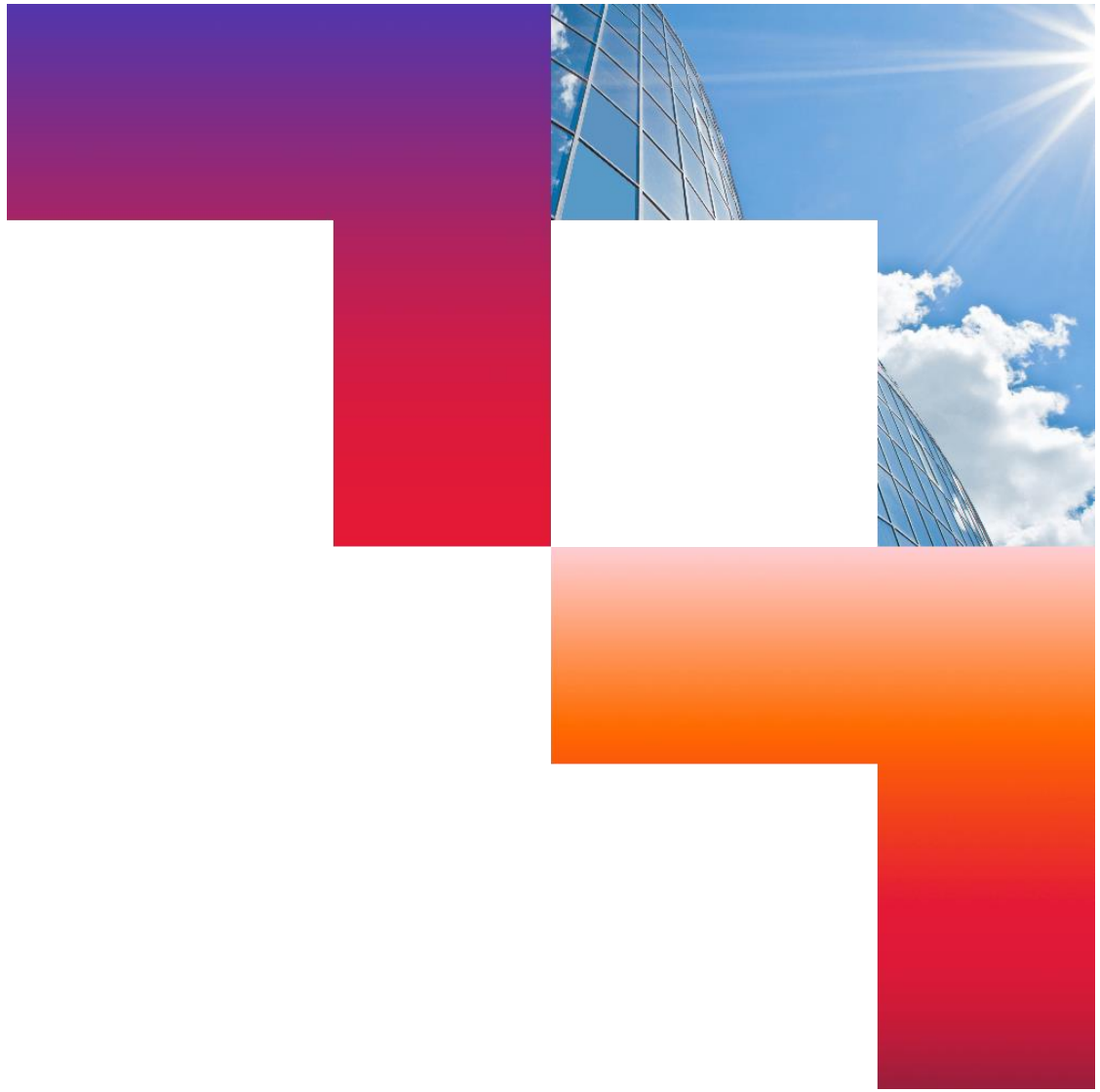
Executive Insights Exchange

Framework for AI Ethics



June 2021

CGI

Introductions

Dr. Diane Gutiw
Vice President, Data Science, CGI

Dr. Parimal Kulkarni
Sr. Consultant, Data Science, CGI

Why AI Ethics

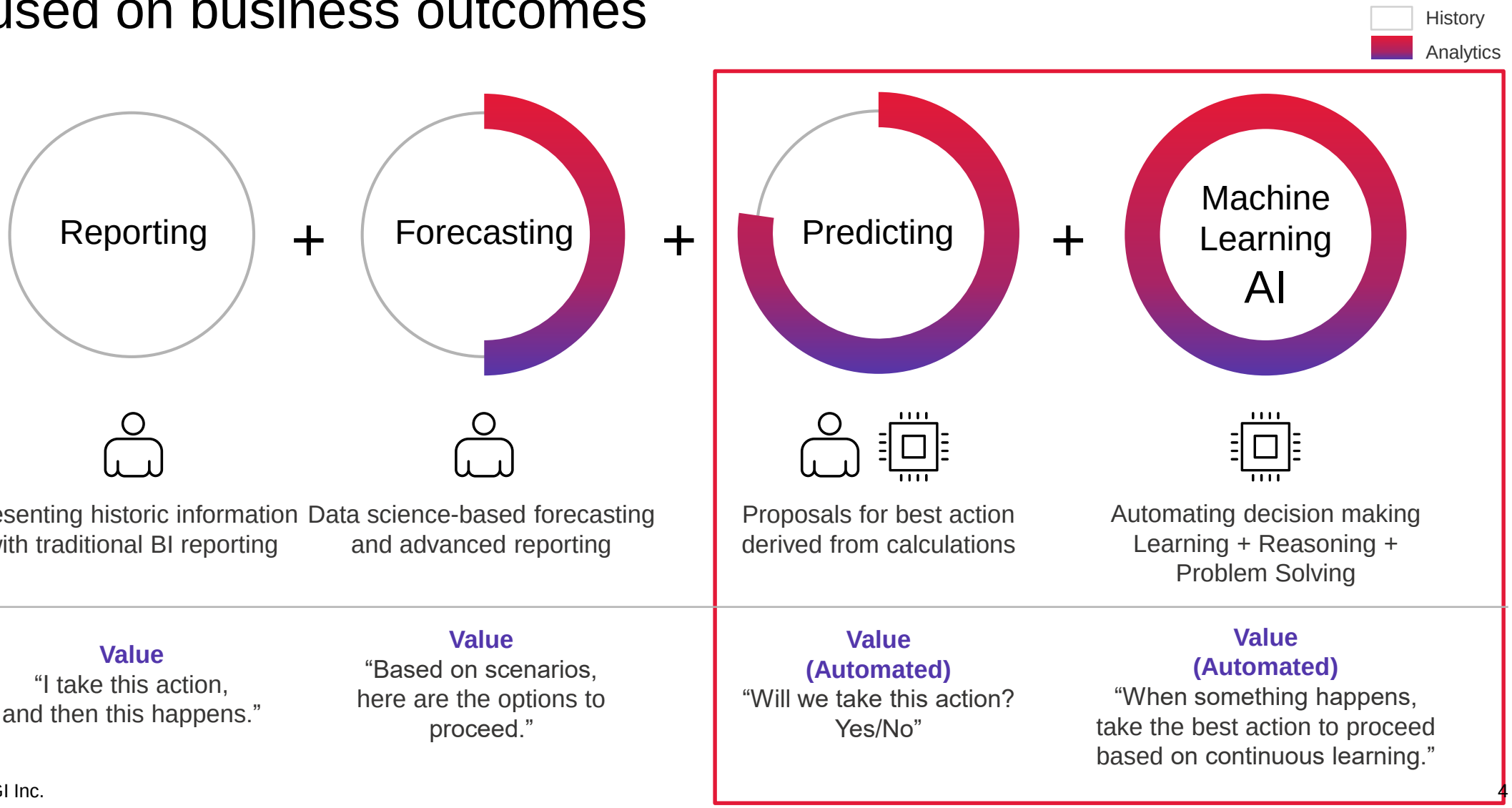
Privacy is a top concern

Rapidly evolving AI developments with low transparency

Behavioral Manipulations

Bias Control

Organizations are adopting more advanced analytics that are focused on business outcomes



Typical AI/ML Maturity Journey

Assessing where we are and envisioning where we want to be

Manual

- Simple rule based
- Scripts/Macros
- Ad-Hoc Implementation

Basic

- Descriptive Analytics
- Basic supervised learning
- RPA
- Project based implementation

Conditional

- Image recognition,Optical Character Recognition (OCR)
- Intelligent document processing
- Basic sstructured and unstructured data analytics
- Simple web chatbot integration (e.g. FAQ)
- Process based implementation

