

Transparent Models or Black Box? Using Generative AI to Generate Real-World Intelligence

Sherrine Eid, MPH

Global Head, Epidemiology, RWE & Observational Research

Samiul Haque, PhD

Sr. Solutions Architect

SAS Institute, Inc

ML34



Agenda

1 RWE In Life Sciences Today

2 Challenges

3 Bias and AI

4 Ethical Considerations

5 Opportunities

RWE in Life Sciences Today



Regulators are leaning in...



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Search

Medicines ▾ Human regulatory ▾ Veterinary regulatory ▾ Committees ▾ News & events ▾ Partners & networks ▾ About us ▾

Home > About us > How we work > Data in regulation: Big data and other sources > Real-world evidence

Real-world evidence

Share

Making greater use of real-world evidence and real-world data can improve the evidence base for benefit-risk decisions. This can help bring medicines to patients. Real-world data are observational data stored in repositories such as electronic health records and disease registries.

HumanVeterinaryData on medicinesRegulatory and procedural guidanceMedicines

Page contents

Darwin EU

Use of real-world evidence

HMA-EMA catalogues of real-world data sources and studies

Guidance on real-world evidence

International collaboration on real-world evidence

Related content

External content

Related EU legislation

The European Medicines Agency (EMA) and the [European medicines regulatory network](#) are working to establish a sustainable framework enabling better integration of real-world data and real-world evidence into regulatory decisions.

For more information, see:

- [Data in regulation: Big data and other sources](#)

Darwin EU

EMA has established a **coordination centre** to provide timely and reliable evidence on the use, safety and effectiveness of human medicines, including vaccines, from real-world healthcare databases across the European Union (EU).

This capability is called the **Data Analysis and Real World Interrogation Network (Darwin EU)**.

For more information, see:

- [Data Analysis and Real World Interrogation Network \(DARWIN EU\)](#)

Use of real-world evidence

A guide is available on how EMA can help generate real-world evidence.

It is meant for EU regulators and decision-makers, including EMA's scientific committees, working parties and groups, as well as [national competent authorities](#), healthcare technology assessment bodies and payers.

It covers:

- how the mentioned stakeholders can request **real-world data studies** from EMA;
- what types of studies can be performed:

FDA NEWS RELEASE

FDA Eliminates Major Barrier to Using Real-World Evidence in Drug and Device Application Reviews

For Immediate Release: December 15, 2025

The U.S. Food and Drug Administration (FDA) today removed a key limitation on the use of real-world evidence (RWE) used in drug and device applications reviews. In new guidance for certain types of medical device submissions, the agency states it will accept RWE without requiring that identifiable individual patient data collected from real-world data sources always be submitted in a marketing submission. The FDA similarly intends to consider updating its guidance for drugs and biologics.

While RWE has been promoted as an advance in regulatory policy, past agency expectations have meant that most RWE could not be considered in product applications. Since 2016, 35 drugs, biologics, or vaccines have included RWE in their applications. The integration of RWE into device approvals has been more extensive, with over 250 premarket authorizations including RWE for the same period, although even for devices the rate of RWE-based authorizations has plateaued in recent years.

Historically, the FDA has insisted that any RWE submitted to the agency include private, confidential information at the individual patient level. This approach makes it impractical to use most large databases with valuable macro-level data.

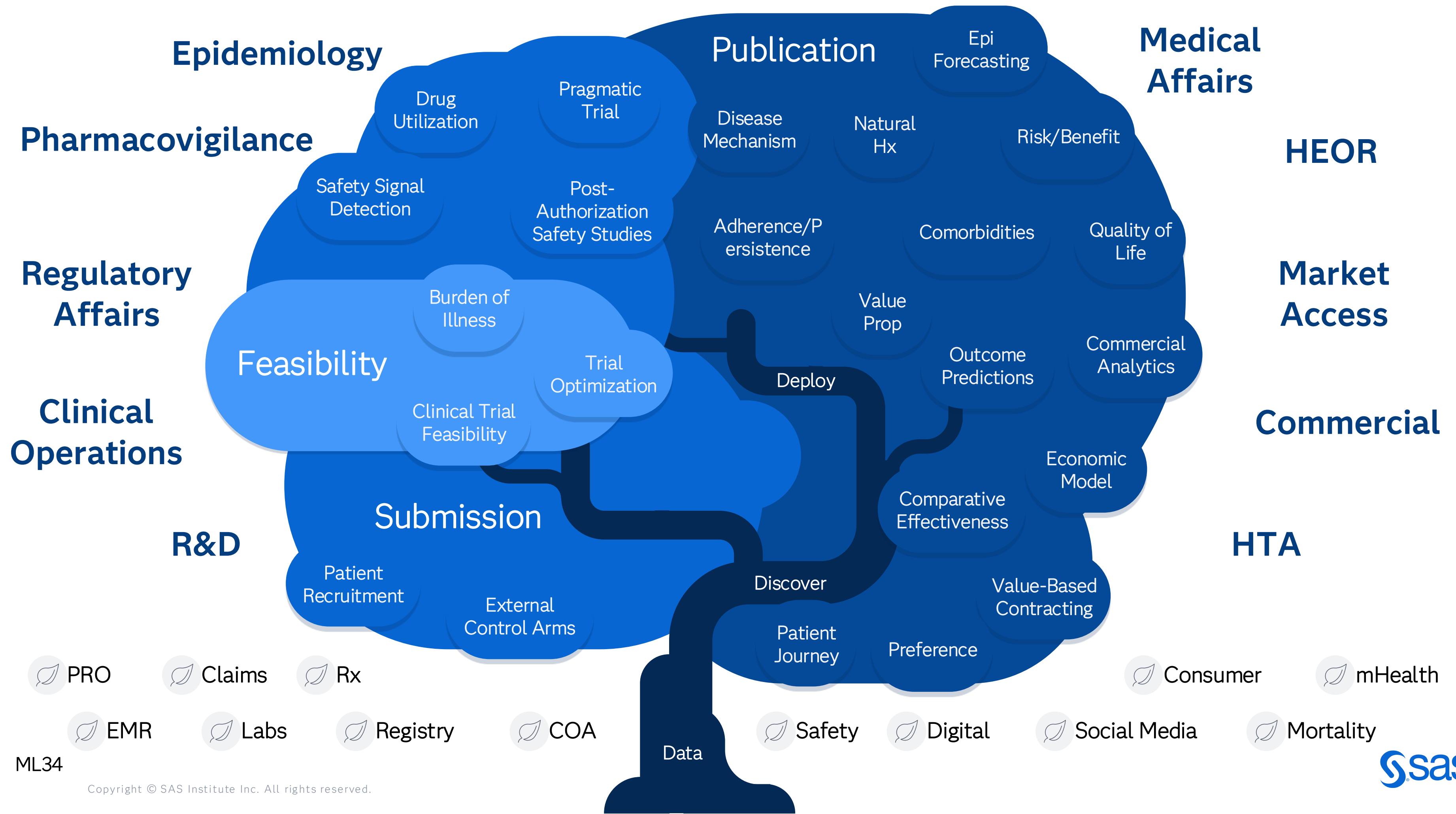
The FDA is responding to the position of many sponsors and data scientists that meaningful information can be extracted from some big data sources without private, individual information. FDA reviewers will now consider the strength of submitted RWE on an application-by-application basis.

