roles for AI. Representatives of each applied FECM program participated in exercises to identify the
 potential roles for AI to help address critical program-specific challenges. This effort produced a series
 of documents that identify promising roles and provide context for the challenges as well as current
 gaps in data or scientific understanding.
 - Key AI approaches. Matrices and a summary chart link the program-specific roles for AI to the mostneeded applied AI approaches in each FECM program to help inform the preparation of program solicitations.
 - Priority Grand Challenges in AI. The high-impact, program-specific roles for AI are linked to the relevant AI Grand Challenges identified by the DOE Office of Science, producing a matrix and summary chart highlighting the AI challenges/solutions that are expected to bring the greatest value to FECM programs, individually and overall.

Collectively, these resources (further described on pages 6–11 and in Appendices D, E, and F) should contribute to the successful deployment of AI across FECM programs by identifying effective use cases.

- Educate and train: FECM has set up opportunities to assist current program staff in improving their AI awareness and upgrading their knowledge of AI developments. Staff can take advantage of a range of ongoing education and staff development offerings to build awareness of AI capabilities, benefits, human-AI collaboration, and best practices (see page 11). Plans are underway to share and leverage insights on best practices from other offices, agencies, and the private sector. FECM business offices, human resource specialists, and managers can take advantage of training opportunities for non-programmers to make the most of their unique human skillsets. AI education and awareness will help FECM build a strong workforce.
- Leverage and Collaborate: FECM programs have many opportunities to leverage resources (datasets, specialized expertise, models) by coordinating and collaborating with other DOE offices, federal agencies, academia, industry, non-profits, and national and international organizations (see page 13).

With proper planning and preparation, FECM will be well-positioned to take full advantage of EDX++ and thereby improve AI application outcomes, reduce AI development time, improve confidence in novel technologies, accelerate technology deployment, and achieve transformative solutions. In making these preparations, FECM staff, management, and diverse support teams will need to continuously assess their plans and activities and ensure awareness of evolving resources, guidance, security measures, partnerships, and other opportunities for transformative innovation.

Follow Official AI Guidance

FECM works closely with the National Energy Technology Laboratory (NETL), other national labs, partners, and other federal agencies to develop data repositories and AI resources in alignment with all federal and DOE guidance. Guidance to date (see inset as a subset) emphasizes the development and application of AI to ensure the following:

- Technology leadership
- Trustworthiness and explainability
- Equity and fairness in access to data and AI
- Cybersecurity and privacy
- Education and workforce training
- Human-AI collaboration
- Sharing and partnership to expedite progress.

Adherence to robust national guidance on cybersecurity and privacy protections is entrusted to specialists in the DOE Office of the Chief Information Officer (OCIO), which (a) provides IT governance, policy, and oversight processes to ensure secure, efficient, and cost-effective use of IT resources and (b) ensures acceptable risk-based cybersecurity by enhancing enterprise situational awareness, developing near real-time risk management, and combating advanced persistent threats (DOE <u>OCIO</u> 2023). Within FECM, AI is applied to facilitate solutions for the priority RDD&D areas identified in the 2022 *FECM <u>Strategic Vision</u>* and alleviate management and business workloads. For more information, please see Appendix A.



Selected National AI Guidance & Initiatives

www.ai.gov/documents/

Executive Order <u>13859</u> launches sustained effort to maintain American leadership in AI.

Executive Order <u>13960</u> promotes the use of trustworthy AI in the federal government.

Executive Order <u>13985</u> promotes equity in science and the elimination of bias in the design and use of new technologies, such as AI.

Executive Order <u>14028</u> promotes improved national cybersecurity, including security principles governing Cloud Service Providers for incorporation into agency modernization efforts.

Guidance on Desirable Characteristics of Data Repositories for Federally Funded Research to make data FAIR and promote equitable access with privacy and security (White House, May 2022).

<u>Strengthening and Democratizing the U.S.</u> <u>Artificial Intelligence Innovation Ecosystem:</u> An Implementation Plan for a National AI Research Resource (NAIRR, January 2023)

National Strategy To Advance Privacy-Preserving Data Sharing and Analytics presents methods to unlock the benefits of data analysis while protecting privacy (NITRD, March 2023).

