International Council on Systems Engineering
A better world through a systems approach

Observations in Establishing AI Practices in Highly Regulated Environments

Dr. Jose Andre Morales, Douglas J. Reynolds, Dr. Matthew Walsh, Joseph Yankel, and Hasan Yasar

INCOSE International Symposium 2025 | Ottawa, Canada

[DISTRIBUTION STATEMENT A] This material has been

SEI Markings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

DM25-0624

Topics For Today’s Talk

- Introduction
- AI and HREs
- The AI Lifecycle
- Observed Obstacles
- Recommendations
- Path Forward and Conclusion

Introduction

- We present obstacles to practicing AI development, deployment, and sustainment.
- Findings come from a multiyear observation of AI in practice at several large well-funded efforts in highly regulated environments.
- Most efforts, if not all, (partially) implemented MLOps, which is DevSecOps for AI focused on machine learning.
- Obstacles can be overcome by following or correctly implementing established guidelines and practices, including MLOps.
- We provide recommendations for overcoming the obstacles.

AI and HREs

AI Is Everywhere and Growing!!!

The AI market:

- Forecast to reach $4.8 Trillion by 2033
- AI advances and adoption by multiple industries are main drivers
- Provides efficiency, improved decision making and automation of multiple processes
- Main targeted industries of very large growth: healthcare, finance, government, transportation, and manufacturing

Word cloud: © European Union. Nativi, S. & Gómez-Losada, Á. *Artificial Intelligence at the JRC: Survey Results.* Joint Research Centre. 2019. 10.2760/54605.

AI in Highly Regulated Environments (HREs)

HRE* common characteristics:

- air-gapped physical spaces
- computer systems with heightened security and access controls
- segregation of duties
- inability of personnel to discuss certain topics outside of closed areas
- inability to take certain artifacts off premises

HRE* published guidance and governance:

- NIST AI Risk Management Framework
- European Union’s Ethics Guidelines for Trustworthy AI
- ISO/IEC 42001
- US federal AI governance
- US DoD Data, Analytics, & Artificial Intelligence Adoption Strategy
- Over 1,000 current in 2025**

HRE* areas of common usage:

- system infrastructure
- policy adherence
- security
- hardware automation
- document creation
- surveillance
- object identification
- many more…

* Morales, J.A.; Yasar, H.; & Volkman, A. Implementing DevOps Practices in Highly Regulated Environments. In *Proceedings of the 19th International Conference on Agile Software Development: Companion (XP '18)*. ACM. 2018. https://doi.org/10.1145/3234152.3234188

** Sherman, N. *AI Regulations Around the World: 2025*. Mind Foundry. 2024. https://www.mindfoundry.ai/blog/ai-regulations-around-the-world

The AI Lifecycle

Consists of Four Main Components

Four main pipelines for creating a complete development, deployment, and monitoring system:

1. Data curation
Receives raw data and converts it to a usable format for AI model development

2. Model development
Applies tools for coding, training, testing, and validating a model in staging and operational environments

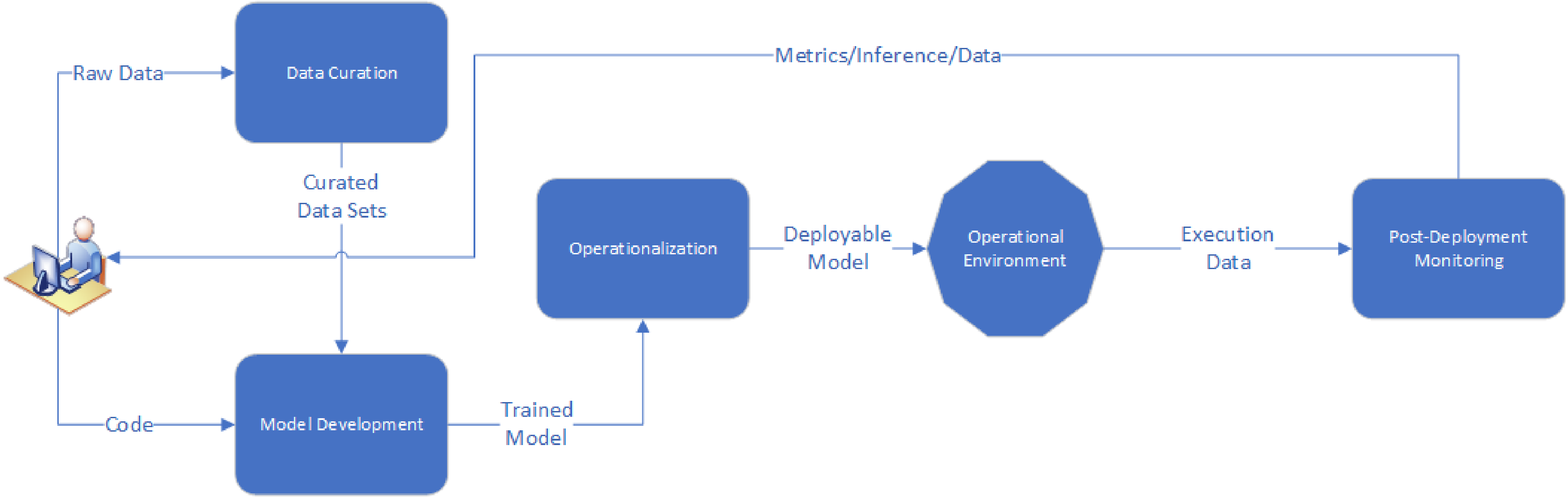
3. Operationalization
Packages a completed AI model in a usable form as a component of some larger operational system

4. Post deployment monitoring
Observes model predictions of deployed models for accuracy and collects real-world data to train models ready for deployment; leverages the other three components

When applied to machine learning, these pipelines are called MLOps.
MLOps is DevSecOps for AI focused on ML.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