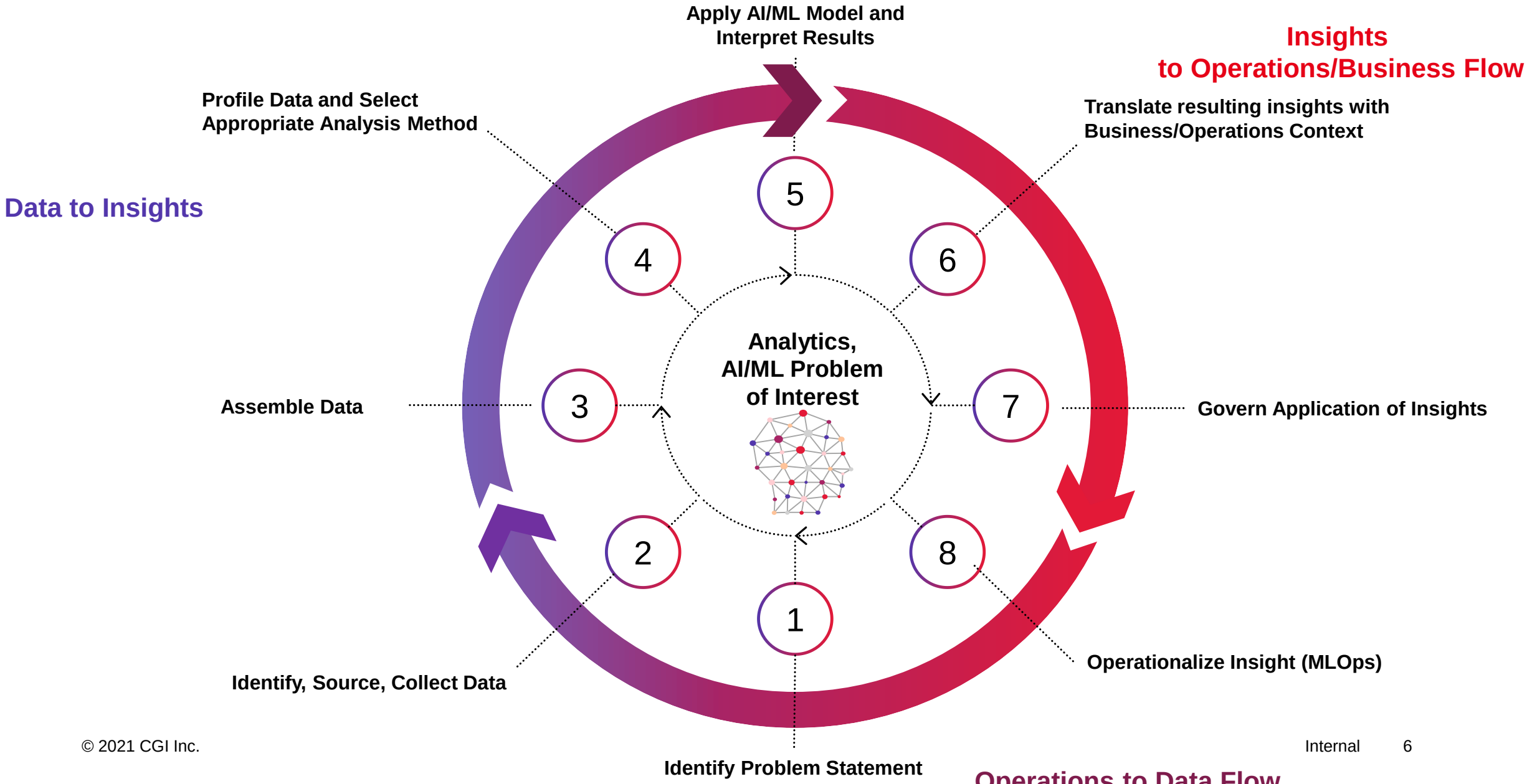
Algorithmic

- Prescriptive analytics
- Complex Supervised Learning-Machine learning (classification/clustering)
- Basic Unsupervised Learning
- Unstructured & Big Data
- Natural Language Processing (NLP), Fully integrated chatbots, Reinforcement Learning
- Domain based implementation

Artificial Intelligence

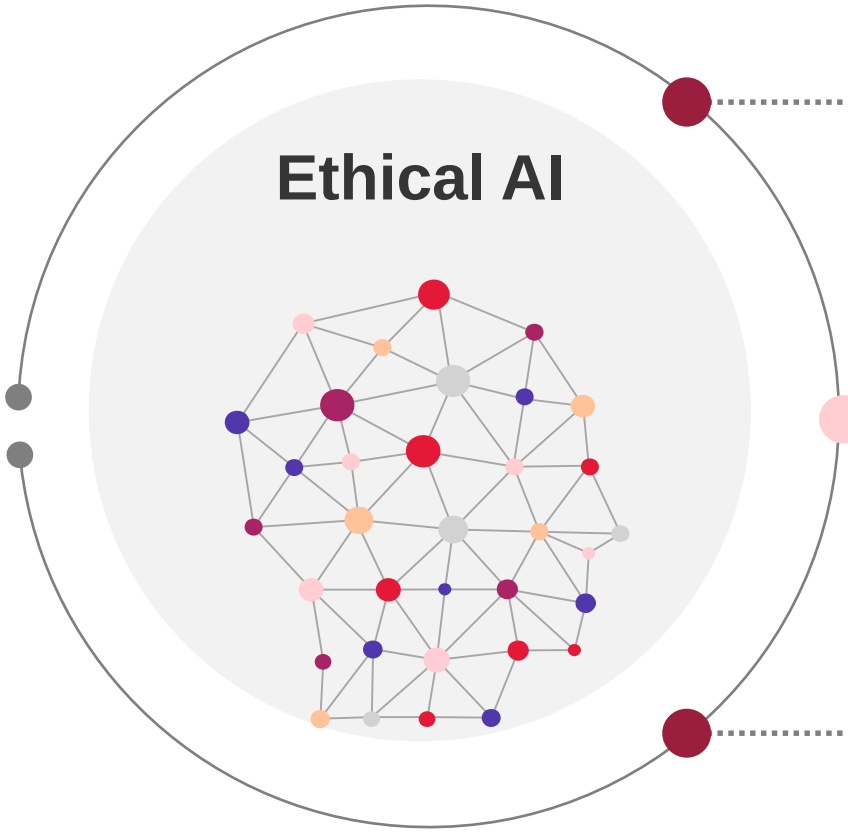
- Cognitive technology
- Full end-to-end autonomy, hypothesizing, reasoning
- Unsupervised Learning, Deep learning
- Augmented & virtual reality
- Fully capable virtual agents, omni-channel
- Enterprise-level implementation

