Machine Learning to Evaluate Governance, Risk, and Compliance Associated with Large Language Models

ITM Central Washington University

Dr. Upakar Bhatta

Outline

- Background of the Study
- Rationale of the Study
- Pre-Requisite Knowledge
- Related Work
- Research Methodology
- ML Implementation Flow
- Data Analysis
- Deploy securely via CI/CD to Azure
- AI Incident Case Study
- Responsible AI
- Embedding AI governance principles directly into the AI lifecycle
- Govern AI Deployment and Use
- Governance Framework

Background of the Study

Rapid adoption of AI and LLMs is transforming business operations and customer engagement.

Organizations rely on LLMs for automation, efficiency, and enhanced service delivery.

Increased use of LLMs introduces new Governance, Risk, and Compliance (GRC) challenges

. Risks include bias, toxicity, data sensitivity violations, and regulatory non-compliance

This research offers a potential solution to predicts GRC risk levels in LLM interactions by integrating cloud services, machine learning, and data analytics services. Study uses Azure OpenAI logs to build a dataset with operational and behavioral features

Rationale of the Study

Several studies have previously explored transparency for LLM, yet a comparative evaluation of compliance risks associated with LLMs remains limited. Previous research doesn’t offer a comparative machine learning framework that evaluates regulatory risks considerations. Furthermore, existing research lacks the analysis of governance models relevant to LLM such as NIST AI Risk Management Framework (2023), which offers structured approaches to AI governance and risk management.

Therefore, further research is needed to demonstrate how machine learning model can be deployed to assess the compliance risks associated with LLM.

Pre-Requisite Knowledge

Cloud computing (NIST 800 145):

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cloud service types: IaaS, PaaS, SaaS

Cloud security

Pre-Requisite Knowledge contd..

Artificial Intelligence (AI)

Machine Learning (ML)

Deep Learning (DL)

Generative AI

Agentic AI

LLM

Related Work

Several studies have previously explored transparency for LLM, yet a comparative evaluation of compliance risks associated with LLMs remains limited.

Zhang, Q., Cheng, L., & Boutaba, R. (2010)

ML communities are primarily aimed at supporting a mechanistic understanding of how the model or system functions by disclosing its components and processes.

Russom, P. (2011)

The author outlined the importance of taking a human-centered perspective on transparency.

Karras, O., & Schneider, K. (2019)

Explanations of machine learning and AI outputs have been proposed as a means to mitigate transparency-related challenges

Zhong, C., & Goel, S. (2024)

The author emphasized the explanations of AI systems which had been identified as contributing to greater system transparency

Research Method

The study demonstrates how various machine learning techniques can be leveraged to evaluate security risks in alignment with compliance requirement.

A modular ML pipeline was designed with the components such as feature extraction, model architecture, and assessing accuracy of machine learning for GRC (Governance, Risks, and Compliance evaluation).

Data Collection: Security log from Azure services were used to construct a sample dataset consisting of 5000 records. Feature selection: A total of nine features were included: response_time_ms, model_type, temperature, tokens_used, logged, data_sensitivity, compliance_flag, bias_score, and toxicity_score.

Data preprocessing: Data cleaning, encoding, and normalization were performed.

Exploratory Data Analysis (EDA): Data visualization techniques such as Correlation heatmap was utilized to detect outliers, observe feature distributions, and identify relations between features.

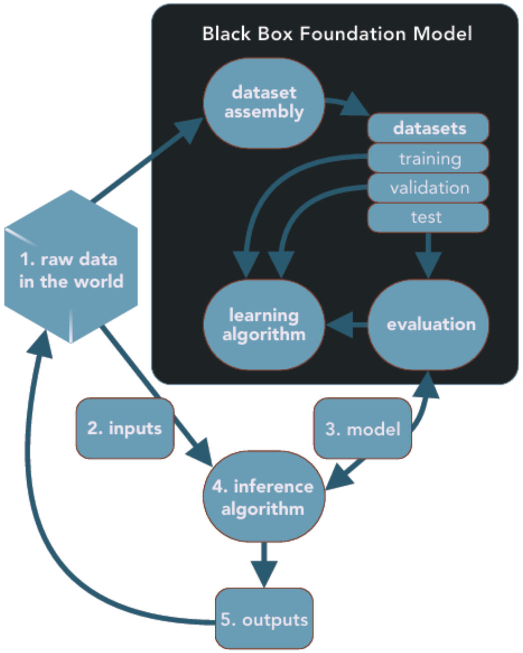
Split the dataset: The dataset was divided into training (80%) and testing (20%) sets.