High-Level Component Interaction



[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

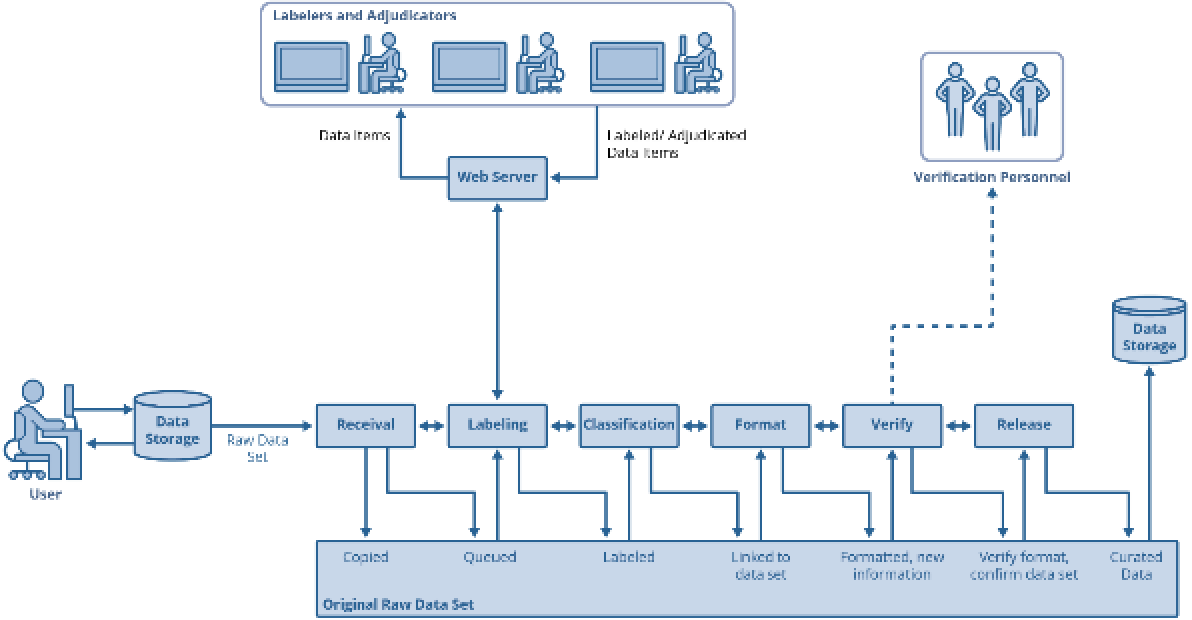
What Is DevSecOps for AI?

- **DevOps** is a set of principles and practices that emphasize collaboration and communication between software development teams and IT operations staff, as well as with acquirers, suppliers, and other stakeholders in the lifecycle of a software system.
- **DevSecOps** is a model for integrating the software development and operational process. It considers the following security activities: requirements, design, coding, testing, delivery, deployment, and incident response.
- **DevSecOps for AI** implements the steps for AI model development and deployment with additional security requirements that focus on “trusting the model.” ➡ MLOps

1. Data Curation

The data curation pipeline

- inputs raw data and outputs curated data
- implements each step to label, classify, and format data
- may need personnel to label raw data
- should run before and parallel to the dev pipeline

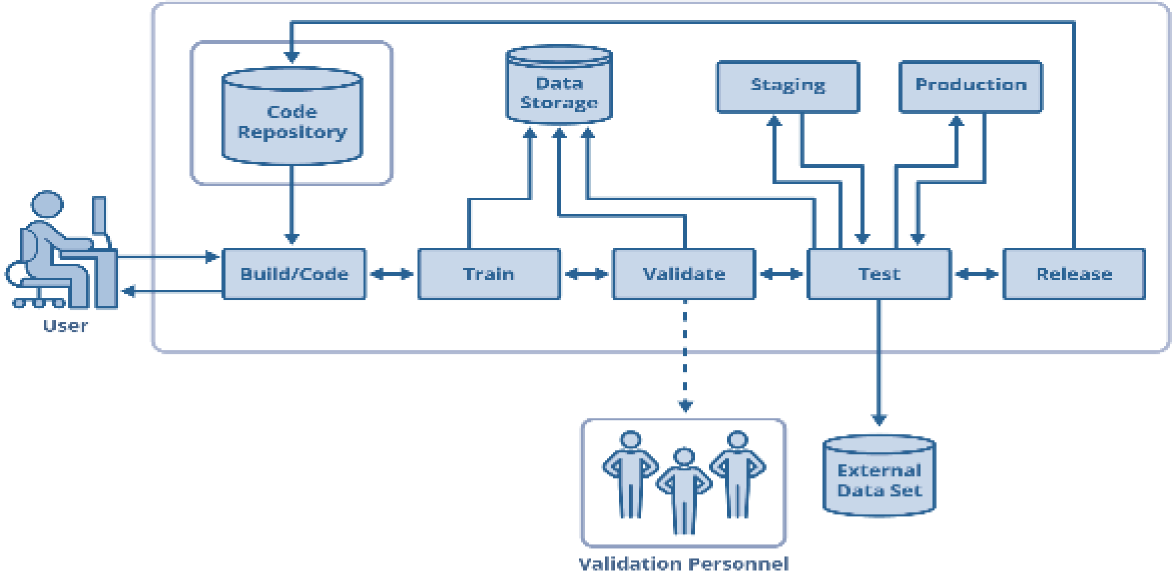


[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

2. Model Development Pipeline

The pipeline

- builds model implementation code
- trains and validates models
- includes a feedback loop at each step
- uses processed data from the curation pipeline via common storage
- outputs a trained and tested AI model