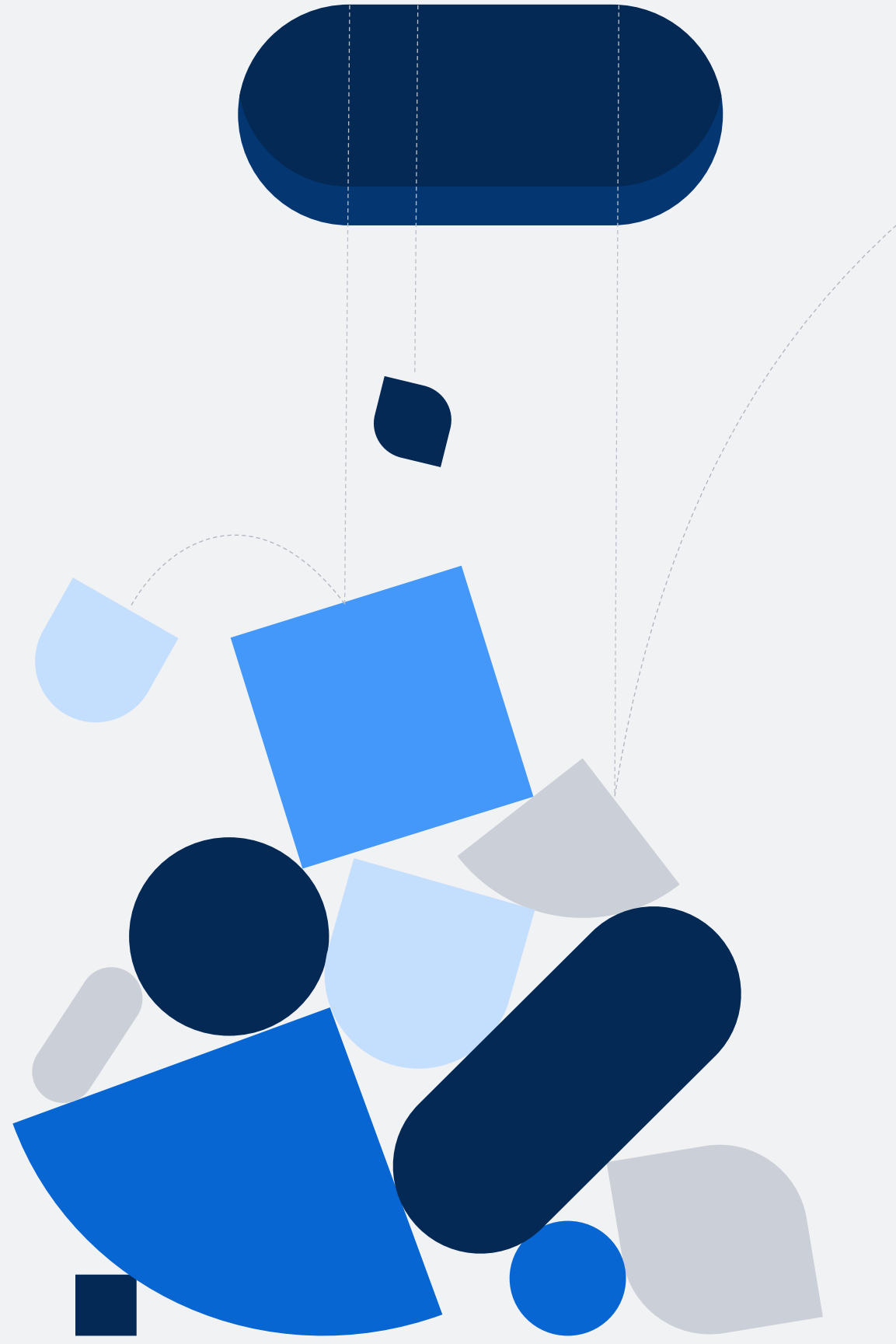
"We're removing unnecessary barriers that have prevented us from using powerful real-world evidence to get life-changing treatments to patients faster," **said FDA Commissioner Marty Makary, M.D., M.P.H.** "This common-sense reform will unlock access to vast databases like cancer and cystic fibrosis registries that contain critical insights about how treatments work in the real world."

This policy change opens the door to using de-identified databases containing millions of

ML34

Copyright © SAS Institute Inc. All rights reserved.





RWE Challenges

■ Time constraints

The process to prepare Real-World Data (RWD), identify populations of interest and generate evidence is slow.

■ Governance

Access to RWD data and insights is often uncontrolled.

■ Fit for purpose data

RWD sources don't always meet the quality needed for rigorous research.

■ Bias

Inherent selection, recall and detection bias undermines the validity of insights.

■ Limited ROI

Only a few experts within an organization have access to RWD and insights.

Data Gaps & Challenges

- Regional regs
- Missing data elements
 - BMI, mortality, smoking status
- Incorrect data
- Limited validation
- Ordered Rx
- IP vs. OP
- Dx can be inconsistent
- Governance

EMR



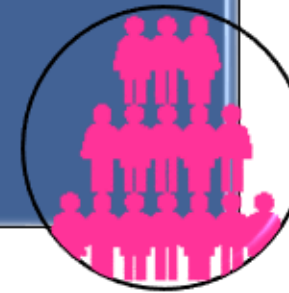
- Rx vs Dx
- Linked to EMR?
- Billing-based
- Limited Validation
- Filled Rx
- Governance
- Changes in providers make fragmented journeys

Claims



- Incomplete data
- Variable differences
- No standardization
- Source documentation
- Limited follow up
- Governance

Registries



- Inadequately documented changes in medical status
- Inconsistent visits/documentation
- Unstructured data capture of key elements
- Limited Patient Reported Outcomes
- Undocumented/fragmented changes in responses to Rx

Overall



Process Challenges

How are you addressing these today?

Key Challenges

- Interoperability and data linking
- Part 11 regulations (21 CFR part 11)
 - Quality, authenticity, and reliability of electronic records
 - Creation to modification, maintenance, archiving, retrieval, or transmission
- GxP Compliance
- HIPAA
- Audit of secondary data sources
 - Informed consent
 - System access
 - Records

Solutions

- Data Access Engines/Connectors & FHIR Adaptor
- Data Access & Management Suite
- Track record with LSAF, CA and Validated deployments
- Data Information Catalogue
- Data Management and Encryption
- Lineage, FWA Solutions and technology
- Probabilistic Methods

BIAS and AI

Applications and operationalization of responsible innovation in sensitive and large data.



The Problem with Bias

DISCLOSE DATA

A person, process, or system creates, and publishes/ shares data

STAGE AND PREPARE DATA

A person, process, or system transforms, moves, or analyzes data.