The National AI R&D Strategic Plan: 2023 Update (May 2023) defines major research challenges in AI to coordinate and focus federal R&D investments and highlights nine strategies:

- Make long-term investments in AI research
- Develop effective human-AI collaboration
- Address ethical, legal, and societal impacts
- Ensure the safety and security of AI systems
- Enable access to high-quality public datasets and environments for AI training and testing
- Develop standards and benchmarks for AI
- Foster an AI-ready American workforce
- Expand partnerships to accelerate AI advances
- Prioritize international collaborations in AI R&D to address global challenges.

<u>Al Guide for Government</u> GSA guide for agency senior leaders and decision makers (June 2023).

Understand the AI Infrastructure

NETL scientists are already using some of the advanced AI capabilities being developed as part of FECM's comprehensive AI framework known as EDX++, and FECM research teams will be able to access the EDX++ pilot in FY 2024. FECM HQ staff (technical and non-technical) currently have access to a natural language processing (NLP) testbed, which enables experimentation and hands-on learning to try out NLP models in a secure environment (sandbox).

EDX++

FECM is working with NETL's Science-based AI/ML Institute (SAMI) to significantly expand the Laboratory's acclaimed Energy Data Exchange (EDX)¹ into a robust AI support framework for launch in early FY 2024. This framework (EDX + model curation + multi-cloud resources) is designed to support the FECM Mission, align with the *AITO <u>FY22 Program Plan</u>*, and improve AI access and efficiency across program offices and business functions. The focus is to simultaneously lower the threshold for using AI to achieve ambitious research goals and continuously expand resources to drive innovation across the applied FECM research portfolio.

EDX++ is designed to provide FECM a virtual platform for AI tools, the curation of data, and computing power to facilitate mission-critical RDD&D (Figure 1). EDX was developed and is maintained by NETL's Research and Innovation Center (RIC) researchers and technical computing teams to support private collaboration for ongoing research efforts and technology transfer of the resulting knowledge and products. NETL will effectively manage the expanded configuration, upkeep, and operation of the integrated EDX++ framework. Now in pilot testing, EDX++ will support FECM and NETL-affiliated research by delivering state-of-the-art data curation, cloud access (aligned with cybersecurity requirements and protocols), secure workspaces, and expert technical support. As summarized in Appendix B, EDX++ also offers the following capabilities and resources:

- Data curation
 - Discoverability and access
 - Legacy data rescue
 - Anonymization and federated learning
 - Quality control and integration
 - Compatibility with future media
- Technical Support
 - RDD&D
 - Business applications
- Cloud Access
 - Virtual gateway
 - High-speed data transfer
 - Git accounts



Figure 1. The EDX ++ framework. EDX++ connects massive scientific datasets with scientific computing resources and cloud access to drive AI-assisted FECM RDD&D. Source: <u>SAMI</u>

¹ See <u>video</u> on EDX

- Specialized computing resources
- Secure workspaces

NLP Sandbox at Headquarters

Al tools that do not require high-performance computing capabilities can also improve the efficiency of management and business functions. FECM has set up a testbed that is now available for HQ personnel to try out various Natural Language Processing (NLP) models in a safe environment (security protocols meet OCIO standards). A range of training on the use of this NLP Sandbox is available to interested parties, depending upon their needs and programming experience (from no experience to extensive use or familiarity). See page 12 and Appendices B and C for more information.

Prepare To Efficiently Utilize AI

Core AI capabilities will start to become available through EDX++ in early 2024, and new tools, data, and resources will be added as the framework is built out for completion in 2026. To maximize benefits, FECM programs can take advantage of the resources available today to begin proactively developing effective RDD&D plans, getting staff trained to leverage AI effectively to meet their responsibilities, and forming strategic collaborations across DOE and with other local, tribal, state, national, and international organizations.

Plan

Many FECM applied program staff and managers may be on the learning curve when it comes to determining when or if AI is the right tool to address a particular challenge. This curve may stretch over time as more AI and machine learning (ML) tools become available, which is why FECM has conducted initial assessments in concert with each program and is pleased to provide the <u>resources</u> described below. The more that program staff understand current and planned AI tools (e.g., generative discovery, transfer learning), capabilities (e.g., prediction, performance optimization, systems integration), trustworthiness (explainability, validation), data requirements (amount, quality, updates), and limitations (applicability to distinct conditions or environments), the better they can plan for their effective use in meeting program goals.