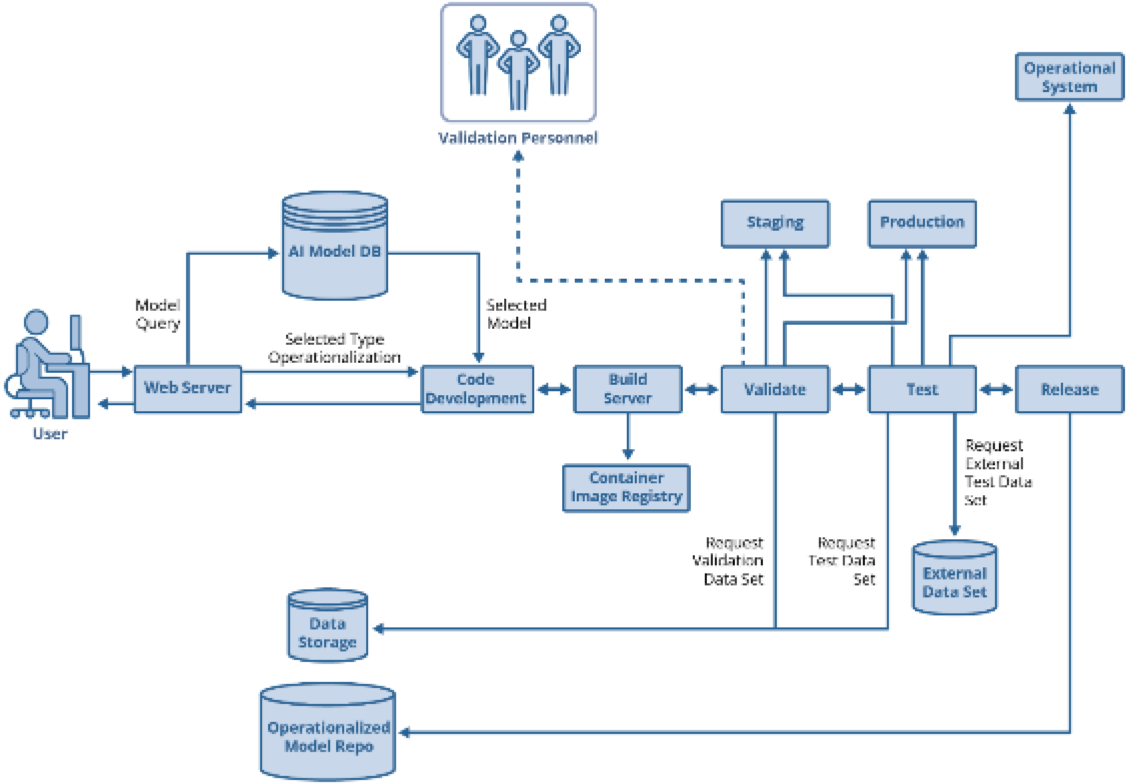


[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

3. AI Model Operationalization Pipeline

The pipeline

- makes models that are usable in the real world
- packages the model in a deployable artifact
- may need a data curation pipeline
- tests in operational environments and systems
- provides public methods for data ingress, prediction, and continuous monitoring

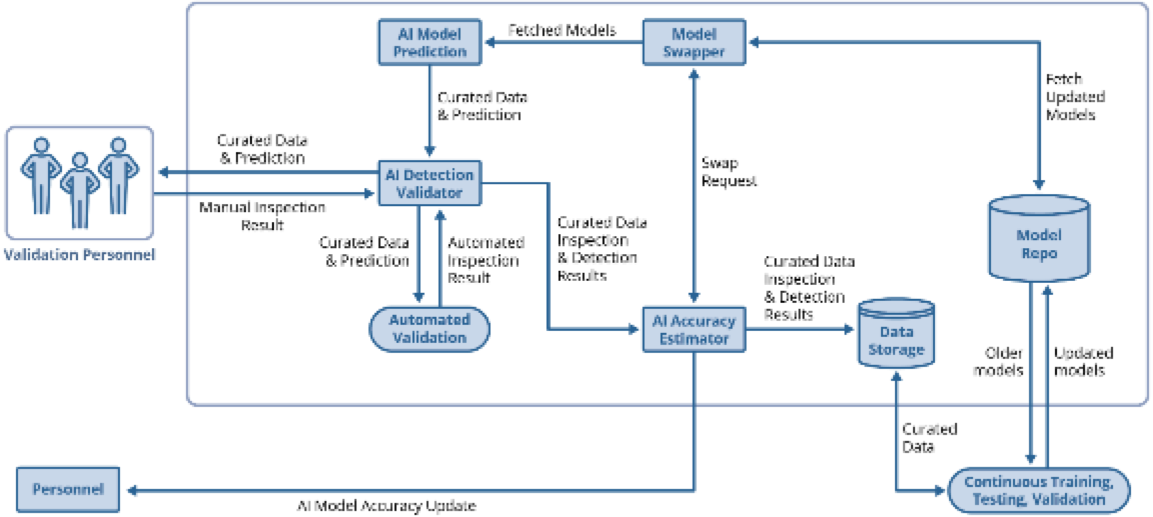


[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

4. Post-Deployment Monitoring

Monitoring

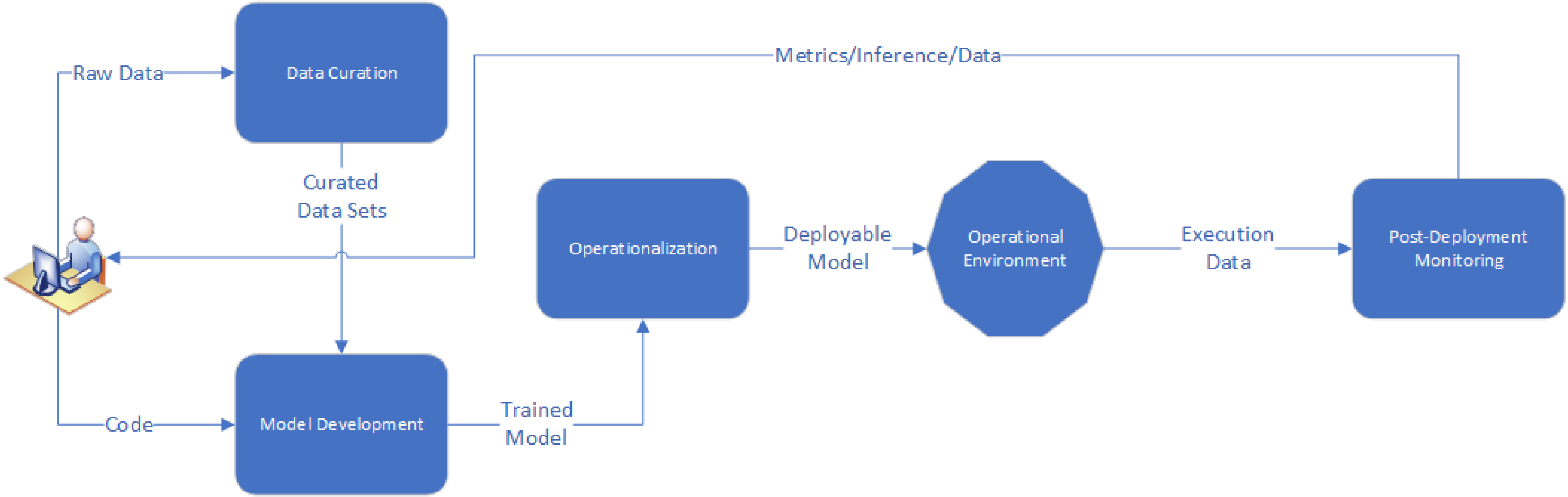
- starts with an AI model prediction
- works with operationally deployed models
- validates each model prediction
- constantly trains models with real-world data
- can dynamically swap models
- may include personnel for prediction validation