CONSUME DATA

A person, process, or system benefits from manipulated data

ACQUIRE

Ingest data from sensors, systems, or humans, recording its provenance and consent for use wherever possible

STORE

Record data to a trusted location that is both secure and easily accessible for further manipulation

PRE- PROCESS

Combine disparate datasets to create a larger dataset that is greater than the sum of its parts

MODEL BUILD

Examine and transform data with the purpose of extracting information and discovering new insights

MODEL DEPLOY
Apply the insights gained from data analysis towards making decisions, affecting change, or delivering a product or service

SHARE/ SELL
Provide Access to datasets or data insights to new sets of data manipulators

DISPOSE
Remove data from servers to prevent future release or use

MANAGE DATA

Availability Bias
Recall Bias

DEVELOP MODELS

Exclusion Bias
Pre-processing Bias
Measurement Bias
Time-interval bias
Historical Bias
Sample Selection:
Selection Bias
Attrition Bias

Confirmation Bias
Cause/ Effect Bias
Confounding Bias
Collider Bias
Prediction Bias
Performance Bias
Hindsight Bias
Chronological Bias
Funding Bias
Proxy Bias

DEPLOY INSIGHTS

Automation bias
Deployment Bias
Drift Bias
Reporting Bias

COMMON BIASES

Ethical AI Solutions





By 2028, enterprises using AI governance platforms will achieve 30% higher customer trust ratings and 25% better regulatory compliance scores than their competitors.

Gartner 2025 Top Strategic Technology Trends



Principle-Driven Approach

Human-Centricity



Promote human **well-being**, human **agency** and **equity**.

Inclusivity



Ensure **accessibility** and include **diverse perspectives** and **experiences**.

Accountability



Proactively identify and mitigate adverse impacts.

Transparency



Explain and **instruct** on usage openly, including potential risks and how decisions are made.

Robustness



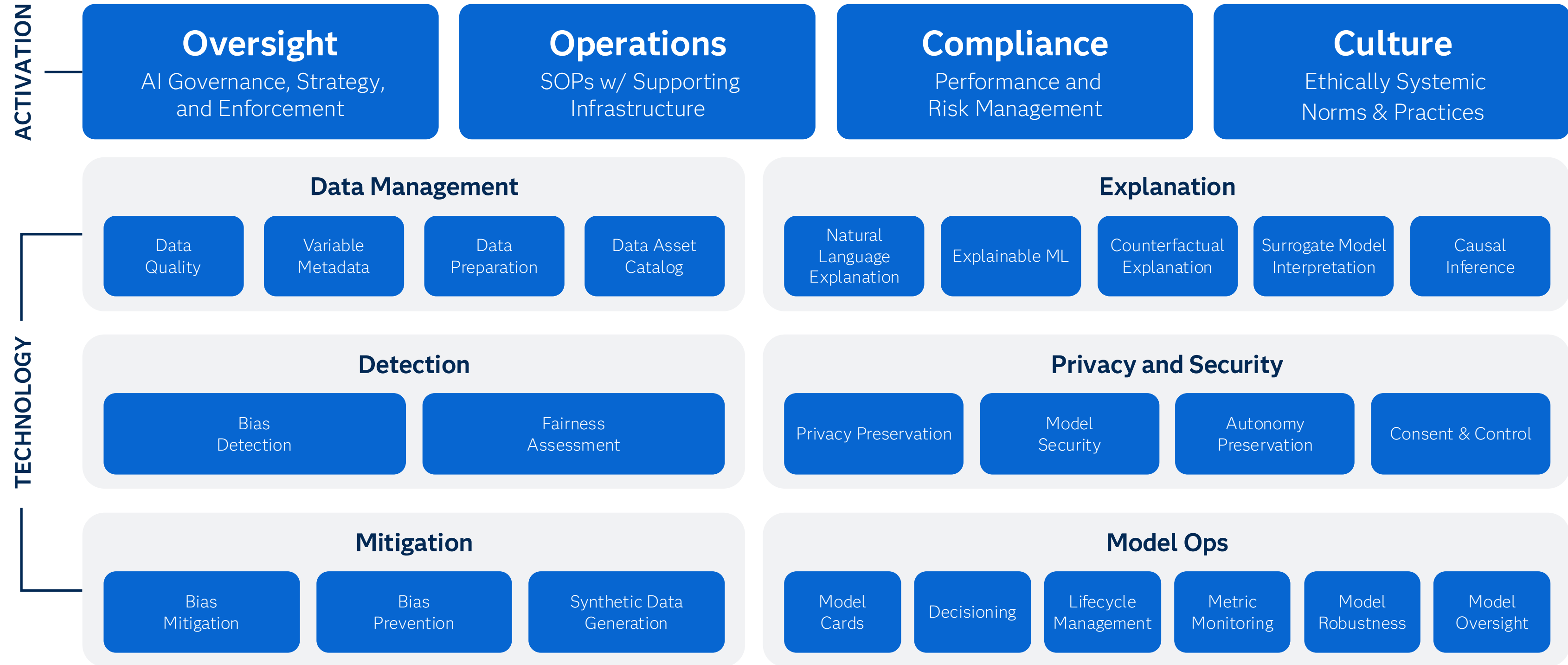
Operate **reliably** and **safely**, while enabling mechanisms that assess and manage potential risks throughout a system's lifecycle.

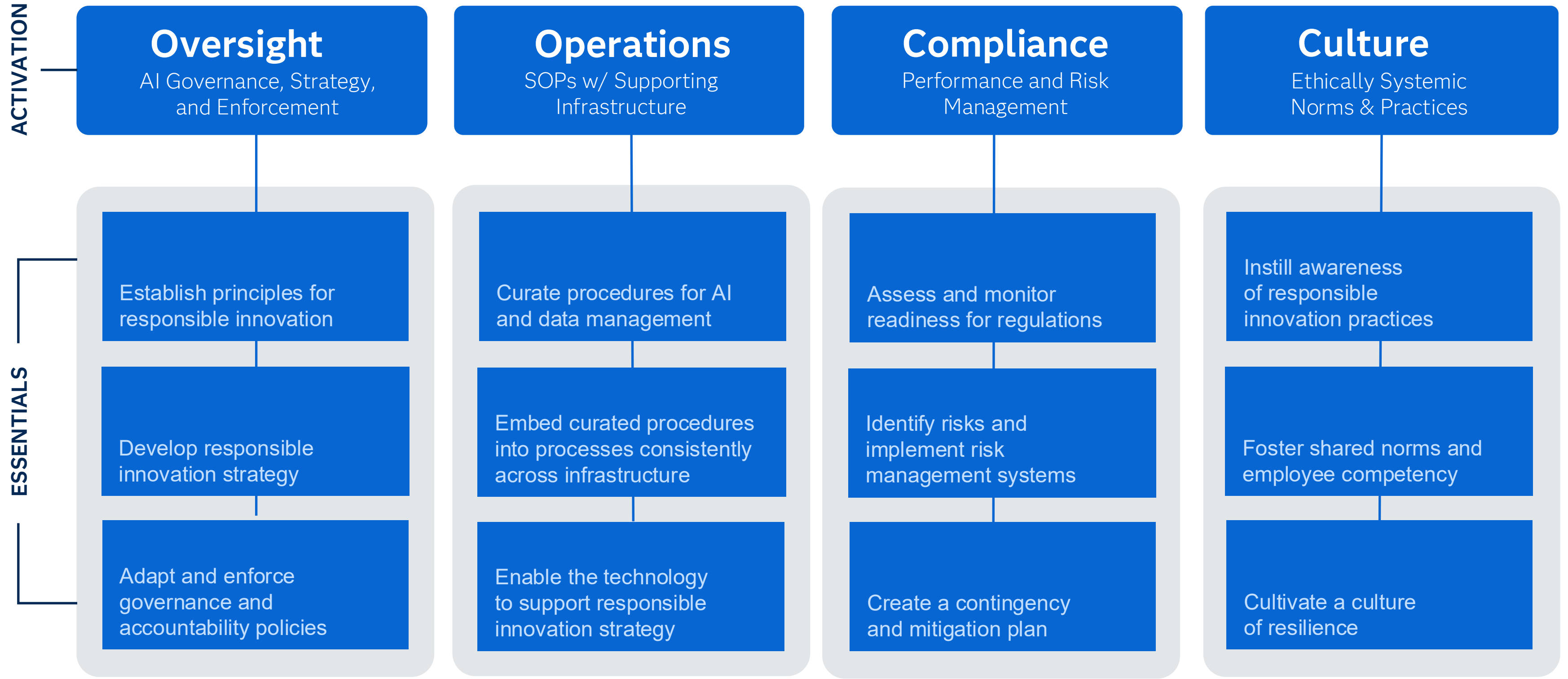
Privacy & Security



Respect the privacy of data subjects.