Using AI judiciously

Al can effectively assess massive amounts of data to detect patterns, provide useful insights, or extract and format meaningful components, but it is not a tool that can be applied indiscriminately—at least not yet. Al typically requires large quantities of reliable data to make predictions with an acceptable level of certainty. While AI has accelerated major scientific breakthroughs and holds great promise to improve understanding of complex systems (or interactions among complex systems), AI remains at a relatively early stage of development—yet it is advancing rapidly. Prospective users need to evaluate the status of the relevant AI, the required accuracy for an intended application, the quality and availability of data for AI training and validation, and the adequacy or strength of more traditional tools. AI tools are evolving, and FECM programs now have access to the following resources to better incorporate AI in their multi-year plans, RDD&D portfolios, and solicitations. The resources described below should help FECM programs answer or begin to address questions such as the following:

- How can AI realistically help attain our research objectives in the near- to mid-term?
- Are the needed AI tools currently available?
- Are our most-needed AI tools a priority for development by the Office of Science?
- Which other FECM programs are also eager to have these AI tools?
- What data are required to validate our applications of available AI tools?
- Do we have the data to enable/validate the needed model development?
- What steps can we take now to improve our data and AI success?

Resources for FECM Applied RDD&D Programs

Roles for AI in Support of FECM RDD&D Priorities series (2023)

Representatives from each FECM applied RDD&D program participated in joint, program-specific exercises with specialists from FECM HQ, NETL, and the Office of Science to explore ways in which AI could potentially expedite progress toward defined program objectives (as identified in the FECM



Strategic Vision). These exercises produced seven documents, one for each program, (see sample cover above) summarizing the promising potential roles for AI and some relevant context, including the complexity of the problems to be solved and the associated challenges for AI developers. Each document provides a summary chart of the key areas in which AI could significantly impact the program (see Figure 2 for one example and Appendix D for the full set). Each FECM applied program can use the document that they helped develop as a starting point to envision AI-powered progress and plan accordingly. The following examples illustrate potential AI applications in specific programs.

- Methane Mitigation: Predict and characterize leaks to prioritize action. Al could integrate information from multiple sources (e.g., spectroscopy, satellite imagery, vegetation succession, terrain, etc.) to rapidly assess the condition or integrity of abandoned or orphaned wells.
- Critical Minerals: Locate and characterize resources. Using geological, geophysical, geochemical, and other data, AI can indicate the likelihood that critical minerals are present in diverse resources, such as shale or mine tailings. The most



Figure 2. Potential AI roles identified. Each FECM applied RDD&D program identified potential high-impact roles for AI.

sophisticated current tool, portable x-ray fluorescence, has limited detection capabilities.

• **Carbon Dioxide Reduction:** Predict optimal materials for commercial-scale operations. Using detailed, labscale materials data (e.g., solubility, permeability, surface area, isotherms, thermogravimetric analysis), AI could identify the key factors for optimizing critical functional properties at scale.

While many of the needed AI tools are still in development, these summaries of potential roles for AI (available at <u>Roles for Artificial Intelligence in Support of FECM Priorities | Department of Energy (doe.gov)</u>) can help FECM programs develop multi-year program plans, build robust RDD&D portfolios, gather the necessary high-quality data, develop FOA topics, and lay the foundation for future projects and investments.

Crosswalk of AI Roles for FECM to AI Approaches

Based on the high-impact roles for AI identified by the FECM programs (described above), subject matter experts at NETL pulled out the key priority application areas specific to each program and sorted them into eight main categories of applied AI/ML, which are listed and briefly explained in the inset on the following page. This exercise produced a series of matrices that show the categories of applied AI most promising for RDD&D in each of the priority applications (ranging from four to nine) in each program. These matrices are included in Appendix

E to provide program managers and planners a perspective on the types of AI/ML that hold the greatest potential to support program objectives and the specific applications they benefit.