[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

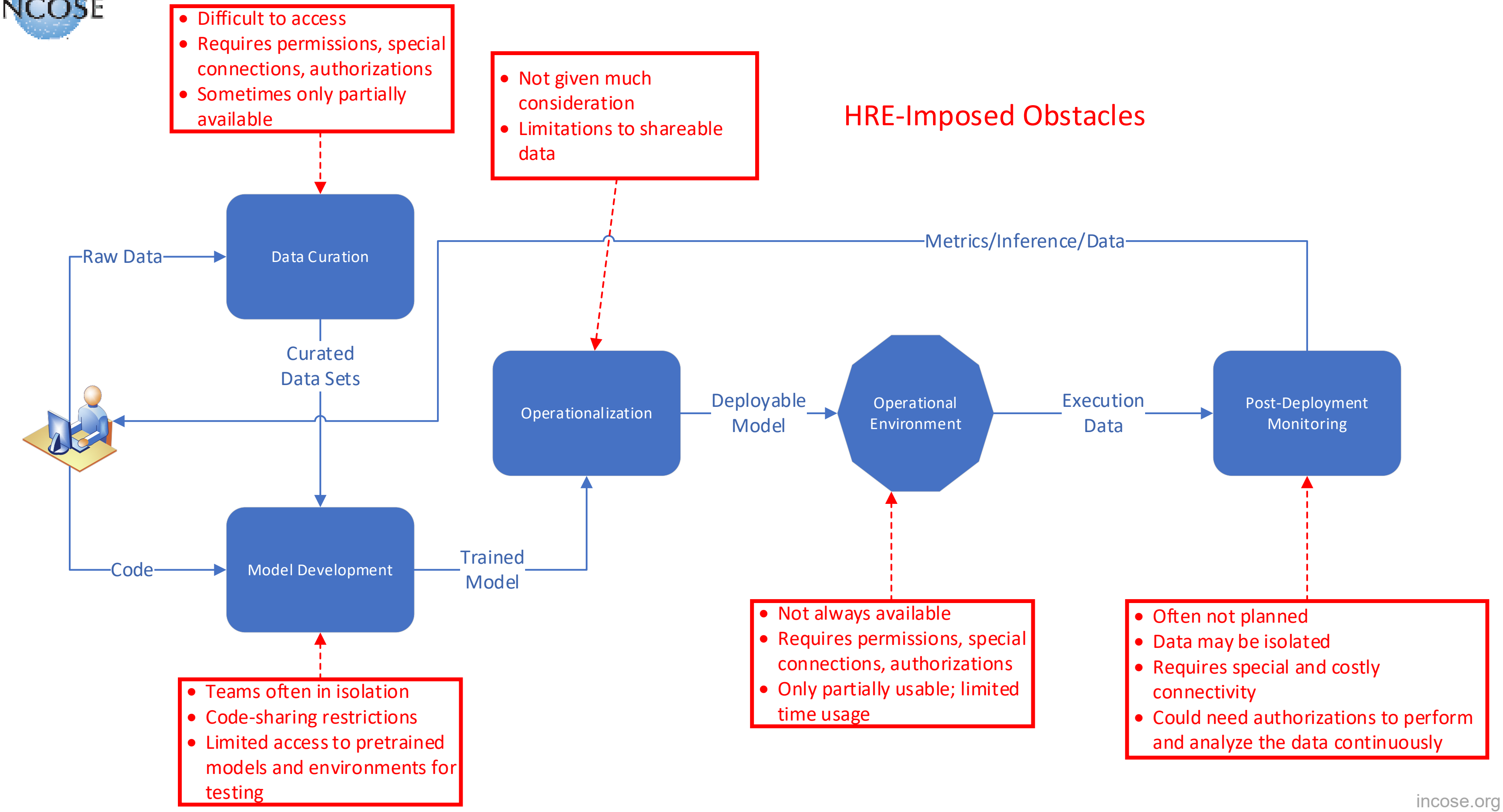
Observed Obstacles

High-Level Component Interaction



[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

HRE-Imposed Obstacles



- Difficult to access
- Requires permissions, special connections, authorizations
- Sometimes only partially available

- Not given much consideration
- Limitations to shareable data

- Teams often in isolation
- Code-sharing restrictions
- Limited access to pretrained models and environments for testing

- Not always available
- Requires permissions, special connections, authorizations
- Only partially usable; limited time usage

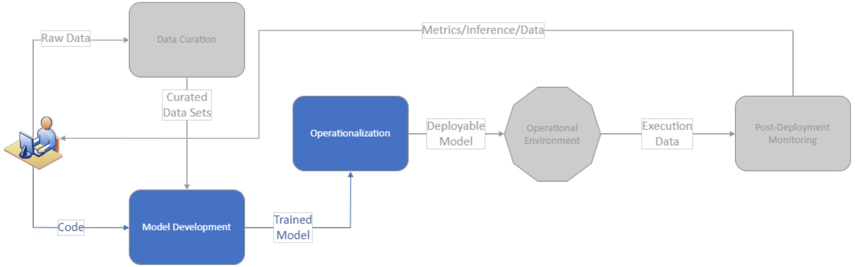
- Often not planned
- Data may be isolated
- Requires special and costly connectivity
- Could need authorizations to perform and analyze the data continuously

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

Lack of a Standard Process and Environment

There is no standard process and environment for AI model development.