AI Adoption and Implementation Strategy



Ethical AI

Artificial Intelligence (AI) should be designed to uphold ethics, have a code of conduct and focuses on human centered design related to clinical and clinical operation decision making



Principles

Guiding Principles i.e. the Code of Conduct for Artificial Intelligence that help ensure it is responsible for the inputs, processes, and outcomes delivered

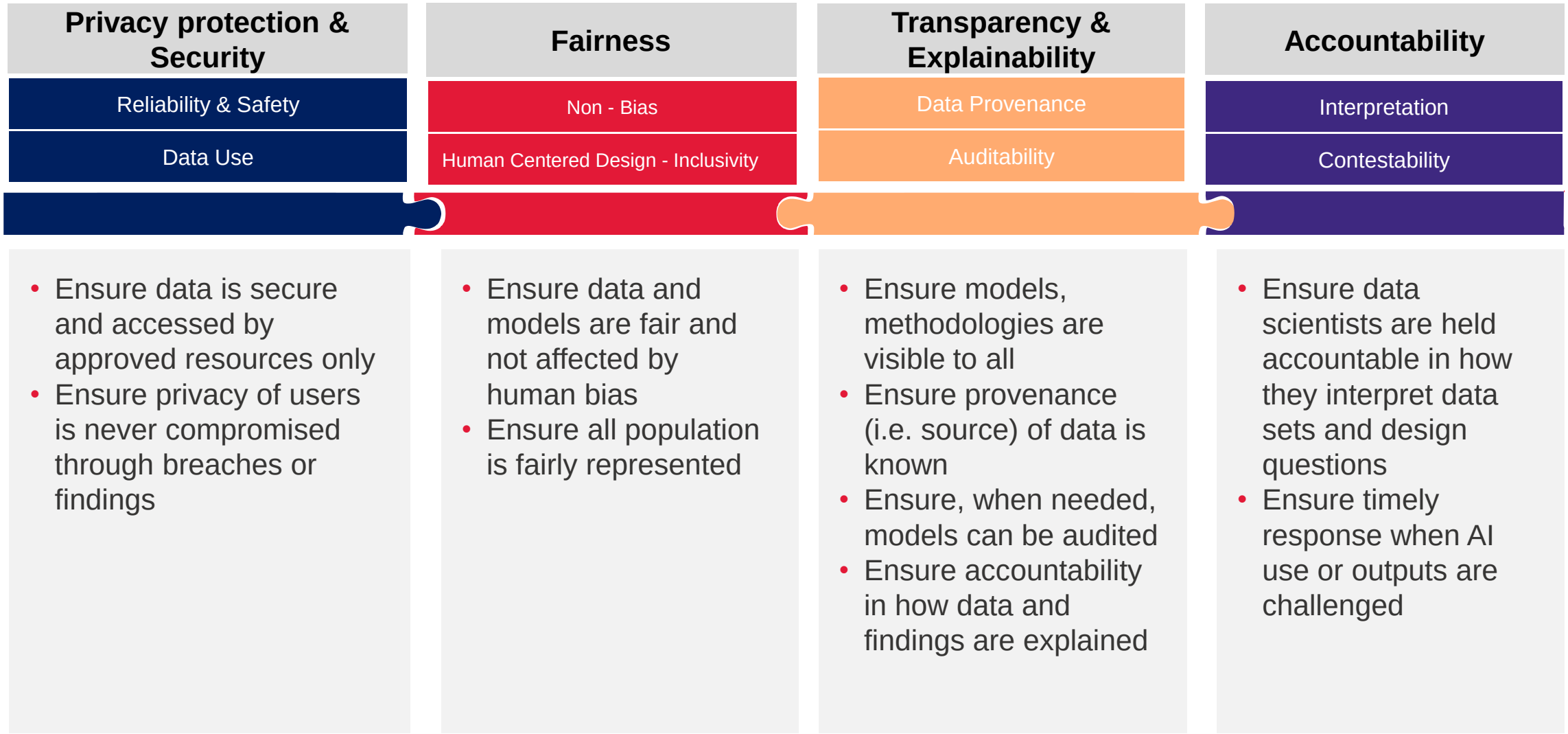
Risks

Traditional methods of risk analysis can no longer handle the ever increasing volume of data, especially in volatile regulatory environments.