Trustworthy AI Landscape



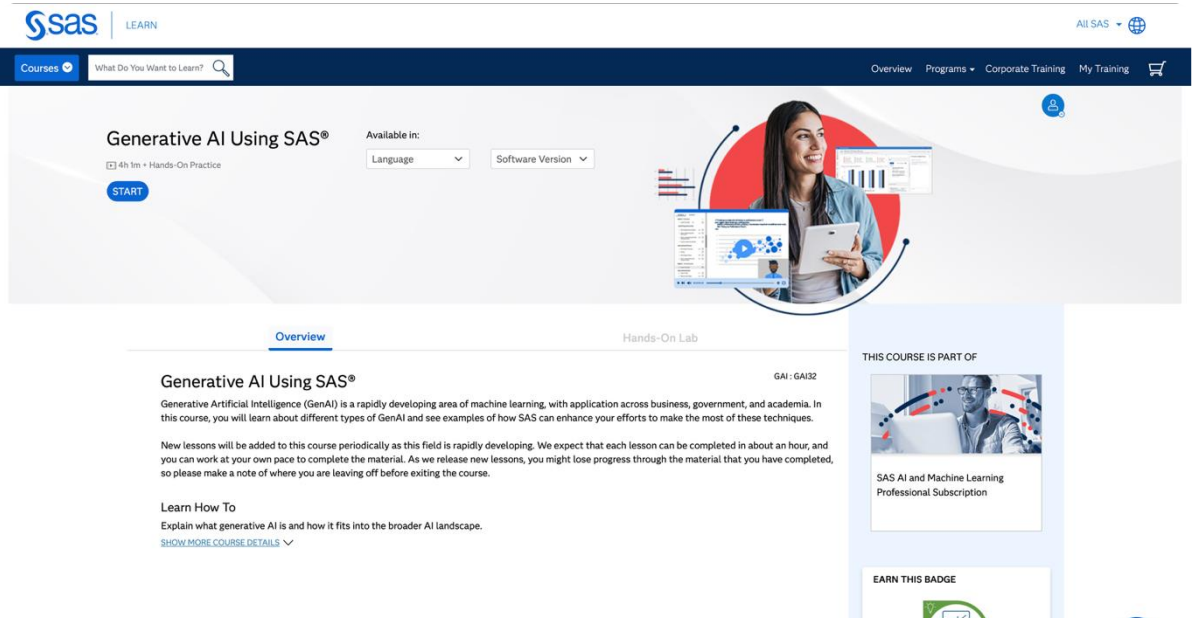


Let's Explore Regulations

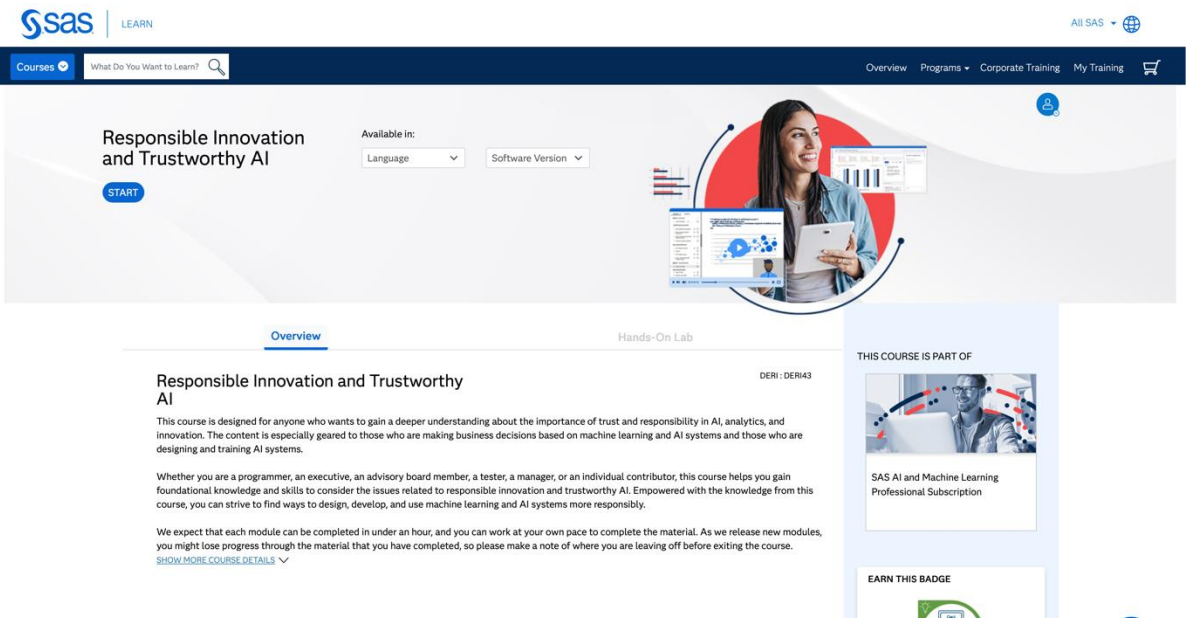
EU AI Act



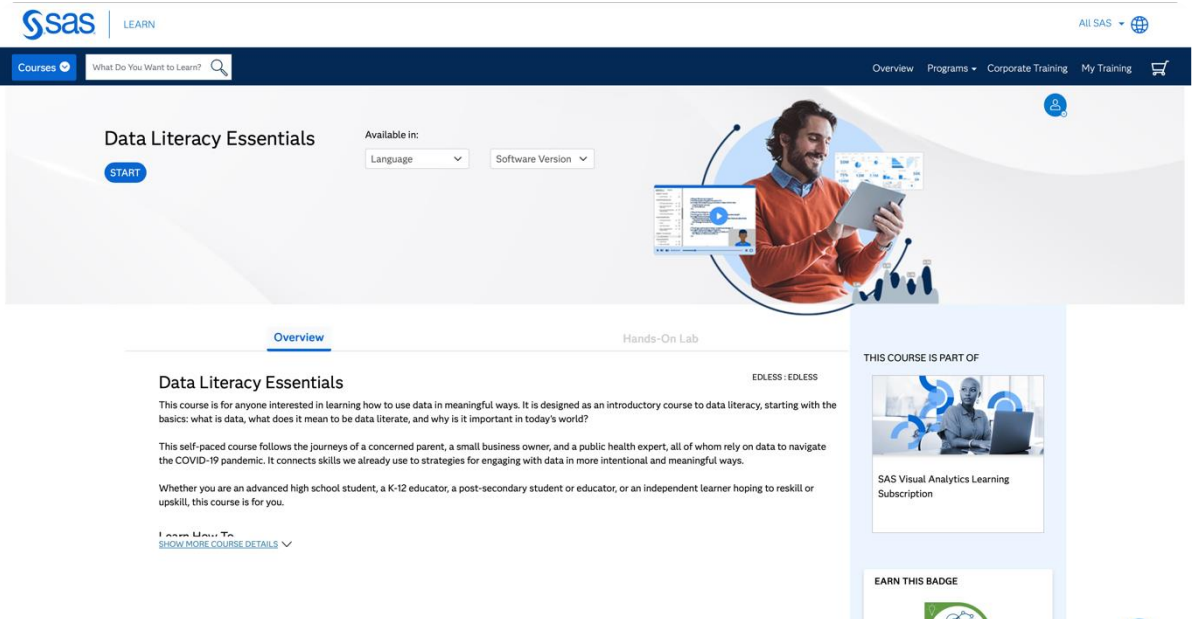
ARTICLE 4: AI LITERACY



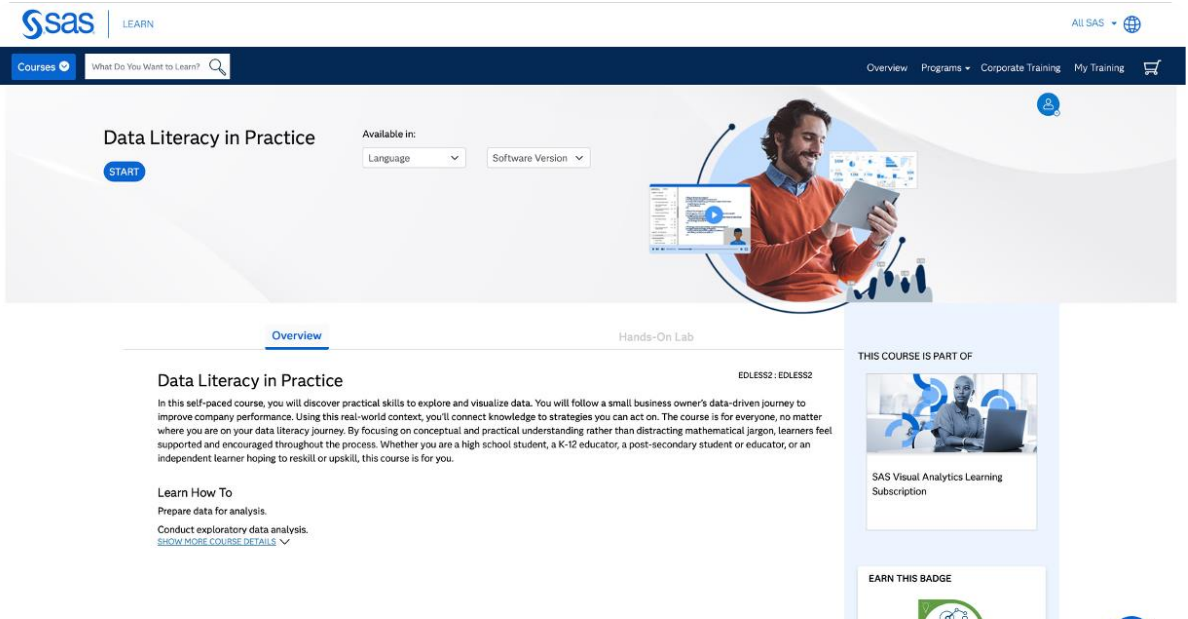
Generative AI Using SAS



Responsible Innovation and Trustworthy AI