- In some cases, this resulted from an HRE requirement for isolation of technical boundaries between projects.
- In other cases, lacking a standard project approach led teams to
  - select their own tools, create environments, and develop practices
  - tailor them based on community best practices and their project needs

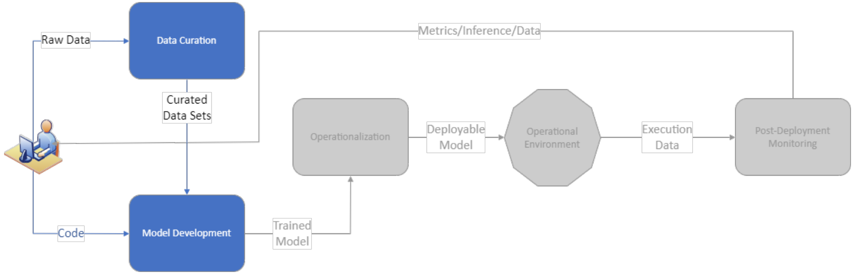


- Diversity of AI requires different approaches, but no standard is available for any approach or even for a narrow domain.
- ML approaches and computer vision domains were most popular.
- Project teams employed their own method of development with their own tool sets and development environments.

Hardened Data Set Acquisition and Curation

Teams are hindered, some by HRE policies, when acquiring desired data sets.

- Very few projects got the data they wanted.
- Several projects had hardened acquisition from HRE-imposed restrictions on data sharing and may only share a portion after several agreements and paperwork.
- Obstacles to data sharing included
  - lack of data owner motivation to share
  - distribution restrictions and requirements by the HRE
  - lengthy amounts of HRE-required paperwork and approvals that take impractically long periods of time

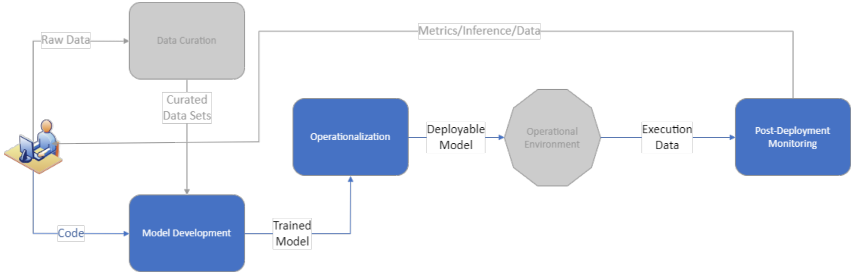


- Acquired data often required lengthy periods of time and resources (funding, equipment, and personnel) to format and label.
- Data was received unlabeled or with unusable labels.
- Some useful labeled data was not shared due to HRE restrictions, mostly based on ownership claims and associated regulations.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

Exclusion of Test Focusing on Unwanted Results

Testing focused on what the model should find.

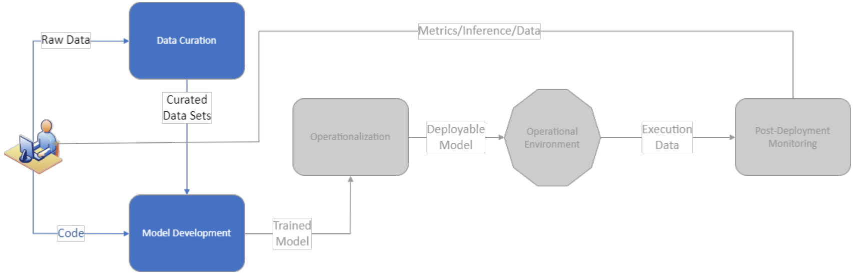


- HRE time restrictions disallowed adequate testing.
- HRE team isolation—via interacting with other teams, the customer, and SMEs—limited learning about what a model should not do.
- Models had minimal to no testing to ensure that unwanted results were not produced.
- Testing focused on confirming the presence of desired elements.
- The test sets were populated with true-positive samples.
- False positives, and true and false negatives, were often left out.
- System integration tests and operational tests were not observed.
- Models were tested in isolation and not as part of an intended larger system.

HRE Guidance Lacked AI Focus

HRE security guidelines lacked focus on AI-relevant requirements.

- Security requirements covered categories typically seen in HREs, including source code flaws, vulnerabilities, malware, and blocked/ allowed file lists.
- Satisfaction was based on test results—including a malware scan, static code analysis, and a vulnerability scan—and a list of all files present in a software bill of materials (SBOM).
- Guidelines did not address the following two AI-centric components:
  1. data used for model training and testing
  2. performance of the trained model once deployed into the real world

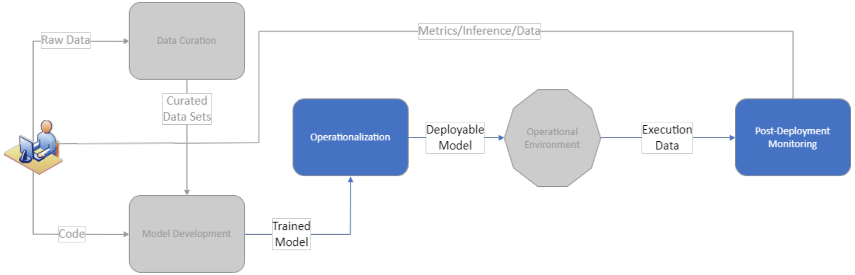


- The guidelines lacked an AI focus and instead treated an AI project as a software/system development project.
- This lack can compromise the data via data poisoning and purposeful training deficiencies, resulting in model drift far sooner than anticipated.

Operationally Ready Models Lacked a Standard Interface

Trained model packaging lacked a standard.

- HRE team isolation and required NDAs limited communication with operators and production environment owners.
- Teams operationalized a trained model for real-world deployment in their own subjective way based on partial information.
- In some cases, models were developed as prototypes, where operationalization is typically not considered. Later, rebuilds for operational use were difficult to do even without standards.



This led to

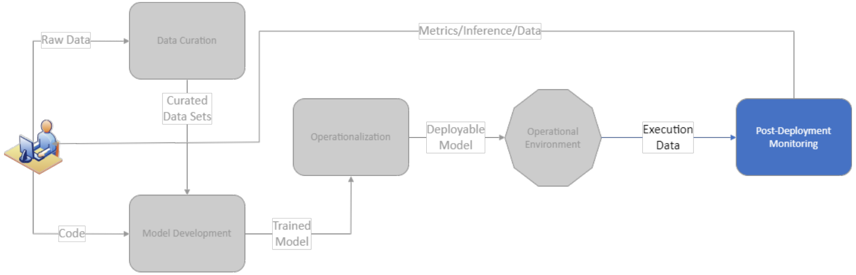
- lots of re-engineering and delays for a model to work with its intended production environment
- needed sustainment of multiple CI/CD pipes for container rebuilding
- customized usage instructions for each model
- inability to issue needed changes, such as for security, across multiple containers

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

Deployed Models Lacked Monitoring Infrastructure

Teams focused on development and deployment and not on tracking a model's performance after deployment in the real world.