Mitigation Techniques

Tools (Methodological, Technical, Strategic) that ensure Principles of Ethical AI are upheld by mitigating risks identified

Principles



Ethical AI Framework

Ensure Fairness, Accountability, Transparency

- Data profiling is done using rigorous testing
- Use packages that are pre-built for Ethical AI

Ensure Privacy Protection & Security

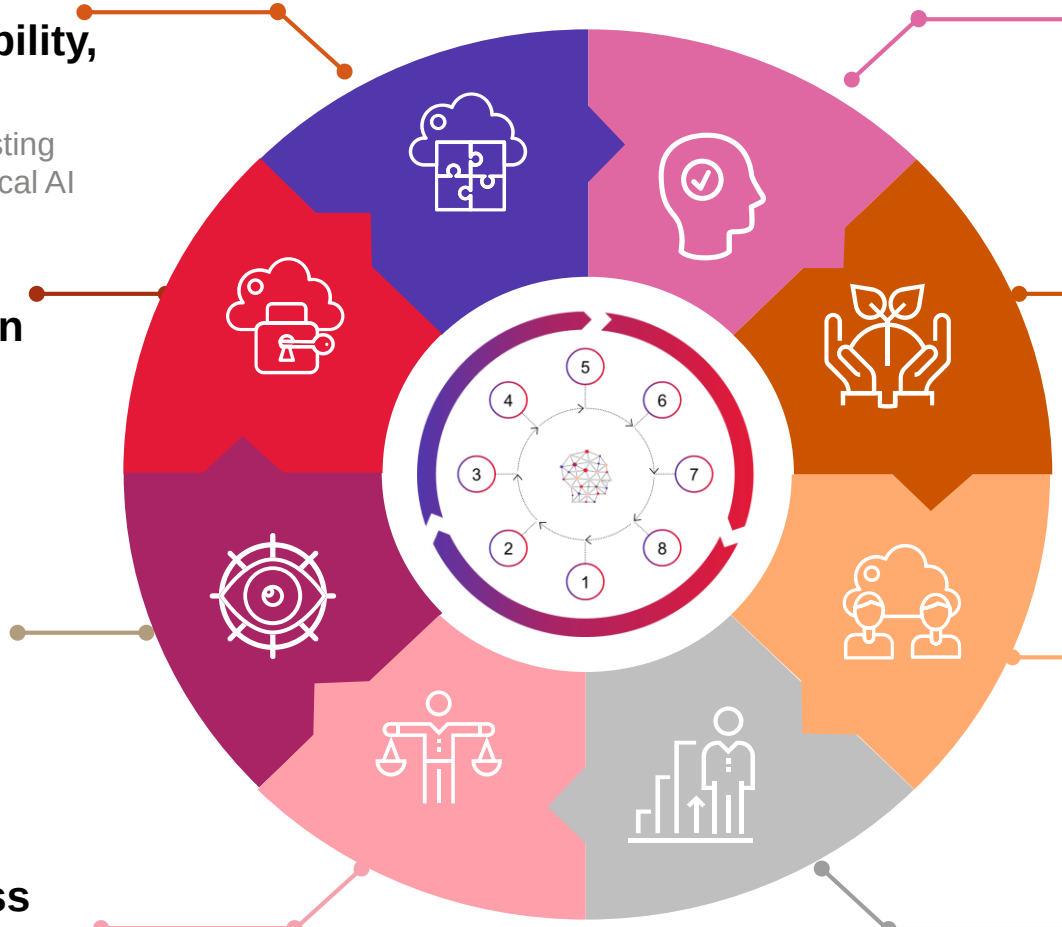
- Monitor Access to & Use of Data
- De-identify data where needed

Ensure Transparency

- Data Provenance is known
- Data collection methodology & assumptions are known

Ensure Fairness

Fairness (Non-Bias) in Problem Statement, data collection (inclusivity)



Ensure Accountability (Interpretation, Explainability)

Provide guardrails for translating and sharing insights from model outputs and its implications

Ensure Privacy Protection & Security

Ensure solutions are de-identified to protect personal data.

Ensure Accountability

Ensure accountability in business stakeholder engagement

Ensure Transparency

Ensure transparency in assumptions made in data sufficiency (e.g. testing scenarios)

AI Ethics Risk Matrix

- Each category expands to further questions for investigation
- Likelihood of occurrence & impact will vary depending on dataset and model being subjectively assessed

Risk Category	Principle	Risk Description	Likelihood of occurrence (0-5)	Impact (0-5)	Mitigation	Post Mitigation Likelihood of occurrence (0-5)	Impact (0-5)
Privacy & Security	Personal Information (PI) should be protected from theft, unintended use, malicious use	Record level data needed to run model and achieve statistical relevancy makes data identifiable	(0-5)	(0-5)	De-identify all record level data with common identifier for matching and linking. No PI data used for AI Security Protocols for data	(0-5)	(0-5)
Bias	Ensure data and models are fair and not affected by human bias	Data or data models reflect the designers bias	(0-5)	(0-5)	Leverage methodologies or packages to ensure data bias is addressed and tested for e.g. Fairlearn package	(0-5)	(0-5)
Inclusivity	Data should fairly represent all population characteristics	Model does not represent all genders and ethnicities, skills, abilities	(0-5)	(0-5)	Test models for factors related to under-representation especially demographic factors	(0-5)	(0-5)
Data Provenance	Data lineage is clear	Data is obtained through un-verified sources or with malicious intent. Findings are not trustworthy	(0-5)	(0-5)	Validate data from source systems, understand recency of data	(0-5)	(0-5)

