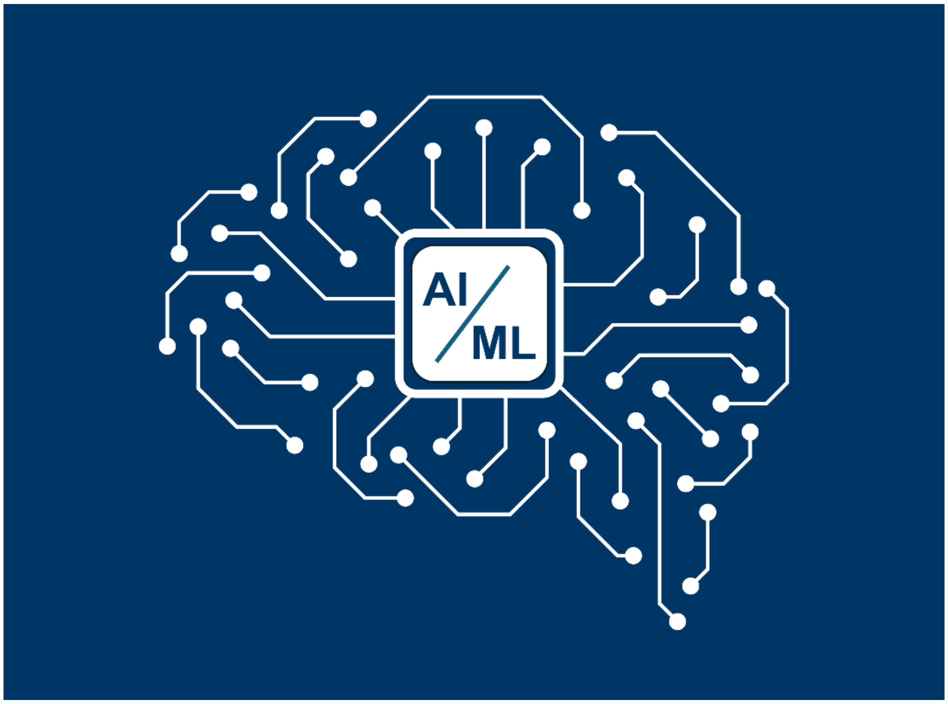
Whitepaper: Utilizing AI/ML in FICAM Environments

Dr. Nnamdi Osia



Adapted from https://intellectualpoint.com.

What Is AI/ML?

Artificial intelligence (AI) and machine learning (ML) are closely related concepts, with ML being a subset under the larger AI umbrella.

Artificial Intelligence

AI refers to the broader concept of machines, computer systems, or devices that can perform tasks or functions that are typically associated with human intelligence. Such tasks include reasoning, learning, problem-solving, self-improvement, perception, natural language understanding, and decision-making.

Machine Learning

ML is a field within AI that focuses on the capability of computers and machines to learn from provided data without being explicitly programmed for a particular task. ML algorithms can learn patterns and relationships from datasets to make predictions or decisions. The decisions are said to reflect how humans think, which over time improves the models and decision-making process.

Figure 1 illustrates the AI-ML relationship.

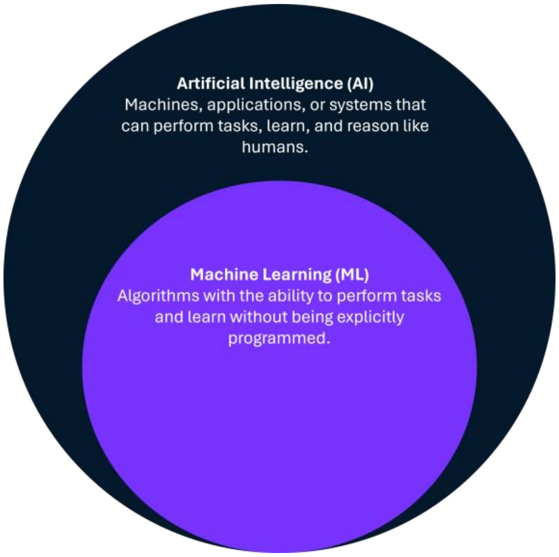

Artificial Intelligence (AI)
Machines, applications, or systems that can perform tasks, learn, and reason like humans.
Machine Learning (ML)
Algorithms with the ability to perform tasks and learn without being explicitly programmed.

Figure 1 - Artificial Intelligence vs Machine Learning.
Adapted from https://images.datacamp.com.

Utilizing AI/ML in FICAM Environments

Federal Identity, Credential, and Access Management (FICAM) focuses on improving security and interoperability of authorization and authentication systems across federal agencies based on approved policy, standards, and regulations. In FICAM environments, federal agencies seek to ensure that entities accessing resources are properly authenticated and authorized with a need to know. FICAM is crucial in enhancing the security posture of federal government systems and infrastructure, especially through the management of the identity lifecycle (i.e., from the provisioning of digital identities to their decommissioning). A successful FICAM implementation enables consistent and effective management of identities, credentials, and access controls across agencies.