- In HREs, long-term connectivity, data collection, and analysis tools were very difficult to achieve and, if feasible, only with lots of required resources, authorizations, and permissions.
- The goal became to develop and deploy the model and leave for others to handle anything post-deployment.
- We did not observe efforts to include metrics gathered within the container to measure container and model performance.



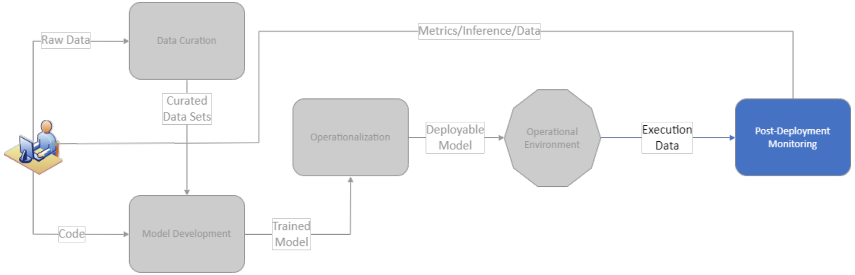
This lack

- makes it difficult to detect drift on a timely basis.
- obscures awareness of suboptimal execution performance for the container and model

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

No Approach to Operational Model Retraining and Redeployment

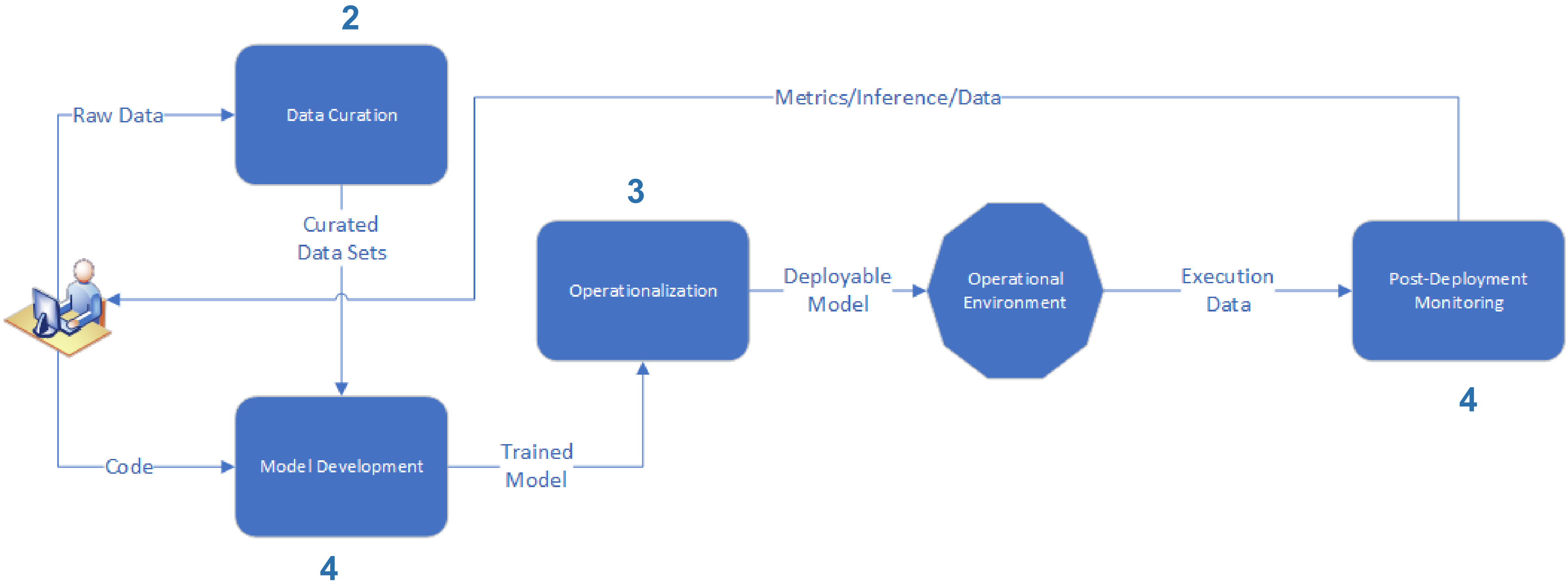
There is no process in place to retrain or update deployed models.



- HREs imposed strong restrictions on data retention, analysis, and sharing and using findings
- Redeploying models to an operational environment required several agreements between multiple HREs.
- Deployed models were in operation without persistent performance monitoring.
- No process exists to collect the real-world data analyzed by the model for retraining or other purposes.
- There is no system in place to retrain the existing model.
- If a determination was made to modify or replace the currently active model, the mechanics to do so were not defined.