AI Ethics Risk Matrix

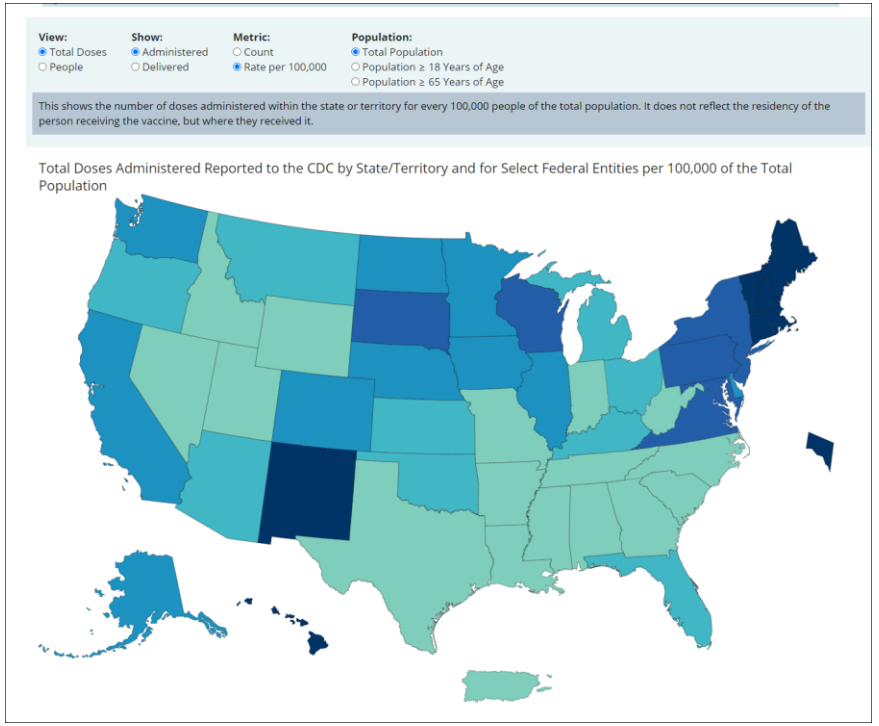
Risk Category	Principle	Risk Description	Likelihood of occurrence (0-5)	Impact (0-5)	Mitigation	Post Mitigation Likelihood of occurrence (0-5)	Impact (0-5)
Auditability	Data models can be audited	Drop in trust of the data, models may be shelved and not used	(0-5)	(0-5)	Create comprehensive business rules for audit and validation of data	(0-5)	(0-5)
Interpretation	Ensure data scientists are held accountable in how they interpret data sets and design questions	Mis-interpretation of data can lead to skewed or bad training sets and models that well trained on skewed/bad data.	(0-5)	(0-5)	Peer-review, departmental review, community of practice	(0-5)	(0-5)
Contestability	Ensure timely response when AI use or outputs are challenged	Contesting data	(0-5)	(0-5)	Peer-review, departmental review, community of practice	(0-5)	(0-5)
OVERALL RISK			(Sum)	(Sum)		(Sum)	(Sum)
AVERAGE RISK			(Average)			(Average)	

Data Set Assessment

Covid – 19 Vaccinations from CDC[2]
State Level – Actual Vaccination numbers are reported

Principles	Risk Assessment Questions	Response	Risk Level
Privacy & Security	Is the data aggregated enough to de-identify population?	Yes, state level data is highly aggregated	1
Bias	Are there demographic factors to be aware of?	Yes, data can be reported by age group. Depending on collection method (Self reported vs vaccination center reported), there may be possibility of bias	3
Inclusivity	Does data represent all population?	Yes, no population exclusion indicated	1
Data Provenance	Is data source clearly stated?	Yes, as stated data sourced from jurisdictional partners, pharmacies etc	1
Auditability	Can data be audited?	Maybe, through contacting the CDC. The CDC is a reputable source of data.	1
Interpretation	Is it clear what the data represents?	Yes, total aggregated state level vaccine counts	1
Contestability	Can outputs of data be contested?	Yes, but data provenance is clear	1