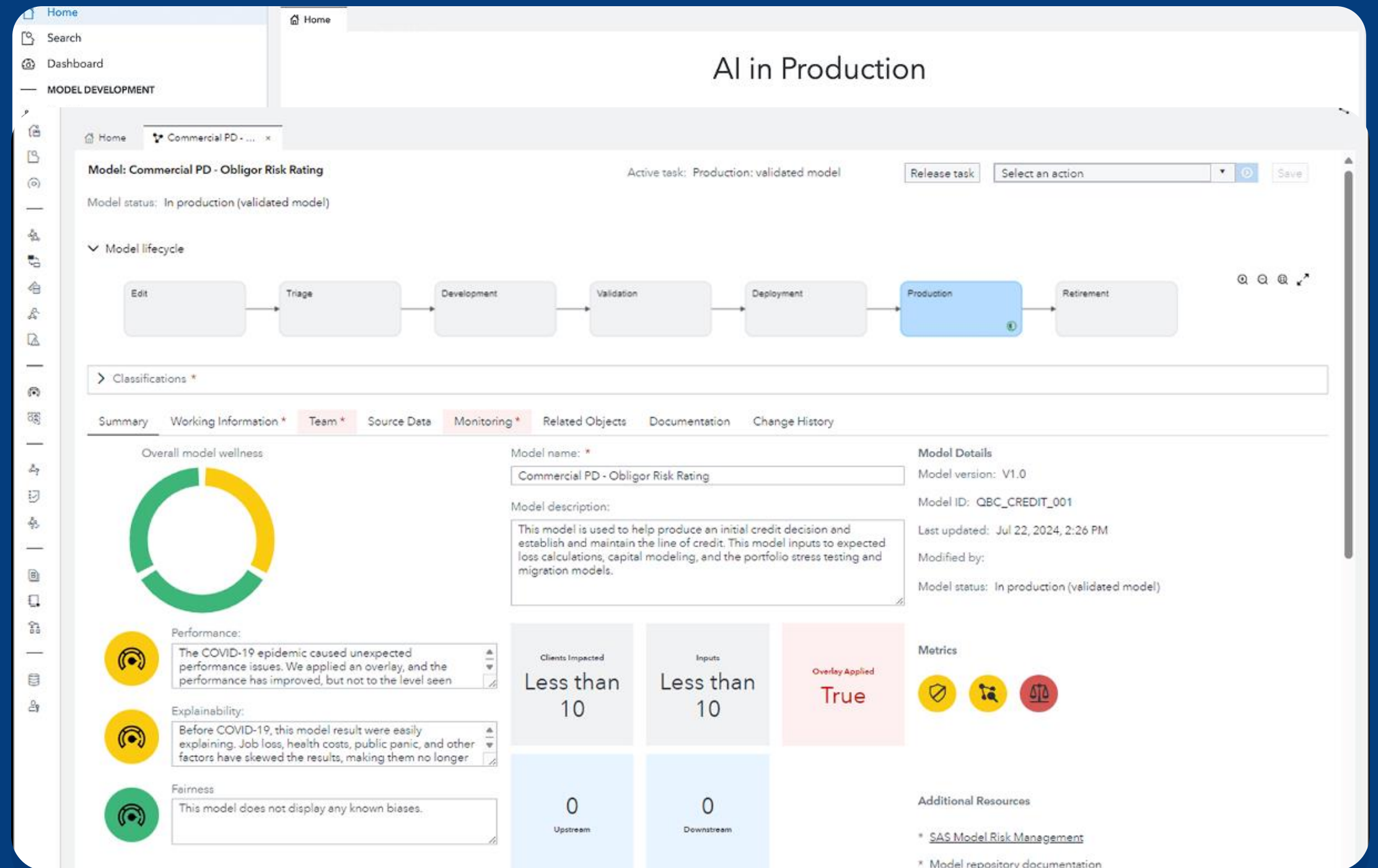


Data Literacy Essentials

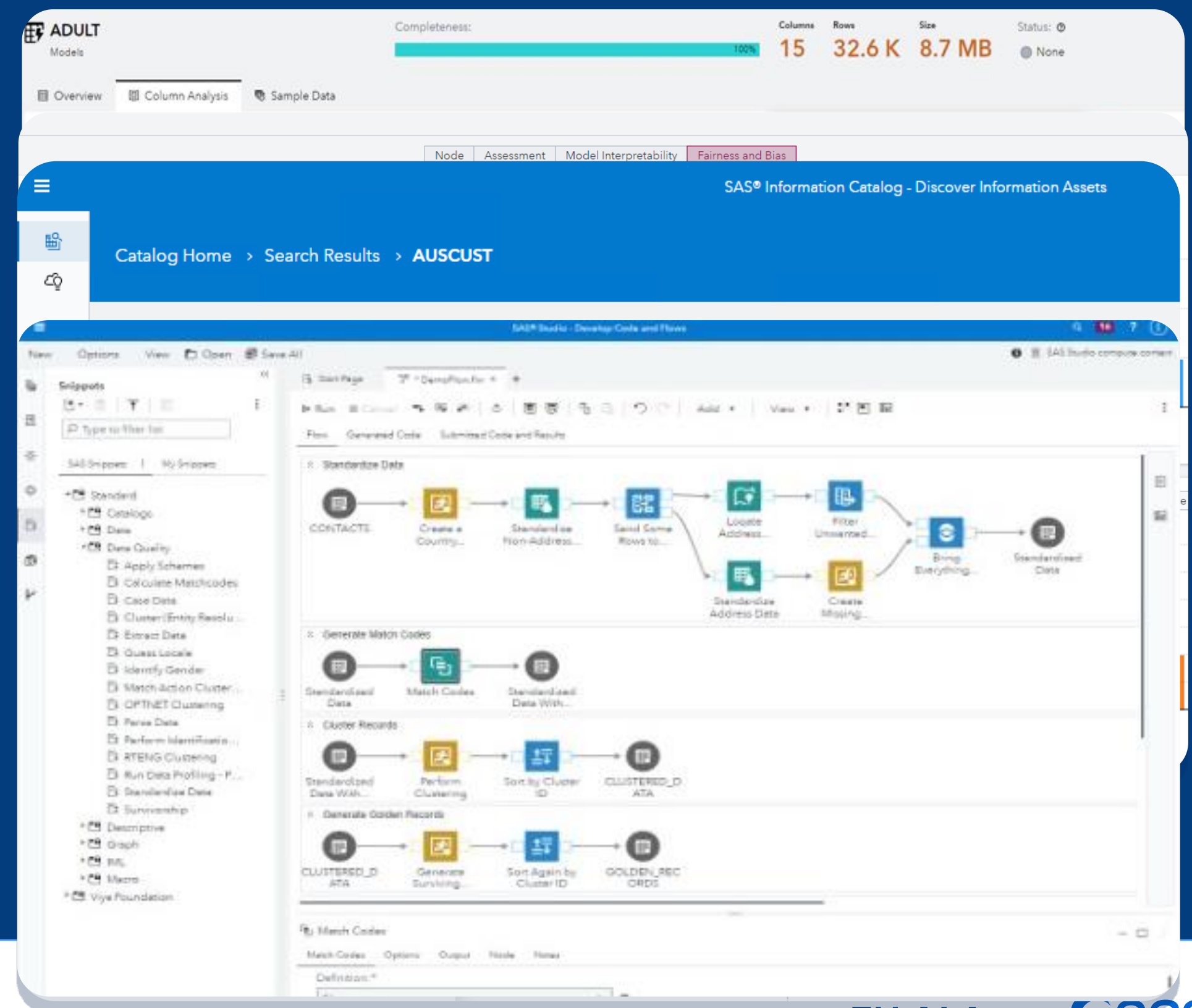


Data Literacy in Practice

- Mitigate model risk with auditable, workflow-controlled processes.
- SAS' Trustworthy AI Workflow reflects the NIST AI Risk Management Framework to minimize risk in model creation.



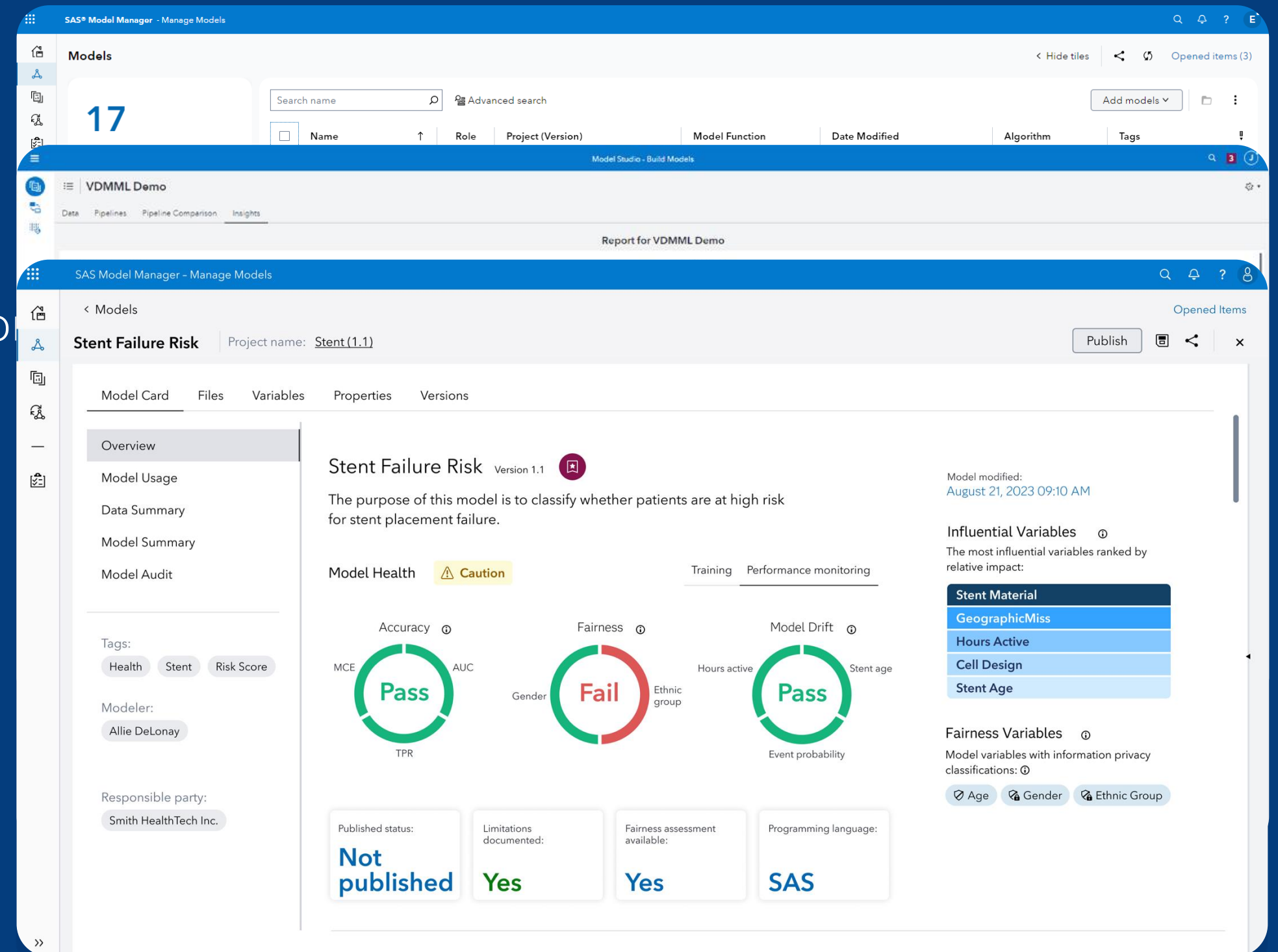
- Build Data Preparation and Data Quality pipelines in a drag and drop or programmatic interface.
- Assess the suitability of data sets and identify relevant data gaps.
- Ensure the security and privacy of data with governed access.
- View information about fairness and bias during data management or take on those actions when building models