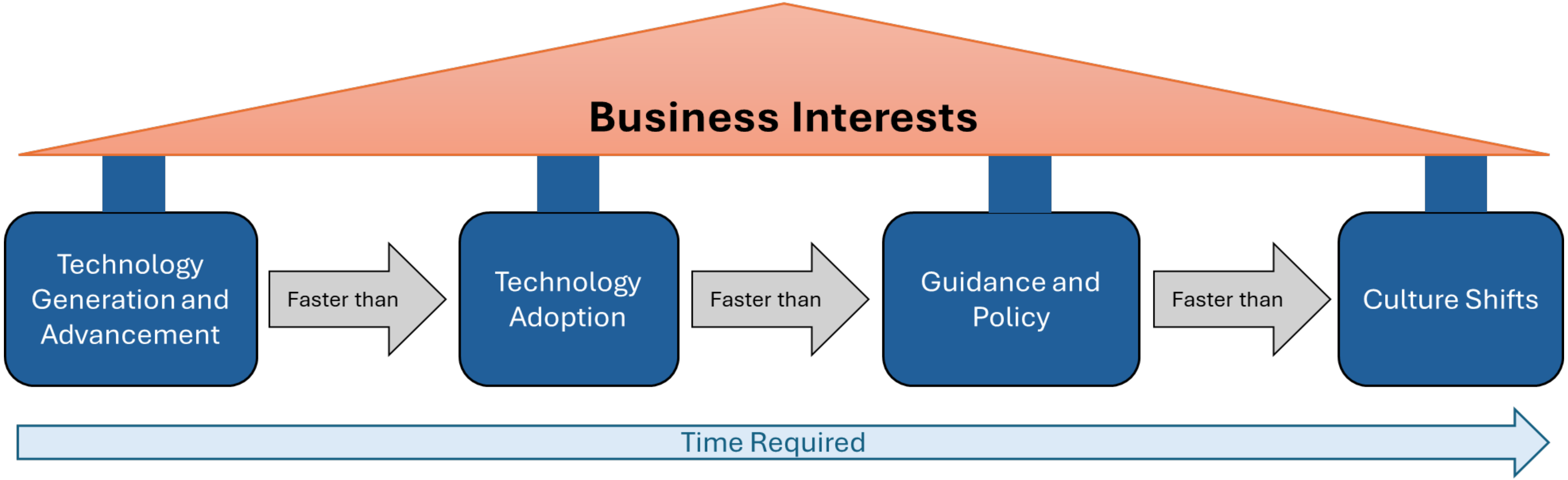
[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

Number of Impacts to High-Level Components



[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution

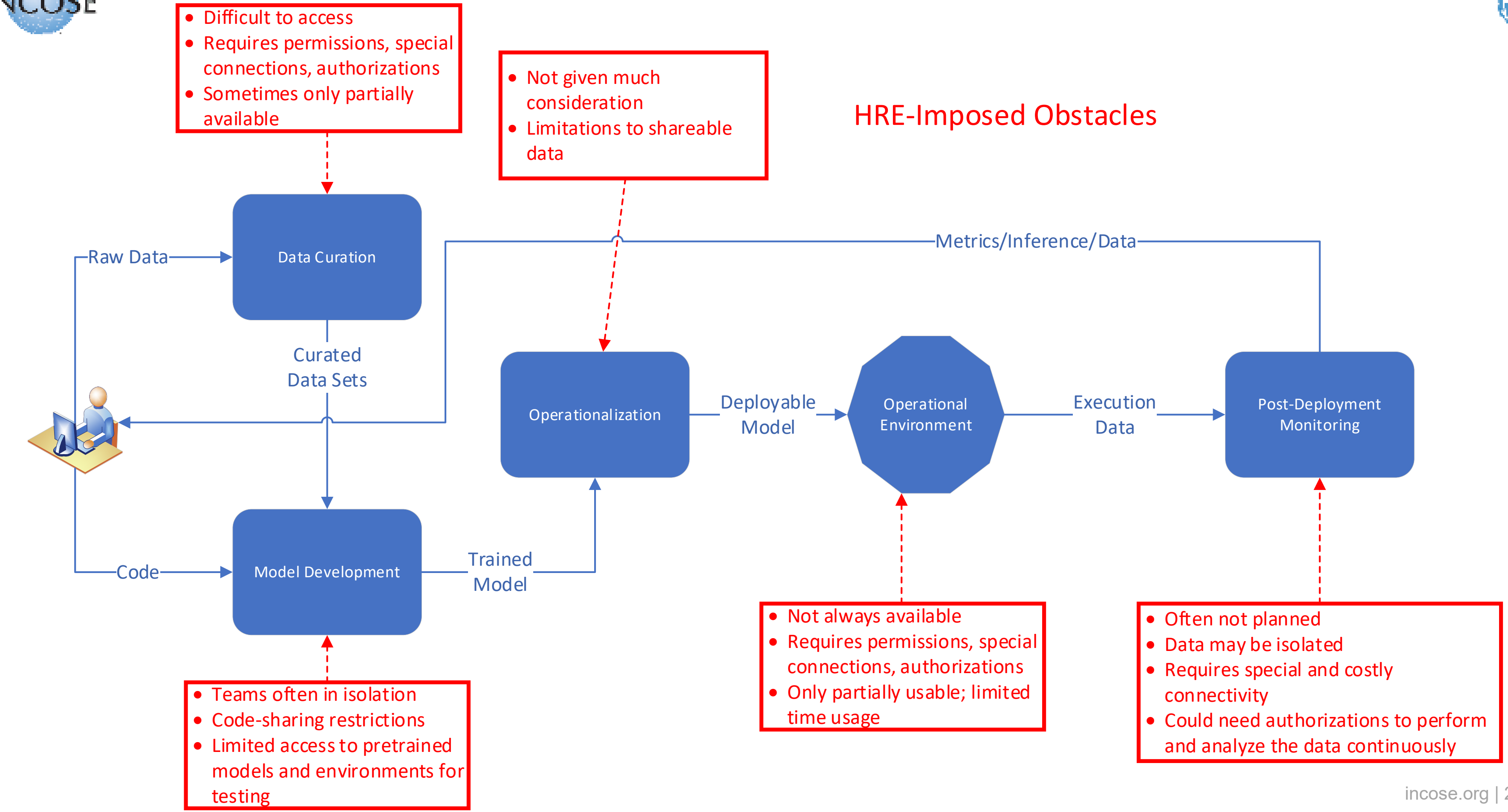
Key Observed Drivers of Noted Impacts



Business interests are overarching driving factors that affect an organization’s ability to change. Technology advancement is often faster than adoption, which is trailed by guidance and policy and moves leaps beyond the culture shifts required to realize true value. The timeline is often exaggerated in HREs, where additional requirements, data sharing, and cross-functionality are even more limiting than traditional environments.