These matrices were subsequently aggregated (as shown in Figure 3) to identify the most needed AI categories across FECM programs. Although the bar lengths are skewed by the varying number of applications in each program, the graph suggests which AI tools broadly support priority FECM RDD&D program needs. These rankings also show where AI needs coincide, suggesting areas for potential collaboration among programs.

The matrices were also used to generate the pie charts in Figure 4 to illustrate the relative extent to which the selected applied AI categories apply to programspecific priorities in general. This information and the detailed charts in Appendix E may inform planning and FOA development.

AI Approaches with High-Impact Potential in FECM

- **Operator Learning**: This relatively new area of research in ML focuses on learning operators or functions that map inputs to outputs, rather than learning the mappings from inputs to outputs directly.
- **Reinforcement Learning**: This approach involves an agent learning how to make decisions in a dynamic environment through continuous interaction and feedback in the form of rewards or punishments.
- Advanced Sensors: Sensing devices to operate in harsh environments, sense minute fluctuations in critical environments, or process readings in real time to flag anomalies or trigger safety measures.
- Edge Computing: This distributed computing approach brings computation capabilities closer to the sources of data, often using simplified versions of complex models to reduce computing requirements.
- Science-based Learning: This includes physics-based hybrid models (i.e., incorporating physics equations into the loss function), chemistry-informed ML (i.e., incorporating chemistry equations in the loss function), or scientific simulations for data generation/validation.
- **Transfer Learning**: ML technique for re-purposing a model built for one task for a different yet similar task.
- Generative Discovery: Using large language models to learn the language of material structure and chemistry to predict new materials, behaviors, or assess properties.
- Unstructured Data Mining via Large Language Models: Data mining and extraction (from research papers, articles, sensor data, diagrams, etc.) can generate meaningful data for human or ML training.



Figure 3. High-impact applications of AI in FECM applied programs cluster into eight main categories. [Note: Bar lengths reflect the varying number of application areas (4-10) identified by each program.]



Figure 4. Relative value of applied AI categories by program. The dominant categories of AI needs vary somewhat by FECM applied RDD&D program. This mapping of anticipated needs may indicate program partnership opportunities. Potential application areas within each program (see Appendix D) may help focus RDD&D opportunities.

AI Grand Challenges of Greatest Value to FECM

Under the guidance of the Office of Science (SC) and the National Nuclear Security Administration (NNSA), NETL and the other DOE national laboratories organized a series of workshops in 2022 to gather input on new and rapidly emerging opportunities and challenges in scientific AI. The workshop discussions are synthesized in the

<u>2023 report</u> *AI for Science, Energy, and Security,* which describes six main foundational categories, each of which contains multiple (two to six) subcategories of AI Grand Challenges. Each challenge could enhance fundamental understanding of physics (or other natural laws) or AI approaches (e.g., better understand the convergence behavior of physics-informed neural networks). Progress in any of the resulting 23 challenges has the potential to transform RDD&D and have broad impacts across DOE missions.

Based on a review of these Grand Challenges, NETL specialists ranked each as offering either High, Medium, or Low value (in terms of positive impact on the quality and/or cost of AI outputs) for the priority AI applications identified by the FECM programs. Nine of the Grand Challenges received at least two marks for high relevance or at least five marks for medium relevance to program priorities. These nine challenges, shown in Figure 5 (x-axis indicates assigned



value for potential impact), would extend significant value to FECM in achieving its goals and objectives. This information could also be useful to the Office of Science when prioritizing its work. The full list of Grand Challenges and their descriptions are provided in Appendix F.

Note that failure to meet these nine Grand Challenge will not hinder progress in applying AI to FECM's key programs. Researchers can apply state-of-the-art tools to achieve progress today; however, advancements in these AI Grand Challenges can significantly expedite discovery, enhance understanding, and improve program outcomes across FECM and DOE.



Figure 5. Grand Challenges of greatest value to FECM applied programs. Advancements in these AI GCs can significantly expedite discovery, enhance understanding, and improve program outcomes.

Business Applications at HQ

Headquarters staff now have the opportunity to try out AI tools that could assist with a range of non-scientific tasks in a secure testbed at DOE HQ. FECM developed this Natural Language Processing (NLP) testbed to give HQ staff a secure area in which to explore a range of natural language technologies that could help generate or summarize reports or resumes, locate and extract needed data, complete forms, translate images to text, answer questions, and perform similar tasks. In addition, the provided AI can help to identify trends and derive insights from the source documents to support better tactical and strategic business decision making.