AI/ML has several use cases within the FICAM environment.

Use Cases

- **Aiding processes related to identity lifecycle management**: Identity lifecycle management involves governance of identities at scale. Typically, it requires some kind of infrastructure or application modernization for legacy systems. The end goal of identity lifecycle management is automation of the entire process. AI can be used for efforts related to identity lifecycle management, such as ensuring digital identities are distinguishable, auditable, and consistently managed. It also may address operations to provision, merge, update, lock, revoke, or decommission credentials for a particular device or application.

- **Analyzing cyber threats and behavioral patterns associated with access**: AI/ML can be used to detect cyber threats and analyze behavioral anomalies associated with device-based or risk-based authentication. Anomaly detection requires data from the device as well as the transaction request. AI/ML can help detect irregularities in actions such as login, authentication, and authorization attempts. AI/ML also can automate communication of incidents or findings of cyber threats and anomalous login attempts.

- **Monitoring risk and internal controls**: AI/ML can aid in monitoring risks and internal controls. Some examples are: (1) data entry and verification; (2) data entry and validation against an expected output; (3) transaction inputs for incompatible processing functions (e.g., input of contractor invoices instead of payment authorizations); and (4) data entry and supervisory authorization functions (e.g., authorizing the continued processing of a rejected transaction or approved waiver when a supervisor's review and approval are required). Other examples of risks and internal controls may be unique to each agency's mission.

- **Automating customer service interactions**: Automated Contact Centers (ACCs) can use AI and combinations of technologies to deliver high-quality customer service interactions to agency stakeholders without human intervention. Thousands of people call various government agency centers every day. AI can perform tasks related to customer interactions including customer service, data collection, customer interaction, and uninterrupted availability. A Security Operations Center (SOC), a highly specialized operation staffed by IT specialists with an information security background, may provide oversight and communication to the ACC if an incident occurs.

- **Enabling digital workers**: A non-person entity (NPE) is a physical device, virtual machine, system, service, or process that is assigned an identifier and is issued credentials to support authentication and authorization. A digital worker is a subset of NPE and can use AI or other autonomous decision-making capability to perform a business task or process like a human user. A chatbot is a common type of digital worker that interacts with human users to provide a service, such as retrieving data, answering a question, or directing the user to a resource. The Digital Worker Identity Playbook supplements guidance from Office of Management and Budget (OMB) Memorandum 19-17.

Deploying ML Algorithms

To train data in ML, a parameterized algorithm known as the model must first be deployed. To create the model, there must be a learning algorithm from which a particular dataset can learn. The model can then be deployed to generate real-time inferences (see Figure 2).

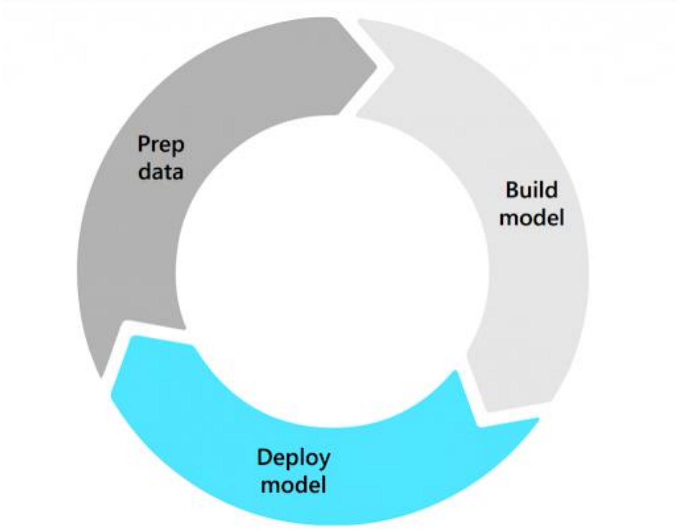


Figure 2 - Deploying ML Model.
Adapted from https://deepchecks.com.

Most ML algorithms belong to one of two categories: supervised or unsupervised. Supervised algorithms are trained using known labeled data where an input corresponds to a known output. An example is a range of Internet Protocol (IP) addresses that are labeled to correspond to a country of origin.

Unsupervised algorithms use uncategorized, unlabeled data and try to establish a meaningful relationship between the inputs and outputs. One example is grouping person entities (e.g., civilian versus contractor), devices, or access requests based on similarities in their characteristics or behavior.

Evaluating the data available – and whether it is uncategorized or unlabeled – can guide agency decisions on which ML implementation best suits a given use case. Regardless of the algorithm used (i.e., supervised or unsupervised), model performance can be retrained over time to achieve improvements, faster decision-making, and better data analysis. It also is possible to implement a Reverse Proxy to make additional authorization decisions based on the results of authentication and authorization attributes in a FICAM environment.

Sci-kit learn (a Python library) is a great resource for a deep dive on common ML algorithms.