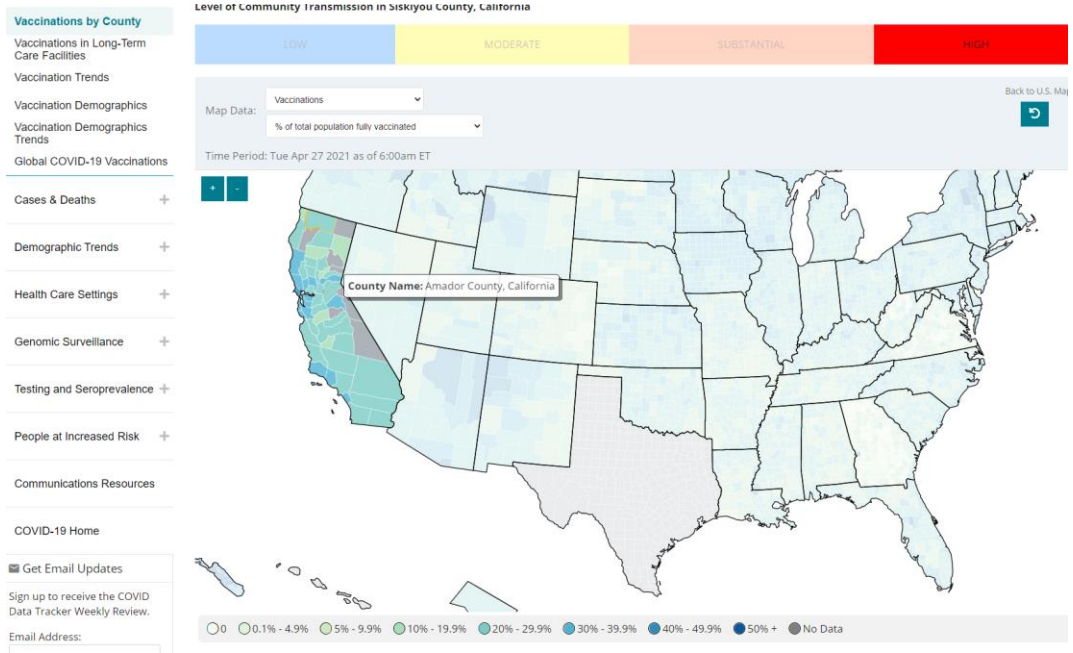
2 https://covid.cdc.gov/covid-data-tracker/#datatracker-home



Data Set Assessment

Covid – 19 Vaccinations from CDC[2]
County Level – Only % Vaccinated Rates are provided

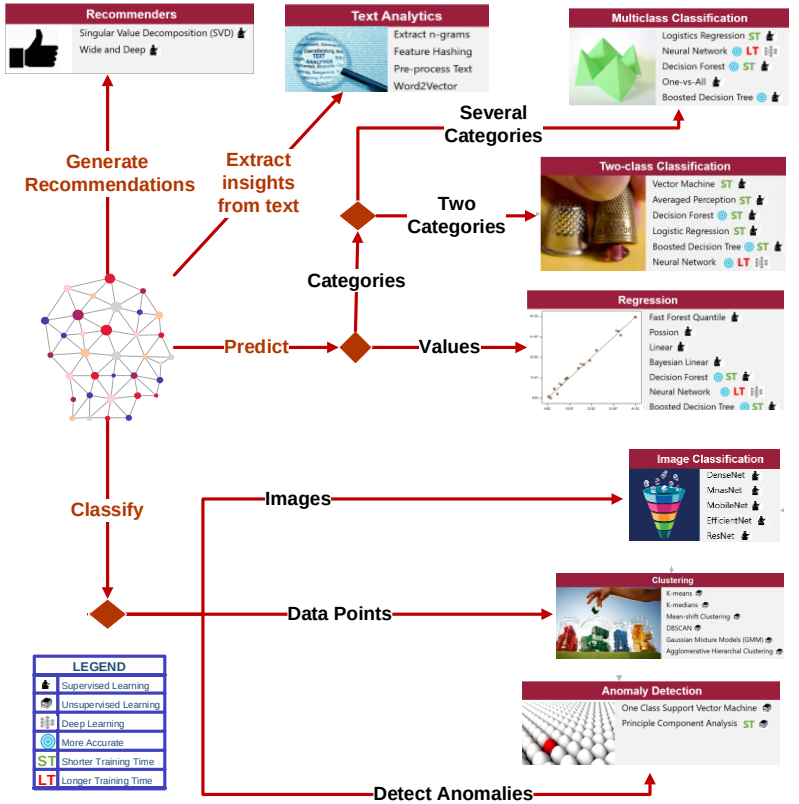
Principles	Risk Assessment Questions	Response	Risk Level
Privacy & Security	Is the data aggregated enough to de-identify population?	County level data is more granular than state level but CDC does not report county level numbers – only % vaccinated rates. This is a good way to protect against identification risk! Adding another data set may make it identifiable	1
Bias	Are there demographic factors to be aware of?	Yes, data can be reported by age group. Depending on collection method (Self reported vs vaccination center reported), there may be possibility of bias. The addition of ethnicity based stats may increase risk of bias	3
Inclusivity	Does data represent all population?	Yes, no population exclusion indicated	1
Data Provenance	Is data source clearly stated?	Yes, as stated data sourced from jurisdictional partners, pharmacies etc	1
Auditability	Can data be audited?	Maybe, through contacting the CDC. The CDC is a reputable source of data.	1
Interpretation	Is it clear what the data represents?	Yes, total aggregated state level vaccine counts	1
Contestability	Can outputs of data be contested?	Yes, but data provenance is clear	1