As described in the following section and Appendix C, staff are encouraged to become familiar with AI tools in this secure environment, known as the HQ NLP Sandbox. Only certain types of public data are allowed in the testbed, alleviating potential security concerns of staff who want to dip their toes in AI by "playing" with these powerful new tools—giving rise to the term "sandbox." Personnel at any level of technical knowledge can access the sandbox's no-code dashboard, attend forums and classes, or receive training appropriate to their skill level.

Train and Educate the Workforce

DOE Headquarters (HQ) staff have diverse options to learn about the types of tasks and problems best suited for the use of AI and to gain direct experience as end users. A range of training formats for different skill levels are available to suit the wide range of needs, knowledge, and experience. For example, the U.S. General Services Administration (GSA) offers the *AI Guide for Government* (see inset) and an online <u>AI training series for federal employees</u>.

FECM anticipates a far smaller requirement for AI training among those who will access the scientifically sophisticated EDX++ framework maintained by SAMI/NETL. The EDX++ is designed to assist RDD&D by the principal Investigators (PIs) selected to work on FECM research projects and other highly trained members of their teams. These experts are expected to be knowledgeable and experienced in the use of AI. As EDX++ nears the launch of its pilot in FY 2024, SAMI/ NETL will develop the necessary rules and protocols to ensure user compliance with established standards for security, equity, and reliable curation of data and models to promote sharing, efficiency, and progress. These protocols will align with safety and security requirements now under development by the OCIO and U.S. Cybersecurity and Infrastructure Security Agency (CISA).

To address the needs of HQ staff, FECM established a DOE AI Community of Interest (COI) in August 2022 to encourage the exchange of ideas, insights, knowledge, and lessons learned as well as data, tools, and other resources. Sponsored by FECM, NNSA, and Science, the community (currently about 70 members based at DOE HQ) has participated in setting up the following opportunities to learn about AI and natural language processing (NLP) models.

Training & Education: HQ Options

Quarterly Forums

- Al-related topics
- Of broad interest
- Distinguished speakers



NLP Sandbox

Self-guidedTailored training

Workforce Development

- Now planning
- 289 289
- EHSS is participating



AI COI Monthly Meetings

- Share news and feedback
- Develop plans and activities

COI Teams Site

- Capture and share ideas
- Post codes and briefings

Images: Adobe Stock 605032130, 227210896 Images: Adobe Stock 605032130, 227210896



Quarterly Forums on AI-Related Topics

Speakers from the government, national labs, industry, and academia are invited to speak at forums that are held quarterly to address AI-related topics of interest to a broad DOE audience with the objective of making AI more intelligible and mission relevant. Recent and upcoming topics include: The Unique Role of Government in Advancing AI, Introduction to ChatGPT, Spotting Deep Fake Images, and Quantum for AI.

Self-Guided Learning in the NLP Sandbox

The NLP Sandbox contains a user-friendly dashboard (see Figure 6) that allows staff members at all levels to try out the selected natural language processing models. The dashboard allows users to choose the types of models they wish to explore and gives them options to either try an existing model "as is," adjust a model for an

intended application, or make programming changes (in Python) to truly customize a model. Through staff use of this novel resource, the NLP Sandbox will enable the following:

- Enhanced staff awareness of AI potential
- · Testing of new models and variations for potential use
- Practical understanding of the capabilities and shortcomings of LLMs
- Results-based comparisons of commercial and open-source models
- Validation of pathways to support responsible, trustworthy AI (see AI Risk Management Playbook (<u>AI RMP</u>).