Recommendations

HRE-Imposed Obstacles



- Difficult to access
- Requires permissions, special connections, authorizations
- Sometimes only partially available

- Not given much consideration
- Limitations to shareable data

Raw Data

Data Curation

Curated Data Sets

Metrics/Inference/Data

Operationalization

Deployable Model

Operational Environment

Execution Data

Post-Deployment Monitoring

Code

Model Development

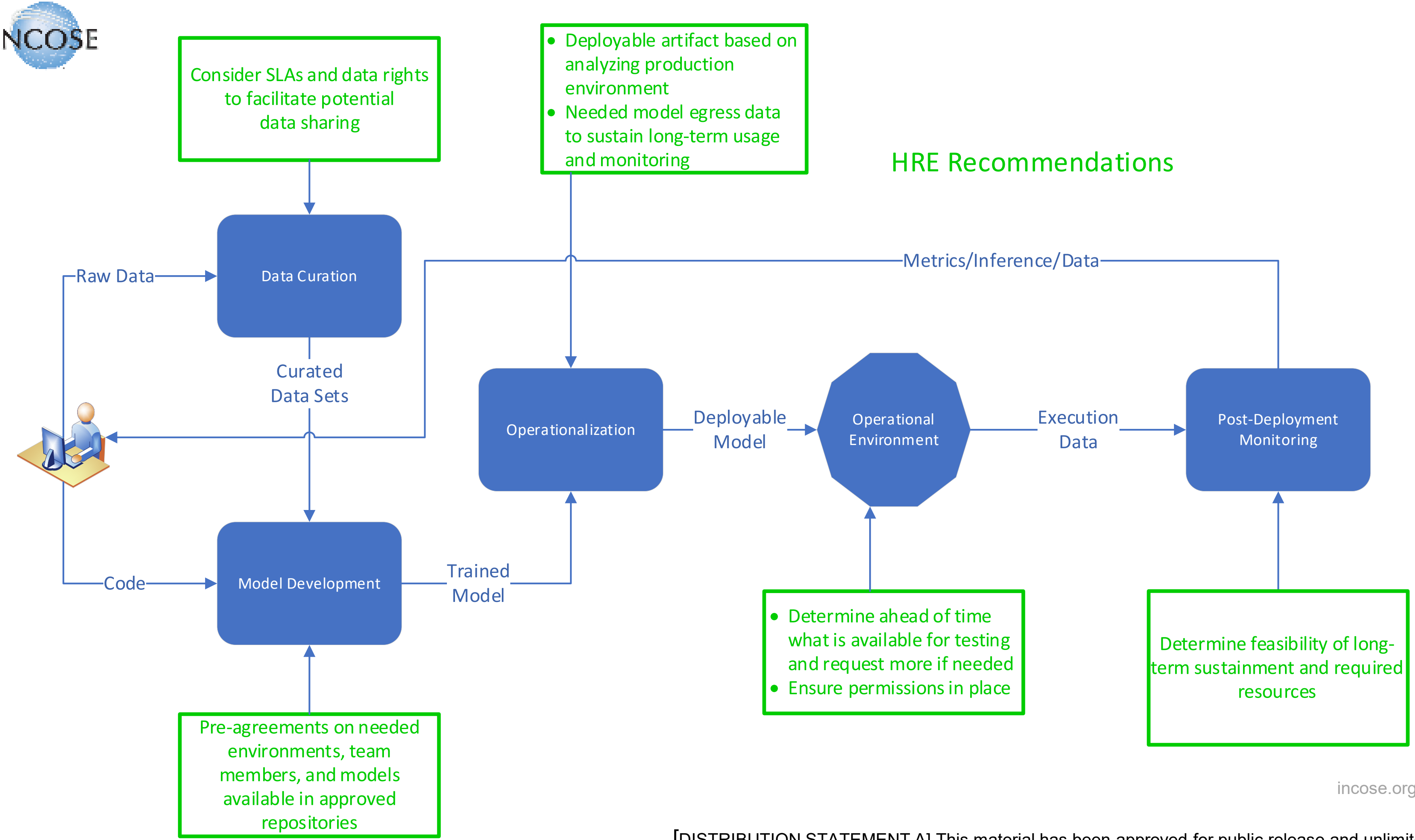
Trained Model

- Teams often in isolation
- Code-sharing restrictions
- Limited access to pretrained models and environments for testing

- Not always available
- Requires permissions, special connections, authorizations
- Only partially usable; limited time usage

- Often not planned
- Data may be isolated
- Requires special and costly connectivity
- Could need authorizations to perform and analyze the data continuously

HRE Recommendations



Consider SLAs and data rights to facilitate potential data sharing

- Deployable artifact based on analyzing production environment
- Needed model egress data to sustain long-term usage and monitoring

Raw Data

Data Curation

Metrics/Inference/Data

Curated Data Sets

Operationalization

Deployable Model

Operational Environment

Execution Data

Post-Deployment Monitoring

Code

Model Development

Trained Model

- Determine ahead of time what is available for testing and request more if needed
- Ensure permissions in place

Determine feasibility of long-term sustainment and required resources

Pre-agreements on needed environments, team members, and models available in approved repositories

Guidance to Alleviate Observed Obstacles

HRE policies need adaptation in multiple areas:

- requirements gathering fully inclusive of all project perspectives, including security and governance needs, at the commencement of the project
- a documented standardization for
  - development, testing, deployment environments, and pipelines
  - data curation, labeling, and formatting
  - operationalization and post-deployment monitoring
- guidance on inclusions in model operationalization to sustain both long-term monitoring and effective model retraining, updates, and replacements
- revised HRE security guidance to include the AI use case and AI-specific elements
- facilitation of an HRE implementation of DevSecOps for AI principles for projects from conception to post-deployment sustainment
- identification of desired data and potential HRE-related obstacles at project commencement

Path Forward and Conclusion

The Future Is Bright

Conclusions

- Observed obstacles can be overcome with established methods.
- Cultural shift is key. Organizational focus should be on the entire AI lifecycle and not a subset of components.
- HREs must amend their policies to remove discussed restrictions and reduce or eliminate some authorizations and permissions.

Future work

- Revisit and inquire whether changes occurred and what their impact was.
- Continue observing several new projects employing AI practices.
- Report on continuance of previously observed obstacles and new ones.
- Create AI practice guidance that reduces chances of these obstacles occurring.

Thank you.

Dr. Jose Andre Morales

Senior Researcher

Software Engineering Institute

Carnegie Mellon University

jamorales@sei.cmu.edu

Douglas J. Reynolds

Senior Researcher

Software Engineering Institute

Carnegie Mellon University

Dr. Matthew Walsh

Senior Researcher

Software Engineering Institute

Carnegie Mellon University

Joseph Yankel

Senior Researcher

Software Engineering Institute

Carnegie Mellon University

Hasan Yasar

Senior Researcher

Software Engineering Institute

Carnegie Mellon University