ARTICLE 11: TECHNICAL DOCUMENTATION

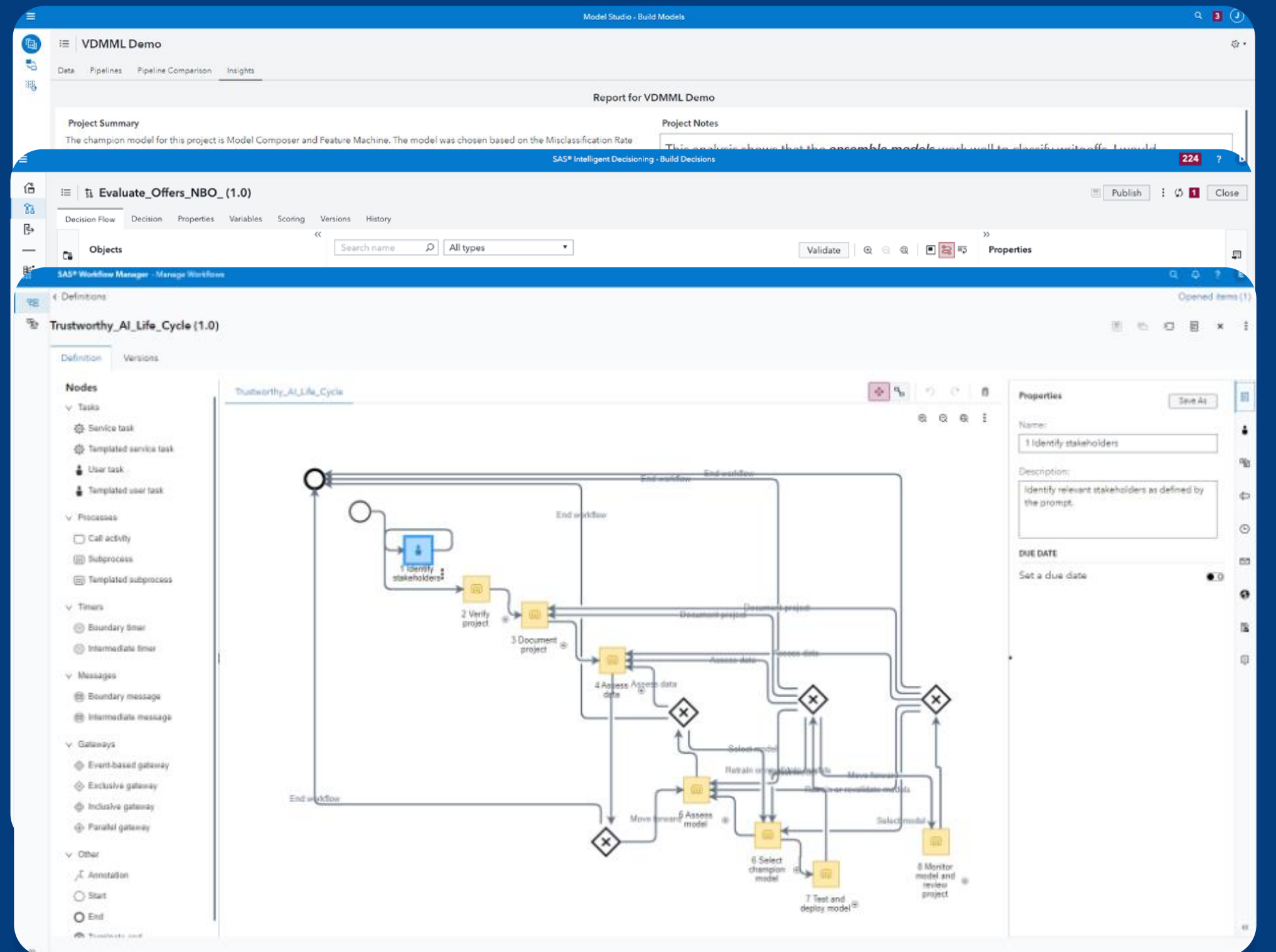
- Monitor performance and govern models.
- View descriptions of datasets used, steps performed for data cleaning and steps used for data preparation.
- Evaluate accuracy measures, explainability and bias assessment in model cards.
- Monitor and control AI system with detailed workflows.