"Avoid centralizing AI practitioners and leaders in one unit. AI talent must be accountable to the business needs and therefore should exist across the organization." *AI Guide for Government* (GSA 2023)

| home | NLP Sandbox No classified data may be shared with the sandbox. Sensitive data (PII, procurement, pre-decisional, etc.) may NOT be used with PaLM, Open Al, or other Azure Cognitive services Open source or publicly available data may be used with any model | | |
|---|--|--|--|
| closed domain QA frequently asked questions | | | |
| question answering summarization | | | |
| Services and Links Introduction Generating hours: 0700-2000 EDT GPU access: Testing phase using endpoint instances defined in personal Jupyter notebooks. GCP potential offerings: interactive sheet Alsorithmic Impact Assessment Tool Data Science Webinars: Contact felix.sonzalez@ho.doe.sov Jupyter notebook accounts: Contact | Natural language processing (I respond to human language. T personalize the user experience Improving search results: website pages and identif user's search query. This c information they are looki Personalizing the user exp browsing history and pref to personalize the content recommending products of content of the series of the series of interact with users in a nat customer questions, provi NLP is a powerful technology t user experience. If you are look is a technology that you should Current Module | NLP) can help websites understand and this can be used to improve search results, and create chatbots. NLP can be used to analyze the content of the keywords that are most relevant to can help to ensure that users find the information can then be used to that is displayed to the user, such as information can then be used to that is displayed to the user, such as information can then be used to track a user's freences. This information can then be used to track a user's freences. This information can then be used to track a user's freences. This information can then be used to that is displayed to the user, such as information can then be used to track a user's freences. This information can then be used to answer id support, or even sell products. Interstep of the websites to provide a better king for ways to improve your website, NLP do consider. Ulles on the Website | |
| steven.wong@nq.doe.gov or sailendra.mahapatra@hq.doe.gov | Module | Purpose | |
| | caption generation | Image analysis models can be trained and provide a first pass at analyzing and documenting images. | |
| | classification | NLP models can be used to provide a first pass at keywords or classification of documents. | |
| | FAQ | Answer questions about this website or natural language processing (NLP). | |
| | j <u>upyter notebook</u> | A chance to test python codes. | |
| | closed domain question answering | Answers questions based on provided text. The purpose is to minimize the generation of false answers. | |
| | open domain question | Similar to closed domain question answering except uses the Internet to get an answer. You can also guide the | |

Figure 6. Dashboard for the Natural Language Processing (NLP) Sandbox at DOE HQ is user-friendly.

Tailored Training in Use of NLP Sandbox

As needed, knowledgeable experts from FECM or NETL will offer hands-on training to novices and beginning programmers at DOE HQ who want to gain more confidence in using the NLP Sandbox. This type of training may be appropriate for those with a clear application for NLP to significantly expedite repetitive, timeconsuming tasks. Classes will be tailored and scheduled according to demand.

Workforce Development Seminar (Planned)

Current Capabilities in NLP Sandbox

- Exploratory area with preconfigured test cases for non-coders
- Testing and workforce development areas for mid-level coders
- Personal websites for interactive demonstrations of concepts and models for experienced coders

The COI is working with the Department's Environment, Health, Safety and Security (EHSS) organization to put on a series of seminars to guide FECM and other interested offices in developing and integrating AI talent across functions. Elevating AI awareness and expertise within FECM and DOE can be achieved by training and educating existing staff and through the hiring process. The seminars will benefit from EHSS expertise in policy development and technical assistance, safety analysis, and corporate safety and security programs. The seminars are scheduled to begin in the first quarter of 2024.

COI Monthly Meetings and Teams Site

Members of the AI COI meet monthly to share input on AI needs, successes, failures, and lessons learned. The group investigates AI impacts and risks, generates potential solutions or recommendations, and encourages productive dialogs with the national laboratories and external groups on topics of concern. The COI set up and regularly updates a Teams site for sharing AI project information, technology updates, impacts, and news. The site facilitates the dissemination of ideas, presentations, and code.

Business Applications

Managers, human resource recruiters, and other DOE HQ staff responsible for digesting, evaluating, and reporting on large volumes of information can safely explore the NLP models available in the Sandbox. Introductory training on the use of specific models is available to all staff.

Leverage and Collaborate on AI

"The purpose of AI and automation is to make low-value tasks faster and easier, thus freeing up people to focus on high-value work that requires human creativity and critical thinking."