2 https://covid.cdc.gov/covid-data-tracker/#datatracker-home

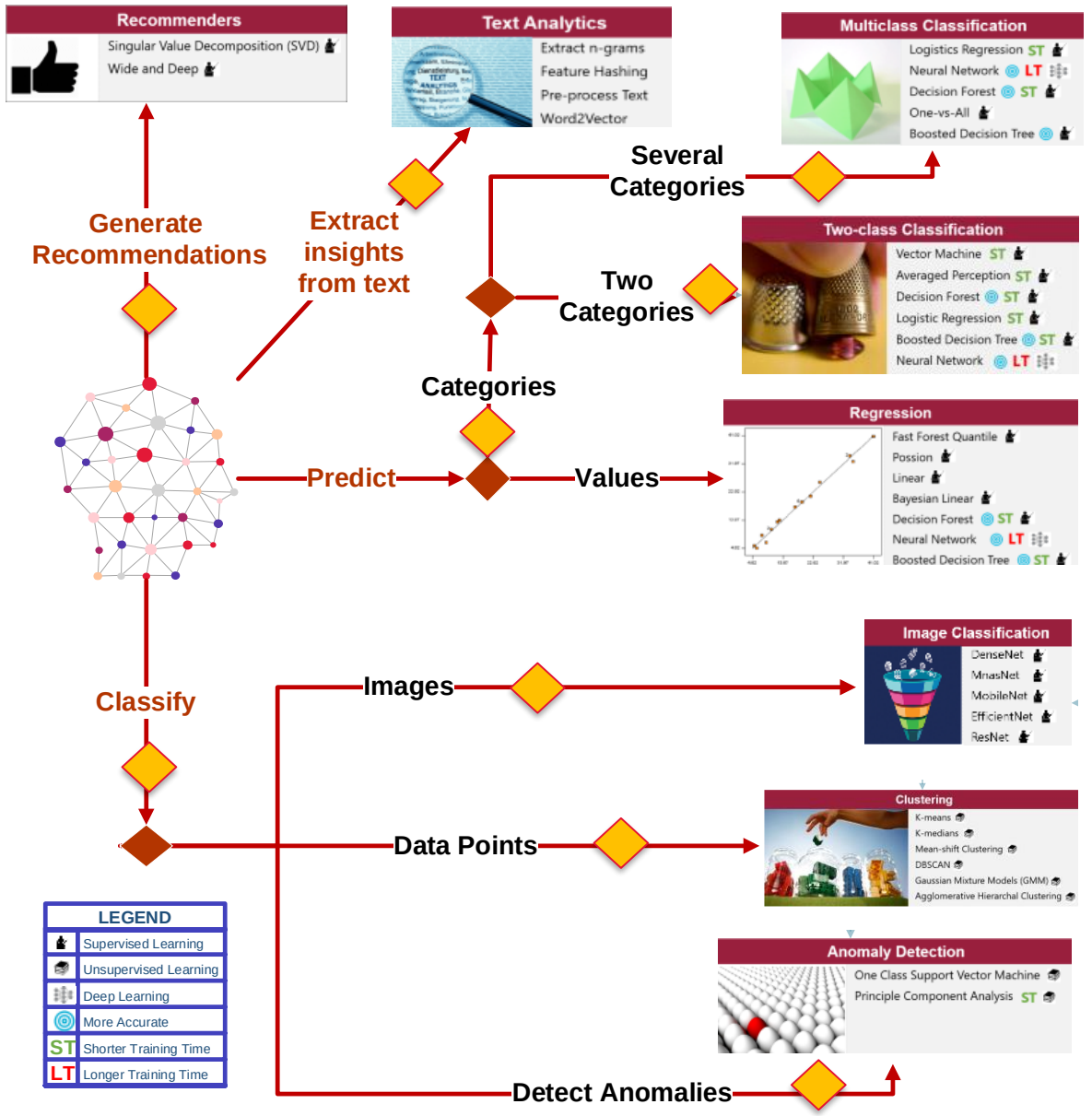
Conclusions and Recommendations

- Risk Model is appropriate for quantifying the AI Ethics Risks across the spectrum of AI implementation
- Mitigations and quantification of post mitigation risk is appropriate to identify remediation
- The model should be applied as part of the Governance Model for all AI development
- Framework is applicable for data analysis, analytics and AI product development
- Sub questions may need adaption for Intelligent Automation vs Data Analytics and Decision Support
- Further analysis across AI Methodologies recommended and integration of ethical framework with AI model decision-tree



Sample Ethical AI Decision Points

- Is the problem statement unbiased and inclusive?
- Is the data source reliable, high quality and trustworthy?
- Is the dataset providing a scope of demographics?
- Is there statistical relevance?
- Can the data be re-identified when new data added?
- Will the model risk re-identification when localized
- Is the approach and methodology transparent to decision-makers?
- Is the output auditable for continuous improvement?



Questions