AI/ML Tools in FICAM Environments

In a FICAM environment, several tools can perform AI/ML tasks such as algorithm training and learning. They include commercial-off-the-shelf (COTS) software such as Microsoft Azure, Power BI, ArcSight, and Splunk as well as custom languages and standalone libraries such as Python, SAS, R, MATLAB, and PL/SQL. Open-source alternatives also exist, such as KNIME, for data analytics, reporting, and integration.

Considerations for tool selection in a FICAM environment include the approval status of a tool based on an approved products list, agency policy, and senior leader endorsement. Each tool comes with trade-offs, advantages, and disadvantages.

Available tools for agency consideration include:

- Microsoft Azure Machine Learning
- Microsoft Power BI
- ArcSight
- Splunk
- Python Machine Learning Packages/Libraries
- SAS, R, and MATLAB Languages
- PL/SQL
- KINME Analytics Platform

Challenges in AI/ML Integration

Challenges accompany the integration and use of AI/ML in FICAM environments. They include adoption, data aggregation, data access, and privacy.

There is a misplaced belief that AI/ML adoption will replace human workers and jobs. We must change this narrative to one where AI/ML is seen as a resource used in conjunction with existing tools to make task performance easier and more efficient. In some instances, agencies may already use AI/ML through various deployed algorithms. The benefits of using AI/ML – and how it can support existing technology – need to be shared with the workforce to promote widespread adoption and implementation.

In a FICAM environment, multiple, disparate data sources (e.g., Identity Manager, Data Repository, and/or Master User Record) may exist. As a result, multiple copies of digital identities also exist. De-duplication and/or aggregation of such identity data is necessary. Data aggregation for sampling requires integration with a Reverse Proxy, Identity Manager, Credential Service Provider, and Authoritative Attribute Source. In addition, such environments require an aggregated data sampling and training approach.

When creating and evaluating AI/ML algorithms, agencies need to consider who can access the aggregated, stored data. Another consideration is whether the developer's permissions grant the correct level of access with the least privileges. For developers to create an efficient AI/ML model, the sample data used in training must be representative of actual user behavior over time. While algorithmic models often are created in a development environment, the best sample data often exists in the production realm.

Because FICAM systems or use cases may deal with Personally Identifiable Information (PII), data cannot simply be copied or shared from production to a lower environment. Specific privacy, regulatory, and ethical regulations govern the handling of PII. Thus, it is necessary to educate developers and the workforce about these policies and associated ethical implications.

Conclusion

AI and ML are interconnected, with ML forming a subset within the broader field of AI.

Data evaluation and analysis for ML can help guide agency decisions on which category algorithm (e.g., supervised or unsupervised) to use for training purposes. Agencies also must consider which sample data to use to train the ML models and how to obtain such sample data. The challenge goes beyond merely replicating data. The data must accurately capture the patterns evident in real-time use cases.

When it comes to a federated solution for AI, the greatest challenge will be adoption and consistent interpretation and implementation of AI standards across agency boundaries. At the very least, federal agencies should concur on a metric for measuring AI trustworthiness. If the implementation meets one agency's requirements, then other agencies may be able to consider a similar approach. The National Institute of Standards and Technology (NIST) created an AI Risk Management Framework as a resource for use in designing, developing, deploying, or using AI systems. It helps manage the many risks of AI and promotes trustworthiness and responsibility in its development and use.

AI/ML has existed for decades in our everyday applications. However, with Zero Trust guidance and policy being applied to FICAM environments, there is space to use and implement AI/ML within established policy and best practices. Some agency considerations when beginning AI/ML implementation involve strategizing and brainstorming approaches for educating the workforce, aggregating data, implementing data access controls, selecting AI/ML tools, and understanding related policy.

Electrosoft is focused on leveraging its broad range of SMEs, industry presence, and innovation to positively impact agency AI/ML FICAM efforts. Electrosoft FICAM SMEs possess expertise in technologies related to Program Management, ICAM, PKI, Cyber, Cloud, and Security. Electrosoft's strong track record of developing and delivering meaningful innovation to its federal customers arises from continuously looking for creative ways to improve outcomes. Strong day-to-day security practices for all employees, coupled with corporate processes that are routinely reviewed and improved in alignment with ISO 9001 and CMMI Level 3, contribute to a consistent capability to delight customers as evidenced by an average CPARS rating of 4.4 out of 5 on all prime contracts.

Contact Us

To learn more information about Electrosoft and our capabilities, contact us at info@electrosoft-inc.com.

About Electrosoft

Electrosoft delivers a diversified set of technology-based solutions and services differentiated by thought leadership and innovation. Fueling the success of our government and commercial customers since 2001 through outstanding value and trust, we couple our domain knowledge and experience with proven, mature management practices to deliver the right solutions on time and within budget. These practices include an ISO 9001:2015 registered Quality Management System (QMS) and Capability Maturity Model Integration (CMMI) Level 3 assessed processes. Headquartered in Reston, Virginia, Electrosoft is an 8(a) certified Small Disadvantaged Business (SDB) and an 8(m) certified Economically Disadvantaged Woman-Owned Small Business (EDWOSB). For more information about Electrosoft, visit our website at www.electrosoft-inc.com .