ARTICLE 13: TRANSPARENCY & PROVISION OF INFORMATION TO DEPLOYERS

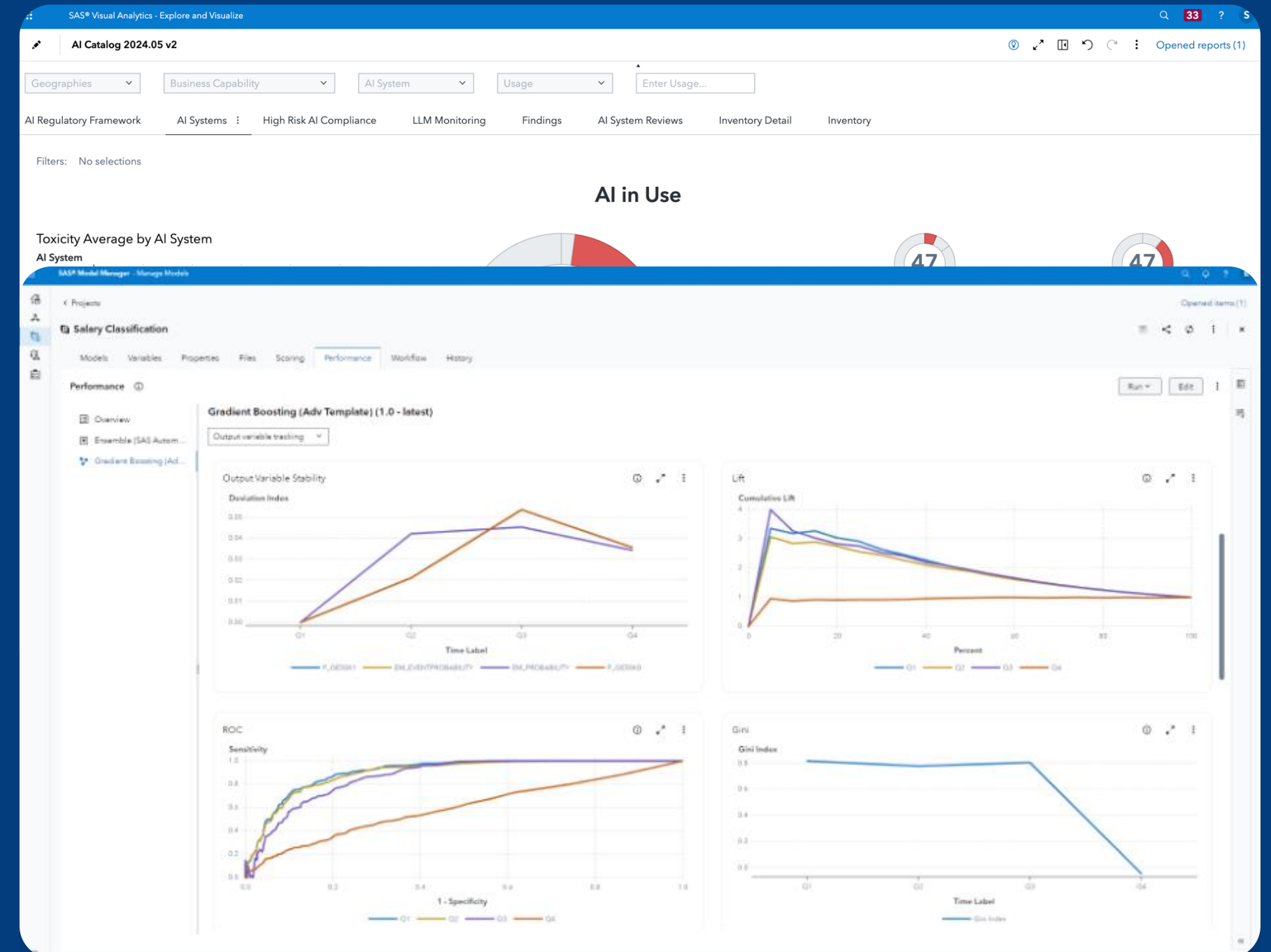


ARTICLE 14: HUMAN OVERSIGHT

- Monitor AI system performance with alerts.
- Easily view Model Interpretability and Fairness Assessment
- Route tasks requiring manual intervention using SAS' Trustworthy AI Workflow.
- Add business rules to existing decision flows for transparency into each decision and multiple approval pathways to implement decisions.



- Nutrition labels for models" provide critical insights like a summary of a model's training data, performance, accuracy, transparency and fairness.
- Deploy, manage and govern models and monitor them for accuracy and relevance over time.



SEPTEMBER 2025

Insights provided by
IDC

DATA AND AI IMPACT REPORT: The Trust Imperative



sas

AI EXPERT PANEL

The Trust Dilemma: Data & AI in 2026

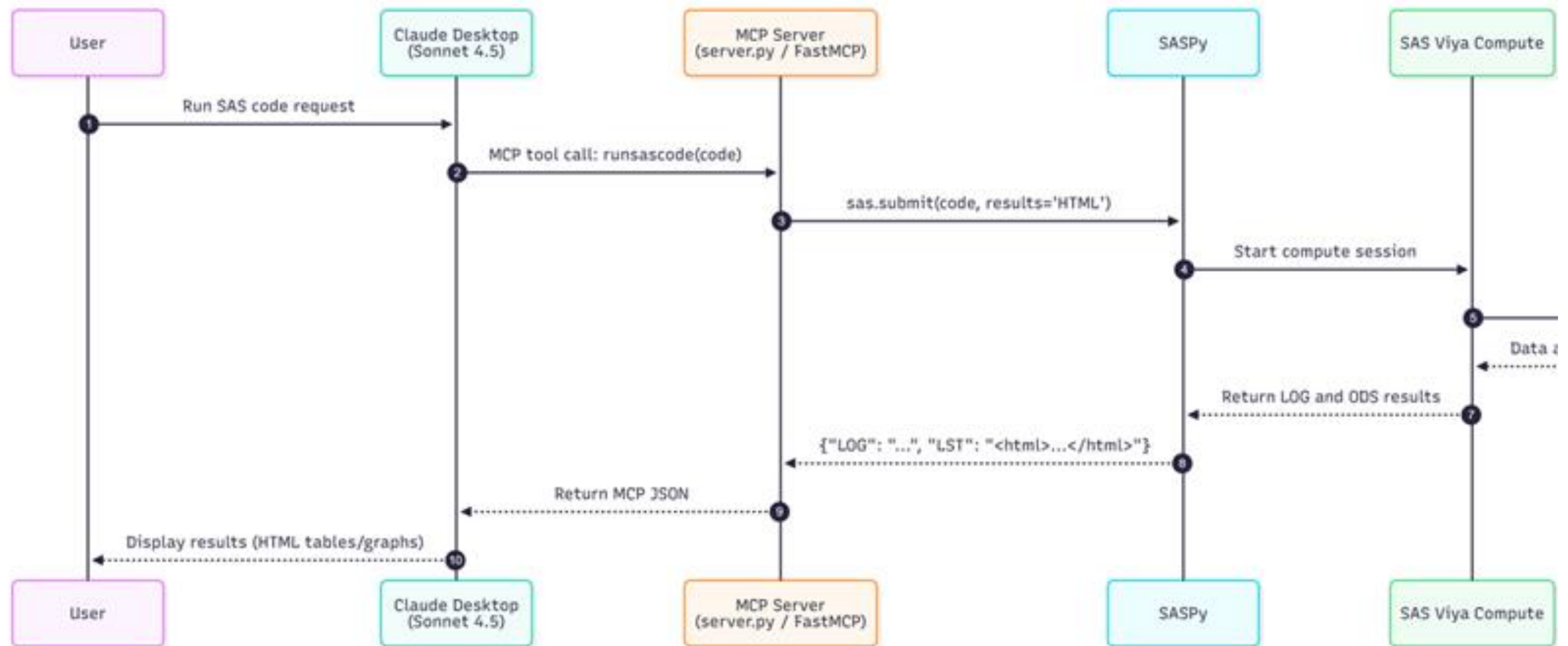
Read the global study



MCP



Agentic Workflow with MCP Server and SAS Viya



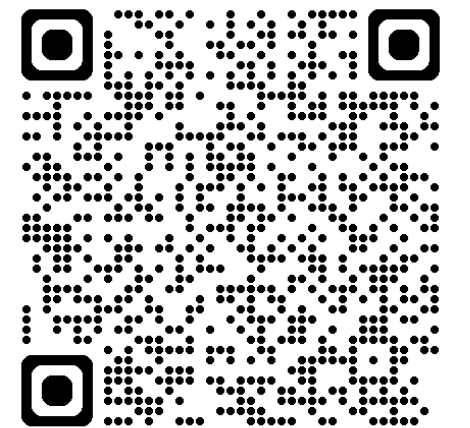
“We have unprecedented access to data, the strongest compute power in history, and cutting-edge analytical methods. So, why are we not doing better by our patients and insist on doing research the same way we’ve done for the past 75 years.”

–Sherrine Eid, MPH

SAVE THE DATE

sas **innovate** 2026

April 27 – 30 | Gaylord Texan | Grapevine, TX



Thank You



Sherrine Eid, Global Head of RWE & Epidemiology, SAS Institute



Samiul Haque, PhD, Senior Data Scientist