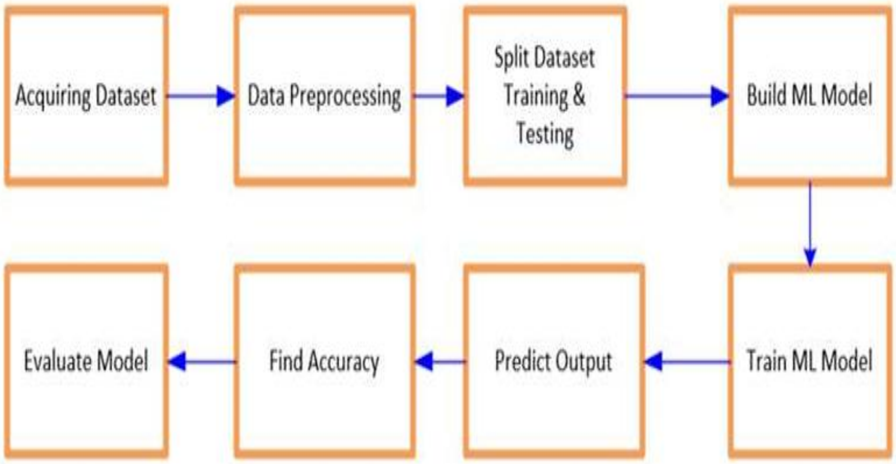
Synthetic Minority Over-sampling Technique (SMOTE): SMOTE was applied to address the class imbalance in the dataset. Model selection:

Mathematical algorithms were selected to train the machine learning models. ompliance and regulatory requirements

ML Implementation Flow

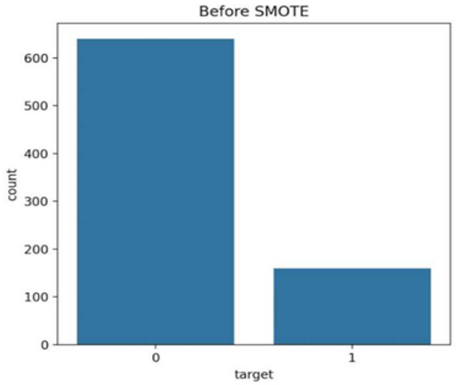
LLM process implementation workflow

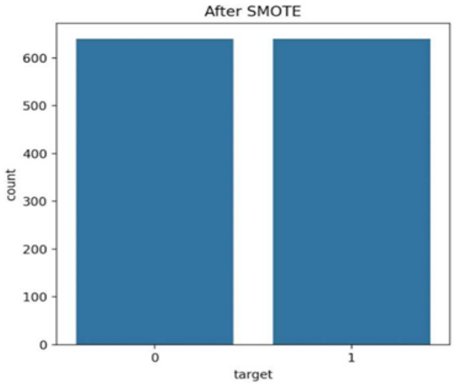


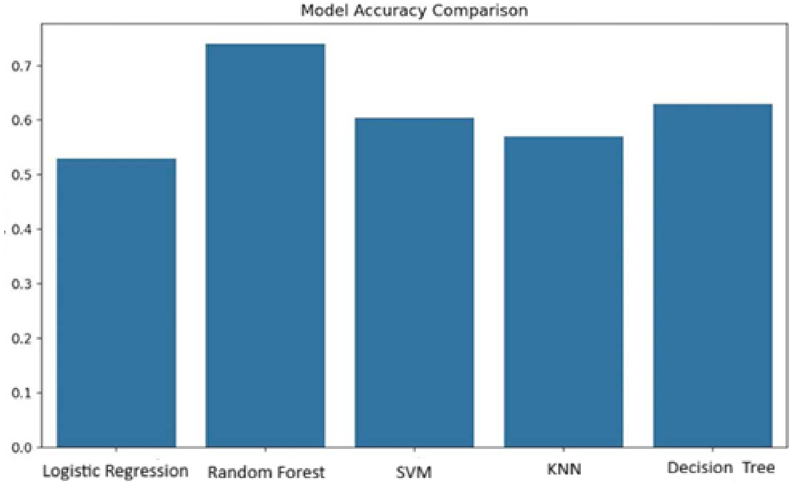


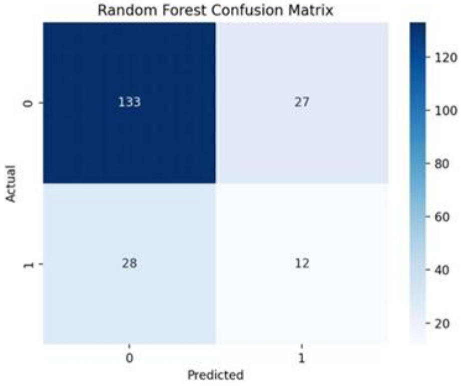
ML Data Analysis and Result

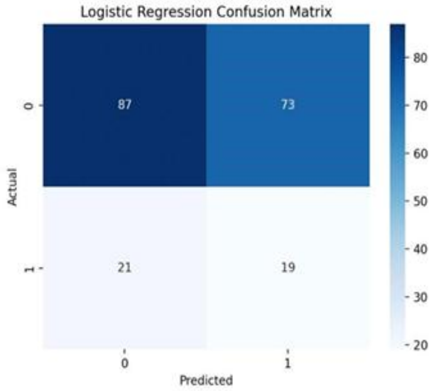
Model Accuracy











Deploy securely via CI/CD to Azure

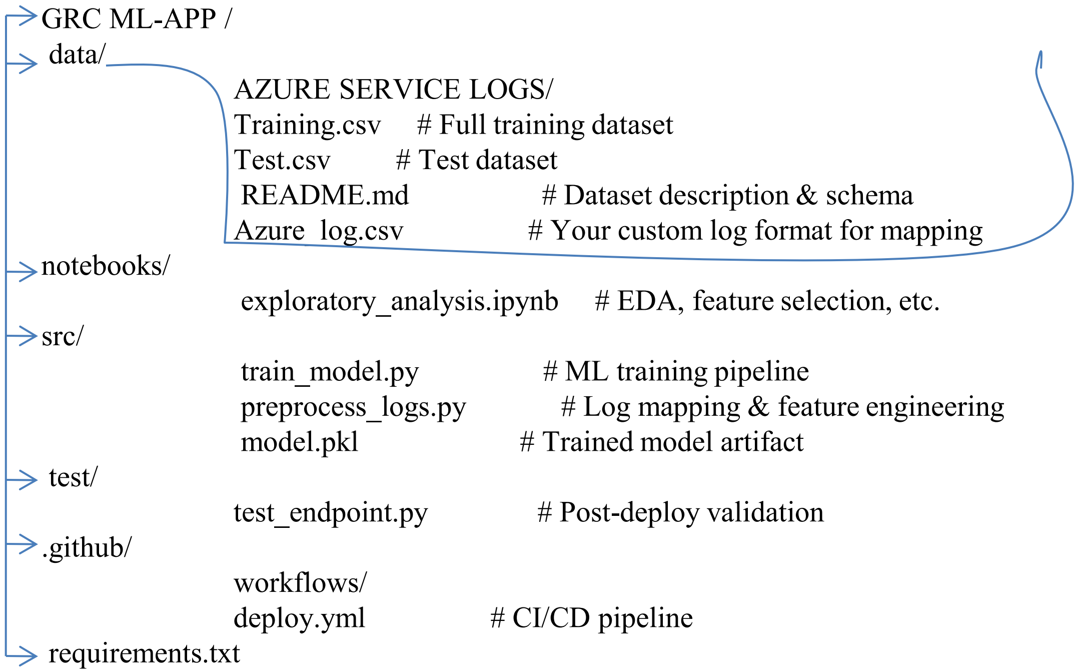
You push your code to GitHub, and your CI/CD pipeline (GitHub Actions or Azure DevOps) automatically builds and deploys the application to Azure App Service.

- You push code to GitHub
- Azure DevOps Pipeline is triggered
- Azure DevOps runs CI (build + test)
- Azure DevOps runs CD (deploy to Azure App Service)

GitHub only stores the code; Azure DevOps does the CI/CD. CI builds it, CD deploys it
CD is the vehicle that delivers the built application to Azure App Service.

Repository Structure

GitHub: GRC ML-APP/



```
GRC ML-APP /
 data/
                 AZURE SERVICE LOGS/
                 Training.csv     # Full training dataset
                 Test.csv         # Test dataset
                 README.md                  # Dataset description & schema
                 Azure_log.csv              # Your custom log format for mapping
notebooks/
                 exploratory_analysis.ipynb     # EDA, feature selection, etc.
src/
                 train_model.py             # ML training pipeline
                 preprocess_logs.py           # Log mapping & feature engineering
                 model.pkl                  # Trained model artifact
 test/
                 test_endpoint.py           # Post-deploy validation
.github/
                 workflows/
                 deploy.yml             # CI/CD pipeline
 requirements.txt
```

Case Studies

OECD's artificial intelligence newsletter!

Generative AI Drives Surge in Sophisticated Email Scams in Catalonia

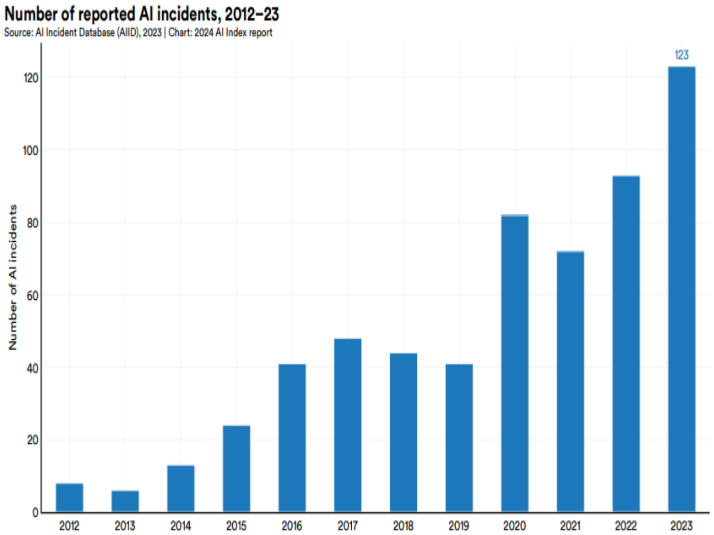
2026-02-01

According to OECD newsletter, 82.6% of email phishing scams now use generative AI to create highly realistic fake texts, voices, and videos, making them harder to detect and increasing cybercrime harms such as fraud, data theft, and ransomware attacks in Catalonia.

AI principles: Robustness & digital security Safety

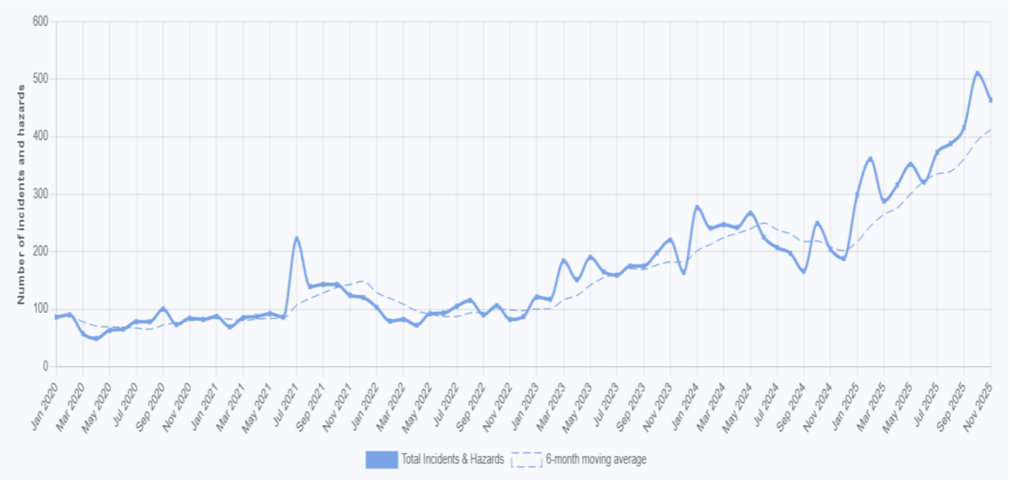
Severity: AI incident

AI system task: Content generation



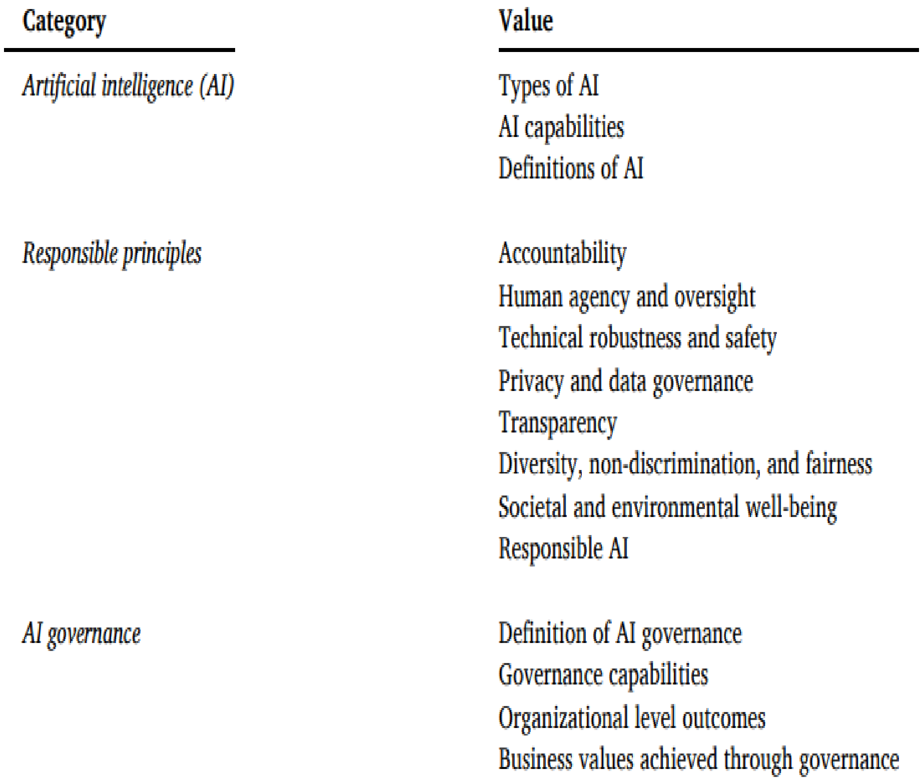
Number of reported AI incidents, 2012–23

Source: AI Incident Database (AIID), 2023 | Chart: 2024 AI Index report

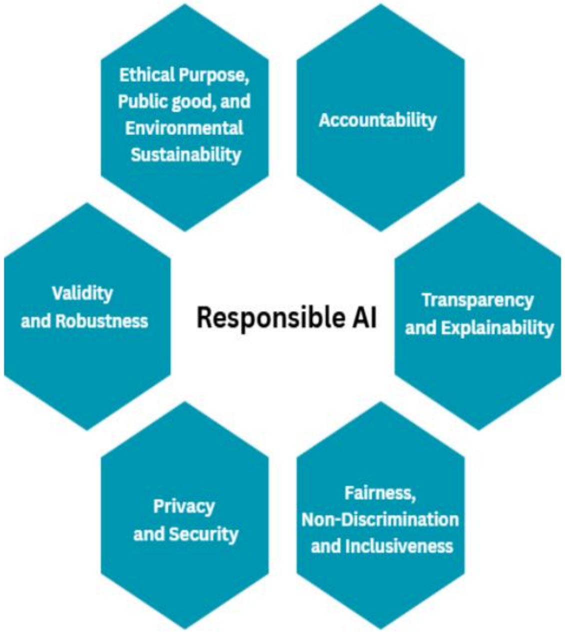


Responsible AI

Liu et al., (2022)



Category	Value
Artificial intelligence (AI)	Types of AI
	AI capabilities
	Definitions of AI
Responsible principles	Accountability
	Human agency and oversight
	Technical robustness and safety
	Privacy and data governance
	Transparency
	Diversity, non-discrimination, and fairness
	Societal and environmental well-being
	Responsible AI
AI governance	Definition of AI governance
	Governance capabilities
	Organizational level outcomes
	Business values achieved through governance

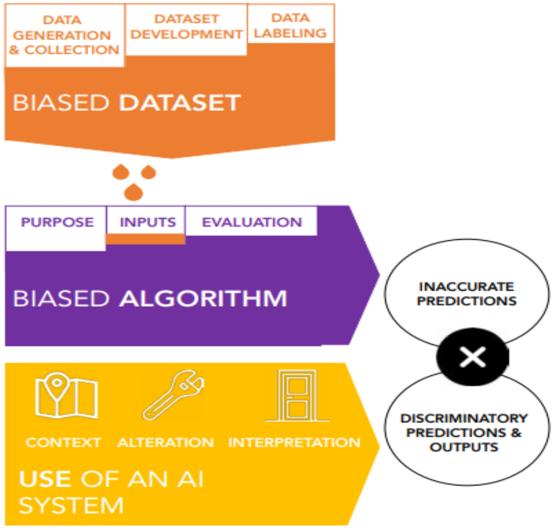

Ethical Purpose, Public good, and Environmental Sustainability
Accountability
Validity and Robustness
Responsible AI
Transparency and Explainability
Privacy and Security
Fairness, Non-Discrimination and Inclusiveness

Embedding AI governance principles directly into the AI lifecycle

Data Collection

EDA (Exploratory Data Analysis) (threat modeling)→Govern AI development

– understand bias, imbalance, quality issues



Data Cleaning/Preprocessing

– handle missing values, normalization, encoding, privacy/fairness compliance

Feature Extraction/Engineering

– create/select features from *cleaned* data

Model Training & Evaluation (risk assessment)→ Govern AI development

– train, test, validate, manage bias & risk

Pre-deployment (impact assessment)→Govern AI deployment

Deployment & Monitoring (threat intelligence)

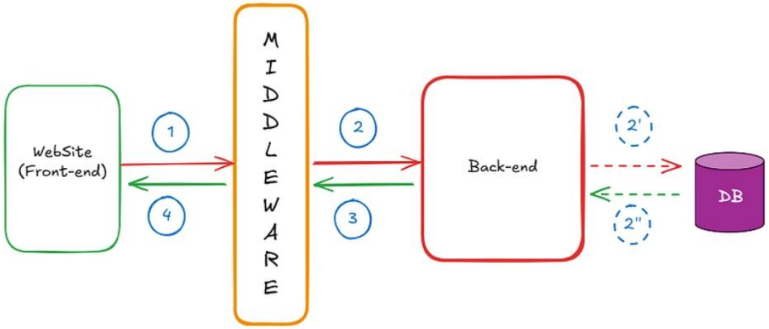
Govern AI Deployment and Use

Responsible AI Deployment

Model training and evaluations monitoring misuse, safety oversight, transparency in deployment

Frontend (Web/App) → Middleware (Flask/FastAPI/API Gateway) → ML Model Backend → Response

[User Uploads Logs] → [Frontend Form]
↓
[Flask Backend Receives File] → [ML Module Processes Lo
↓
[Model Predicts Normal/Attack] → [Recommendations Ge
↓
[Results Returned to Frontend] → [User Sees Output]



Generative AI Chat Application or Vulnerability Mgmt Tool

Amazon Bedrock is a fully managed service

Middleware EC2 instance

Frontend S3

Users (Frontend S3) → Orgaization AI Gateway → Middleware → AI Providers
↳ Copilot
↳ ChatGPT
↳ Grok
↳ Azure OpenAI

Governance Framework

Regulation Standard Policies Procedure

Framework → Policy → Procedure

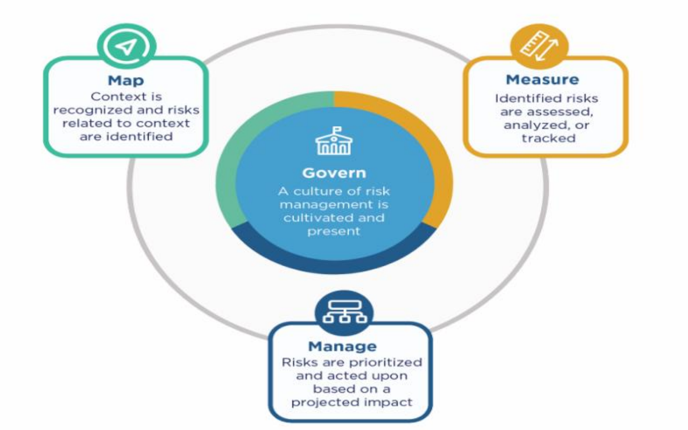
Key aspects when creating AI policies

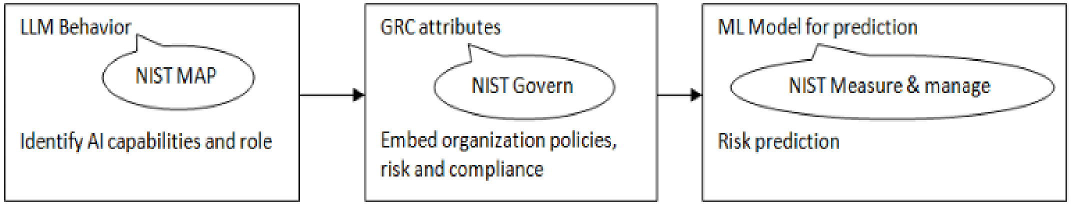
Trustworthiness Characteristics = ethical use which includes:

valid and reliable, safe, secure and resilient, accountable and trans parent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed)

Map each principle to NIST AI RMF or ISO/IEC 42001

The following figure shows the four pillars of NIST AI Risk Management Framework include Govern, Map, Measure, and Manage


Map
Context is recognized and risks related to context are identified
Measure
Identified risks are assessed, analyzed, or tracked
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact



Thank You!