Voluntary Reporting for Enterprises

Triveni Gandhi, Ph.D
Responsible AI Lead, Dataiku

Voluntary Reporting - On what? For Whom?

Proprietary General Purpose AI Models (GPAI)

Publicly available or web accessible GenAI

3rd Party Model Providers

AI Systems

Public Entities, Gov't Agencies

Informal AI usage

Everyone

AI Use Cases

Public Entities, Gov't Agencies, Large Enterprises, SMBs, Non-Profits…

Typical Generative AI Use in the Enterprise

Orgs determine AI use cases that would benefit from Generative AI

3rd Party Model Vendors provide pre-trained general purpose models (GPAI)

Orgs build out pipelines to supplement GPAI with RAG, fine-tuning, additional non-GenAI models and data

Pipelines are tested, evaluated, deployed and monitored across the organization

Governing Generative AI Use in the Enterprise

Orgs determine AI use cases that would benefit from Generative AI

3rd Party Model Vendors provide pre-trained general purpose models (GPAI)

Orgs build out pipelines to supplement GPAI with RAG, fine-tuning, additional non-GenAI models and data

Pipelines are tested, evaluated, deployed and monitored across the organization

Pre-Assessments, Risk Profiling, and Scoping

Model Cards and assessments, Other Types of Reporting Frameworks

Map, Measure, Monitor use cases in accordance with TEVV principles from the AI RMF

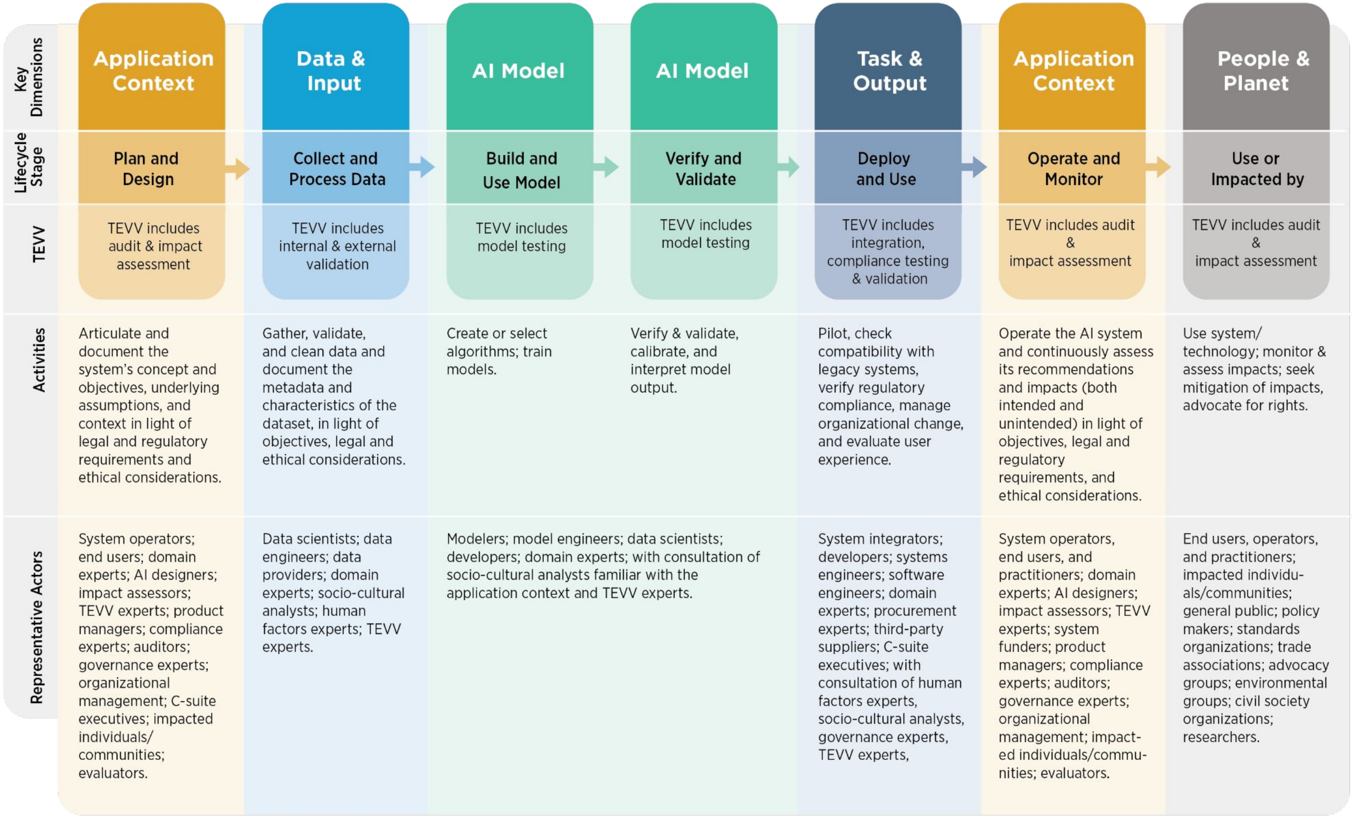
Buy-in to a VRT

How do you convince a kid to eat their broccoli/brussel sprouts/spinach?

Make it delicious by adding butter,cheese, prime energy

How do you convince an organization to use an voluntary reporting template?

Show them how it can increase the value gained from their use case, aid in change management, **and** ensure the AI system is safe and reliable for users



Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design →	Collect and Process Data →	Build and Use Model →	Verify and Validate →	Deploy and Use →	Operate and Monitor →	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impacted individuals/communities; evaluators.	End users, operators, and practitioners; impacted individuals/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

NIST AI RMF AI Lifecycle

Application Context

Will the output of this use case be shared as an automation or AI Assistant?

Who are the expected users of this project, and who might be impacted as a result of this project's use(directly or indirectly)?

What kind of data will end users have access to through the user interface?

Data & Input

What kind of data (sensitive, personal, business related) will be used to develop this project?

Who will validate the appropriateness and potential for bias in input datasets?

Are security measures required to protect the privacy of individuals or corporate information with datasets? What mechanisms to protect data are in place?

AI Model

Is the selected GPAI suitable for the task at hand? Does the usage of this model for this specific use case fit the acceptable use policy?

What evaluations will be run to ensure the model is producing accurate and unbiased results?

Will the GPAI interact with other models in the same project? Have those models been vetted for accuracy, bias, and reliability?

Task & Output

Will prompts and outputs be checked for toxicity, PII or manipulation? What mitigations for problematic content is in place?

What kind of feedback or remediation is possible for end users that do or do not like the outputs provided?

How often will feedback be reviewed and used to improve the project outputs?

Pre-Build Assessment, Scoping, Risk Profiling

Measure, Monitor, Manage the AI Lifecycle