SAS

Trustworthy AI

Taking a trustworthy approach to the AI and analytics life cycle with SAS® Viya®

Robust data governance
for trustworthy decisions
that start with good data

Transparent AI
explains its decisions

Secure ModelOps processes
deploy models into production
with confidence

The Issue

Organizations that embed AI solutions within their business processes also take on the responsibility of mitigating the inherent risks of AI.

AI models are trained on historical data and can perpetuate biases that were hiding in the training data. AI technology often uses sensitive data such as private personal information, health care data or business use cases for creating, training and using models. Unauthorized access to or inappropriate disclosure of these types of data can cause harm to individuals or organizations. Opaque AI systems can foster distrust by masking the decision flows and can increase the challenge of understanding the basis of a decision. Plus, they limit the auditability of a system.

All these issues can lead to AI systems that make unfair or discriminatory decisions, becoming a legal, reputational and ethical liability for the business. The unintended consequences resulting from not addressing and mitigating these risks pose real threats to organizations.

The Challenge

AI models are only as good as their data. Trash in, trash out is extremely accurate when applied in the AI context. If the data is not good, the AI model will not be good – leading to problems such as bias, discrimination and inaccurate predictions.

AI models are opaque and difficult to understand. Lack of transparency in AI makes it difficult for people to understand how decisions are made, the intended scope and outcomes of the AI models, and if there are potential biases in outcomes. Transparency of the model decision-making process fosters trust and understanding.

AI models can be difficult to deploy, govern and monitor. It can be difficult to ensure that AI models are deployed to production in a reliable and secure way, governed in a way that is consistent with organizational policies and procedures, and monitored for performance and accuracy.

Our Approach

Trustworthy AI is critical because it helps build trust with customers. It helps promote transparency and accountability, safety and security, which leads to fair and equitable outcomes.

We approach the problem by providing capabilities to help you with your trustworthy AI journey:

- **Robust data governance.** Support data access, preparation and governance. Private information is automatically tagged, masked, anonymized and suppressed. Automated data exploration uncovers completeness, uniqueness, distinct values, semantic types, correlated variables, mismatched values and outliers. GANs (generative adversarial networks) and SMOTE (synthetic minority oversampling technique) algorithms generate tabular synthetic data.
- **Transparent AI.** Build, optimize and validate models. Understand the outputs of the models with partial dependence (PD) plots, individual conditional expectation (ICE) plots, local interpretable model-agnostic explanation (LIME) and Kernel Shapley values (Kernel SHAP) and explain them using natural language-generated outputs. Assess model fairness and mitigate biases.
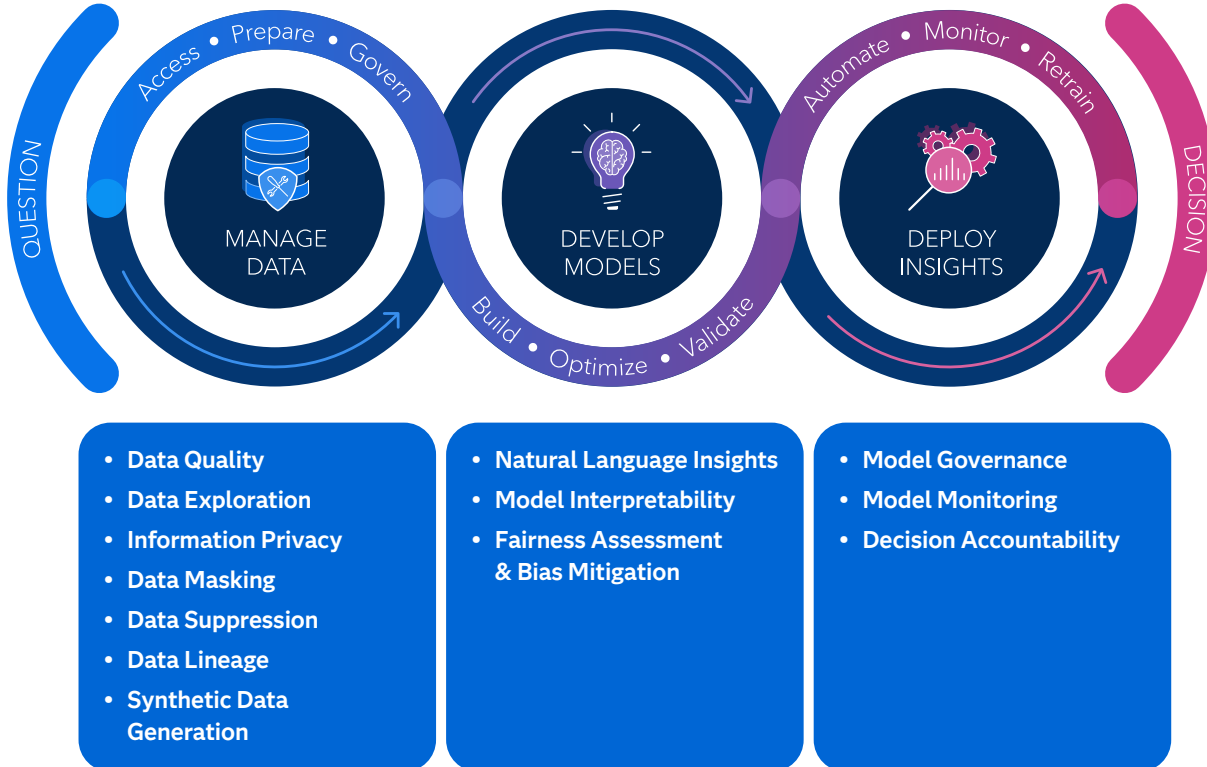
- **Secure ModelOps processes.** Models go through validation and testing to the deployment phase by automating, monitoring and retraining. A centralized, searchable repository for all types of models and analytical assets ensures transparency, traceability and complete visibility into analytical processes.

The SAS® Difference

Our trustworthy AI approach injects fairness and oversight at every step of the AI and analytics journey, so you can tackle complex data and analytics challenges in a responsible manner.

- **Comprehensive, trustworthy AI capabilities.** SAS Viya is an end-to-end data and analytics platform that enables trustworthy practices at every stage of the AI and analytics life cycle – from development to deployment.
- **Principles-based approach.** Our trustworthy AI capabilities are designed with human centricity in mind and in collaboration with diverse stakeholders. It incorporates appropriate levels of accountability, inclusivity and transparency, including features to promote human agency and prevent human harm in a safe and reliable manner.
- **Trusted solution for everyone in the AI and Analytics life cycle.** With intuitive visual interfaces, a centralized platform that enhances and encourages collaboration, AI-driven automation and more, anyone can generate trusted outcomes from their data.

SAS® Viya® trustworthy AI end-to-end capabilities

QUESTION → **MANAGE DATA** (Access • Prepare • Govern) → **DEVELOP MODELS** (Build • Optimize • Validate) → **DEPLOY INSIGHTS** (Automate • Monitor • Retrain) → **DECISION**

- **Data Quality**
- **Data Exploration**
- **Information Privacy**
- **Data Masking**
- **Data Suppression**
- **Data Lineage**
- **Synthetic Data Generation**

- **Natural Language Insights**
- **Model Interpretability**
- **Fairness Assessment & Bias Mitigation**

- **Model Governance**
- **Model Monitoring**
- **Decision Accountability**

Learn more about how SAS is committed to trustworthy AI.

§sas