AI Guide for Government (GSA 2023)

In accordance with national guidance on U.S. AI development

and leadership (see Appendix A), FECM programs are cultivating relationships with a range of other organizations to share data, findings, models, and best practices. Many FECM programs and their RDD&D partners currently leverage or have identified opportunities to leverage resources by collaborating with other DOE programs and offices, national laboratories, federal and state agencies, academic institutions, companies, and or international entities. The benefits of collaboration include the following:

- Rapid identification of and access to the data required to drive and validate the AI
- Deeper understanding of relevant foundational science
- Leveraged investments in RDD&D to achieve shared objectives
- Access to AI tools for materials discovery, system prediction and optimization, and systems integration.

Programs that require deep knowledge of the subsurface (e.g., critical minerals, carbon storage, enhanced oil recovery) have identified shared data and research needs across FECM and DOE for years through the Subsurface Intra-Agency Coordination Team (IACT). In addition to FECM, members include the DOE Offices of

Science, Geothermal Technologies, Nuclear Energy, NNSA, Environmental Management, and others. By upgrading data curation practices and sharing niche data and best practices, these offices are laying the groundwork for future opportunities to share new Al-driven insights and leverage neural networks and transfer learning.

Data is of pivotal importance to AI and, as documented in the program-specific Roles for AI in Support of FECM RDD&D Priorities series, FECM programs have identified the need for more sensitive or smart sensors to collect more and better data to support planned AI roles. Some FECM programs are already envisioning broad application of the AI models to be developed and validated, prompting closer scrutiny of the conditions under which data points are collected. Establishing consistent conditions (e.g., isotherms at defined temperatures) for the collection of certain datapoints could improve comparability across datasets, potentially facilitating model assessment for wider use across geographies or environmental conditions. At a minimum, such data collection issues should be considered in discussions with PIs, AI experts, and outside organizations with an interest in collaborating and sharing data. Current alternatives to obtaining massive datasets include synthetic data or sparse data models, but even these options require accurate training data. Identifying and leveraging collaborative opportunities at all levels (beginning with data) can significantly improve AI-driven RDD&D outcomes.

Example Opportunities for FECM to Share Data or Collaborate on AI

Within FECM



- Sorbents: Point Source
 Direct Air Carbon Capture

Other DOE Offices

Subsurface IACT



- Materials for sensors, equipment, and sorbents: Advanced Materials & Manufacturing Technologies (AMMT)
- Hydrogen Management: Hydrogen & Fuel Cell Technologies Office
- Negative Carbon Earthshot (planned Roadmap)

Federal Agencies

 Ocean-based Carbon Capture: NOAA, NASA



- Interagency Adv. Power Group (IAPG): NASA, Army, Navy, Air Force
- Geospatial data: USGS, EPA, NASEM, NIST,



State & Local Agencies

- Well permits (methane leaks)
- Environ. & econ. dev. agencies
- Community leaders (impacts)

Private Companies

- Pipeline operators Sensor developers
- Mining companies



International

- GEOMAR Helmholtz Centre
- World Bank

The EDX++ framework itself will enhance collaboration among FECM programs and across DOE offices as it develops and curates datasets, algorithms, and models for multiple RDD&D efforts. All FECM programs are encouraged to plan ahead to identify critical data gaps and take steps to secure the data they will need to make the most of the uniquely powerful AI tools soon to be available. Small changes to the language governing data resources in FOAs or MOAs today could have a major impact on how well a program can harness AI tomorrow.

Sources

- CISA 2023. U.S. Cybersecurity and Infrastructure Security Agency, *Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Security-By-Design and -Default*, April 13, 2023. <u>www.cisa.gov/sites/default/files/2023-06/principles_approaches_for_security-by-design-default_508c.pdf</u>
- CISA 2023a. U.S. Cybersecurity and Infrastructure Security Agency, "Software Must Be Secure by Design, and Artificial Intelligence Is No Exception," blog by Christine Lai and Dr. Jonathan Spring, August 18, 2023. <u>https://coe.gsa.gov/coe/ai-guide-for-government/identifying-ai-use-cases-in-your-organization/index.html</u>
- NETL 2023. National Energy Technology Laboratory, NETL and AI Partner Enhance Energy Modeling Techniques, web page, June 20, 2023. <u>https://netl.doe.gov/node/12621</u>
- NIST 2023. National Institute of Science and Technology, U.S. Department of Commerce, Artificial Intelligence Risk Management Framework (AI RMF 1.0), January 2023. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf