Infosys topaz

VIEW POINT

GENAI INTEGRATED DATA MANAGEMENT SOLUTION FOR INTELLIGENT AND SECURE DATA TRANSMISSION – A GOVERNANCE FRAMEWORK

Abstract

Multiple B2B players exchange data to perform business operations that benefit the end customer. However, Industry problems like Bad data quality, out-of-sync data, and data translations (inter-operability) between providers cause significant operational overheads and long onboarding times. Moreover, the humongous amounts of data that is consumed at various stages of any AI platform raises legitimate concerns around governance and acceptable usage. Simultaneously, compliance with varying regulatory and legal requirements across geographies and domains require a comprehensive framework to Scan, Shield and Steer for Responsible AI design, deployment, and monitoring.

Infosys®
Navigate your next

Data Exchange Platform Solution

Leveraging Autonomous Engineering and OpenAI to address data management for data exchange platform

Solution Enablers

GenAI, Machine Learning, Artificial Intelligence, Privacy First

Solution Summary

A cloud-based data exchange platform provides a secure, distributed, real-time, industry-specific data collaboration environment that enables multi-party exchange of data with configurable data types and configurable rules. The platform offers standard data formats, data cleansing tools, scheduling tools for data aggregation, and migration to/from source and target systems.

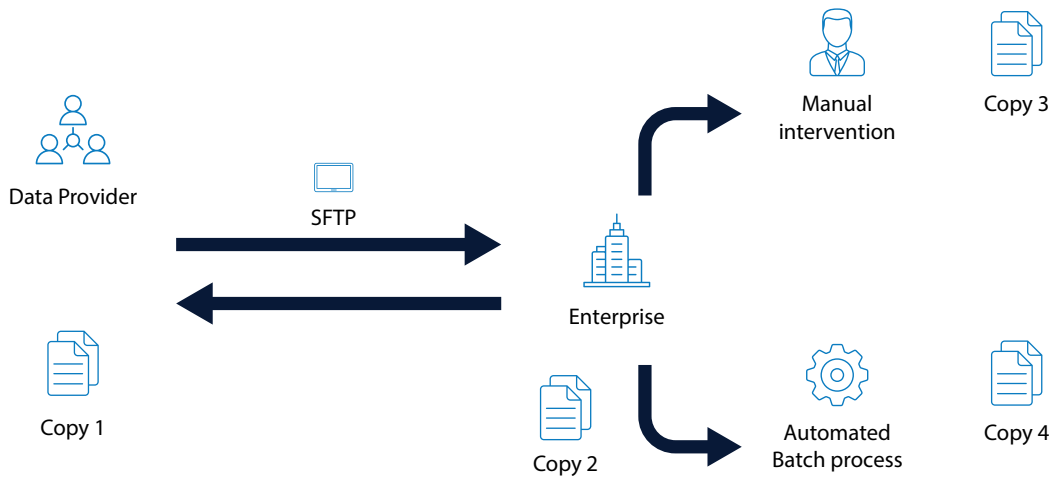
Snapshot of Solution

Leverage the Infosys data engineering IP, data connectors, industry knowledge and cloud collaborations to build a plug and play data exchange platform. Leverage industry consortiums and regulations to standardize the formats. Offer translation service to transform data from one vendor format to another seamlessly. The onus of data quality is on the provider based on mutually agreeable (configurable) rules, thus, significantly eliminating bad data and operational overheads.

Key Value Propositions

- Data enrichment services: Data cleansing, harmonization, deduplication, integrated third-party API address correction, golden record generation and enforcement.
- Selective sharing: Allows data suppliers to selectively share data based on access control and filters.
- Data mapping: GenAI based attribute mapping between heterogeneous databases based on attribute definition.
- Multi-tenant data sharing: Replace FTP and API-based sharing mechanisms with cloud-based, multi-tenant solutions
- Alerts: Notifications and alerts for consumers when there is any change or update to existing data.

Current State Data Exchange



External Document © 2024 Infosys Limited

Future State Data Exchange

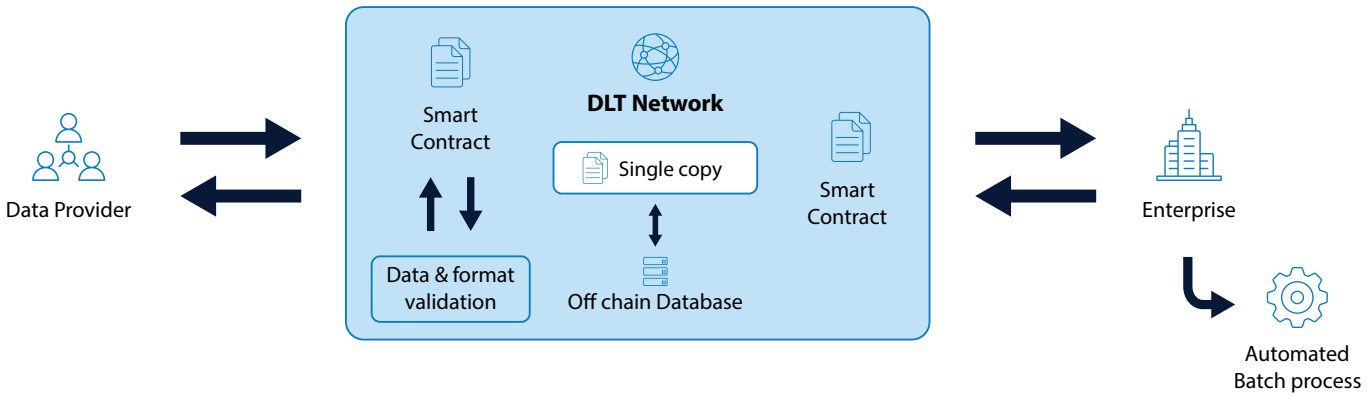
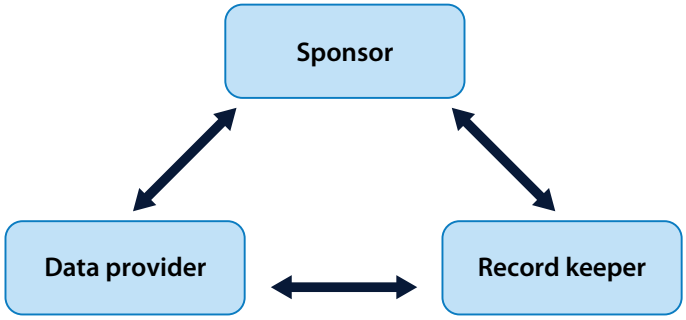
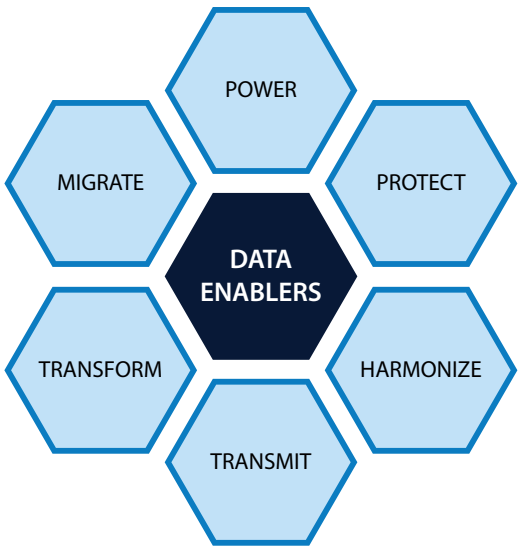


Figure 1: Operational Data Flow between Entities





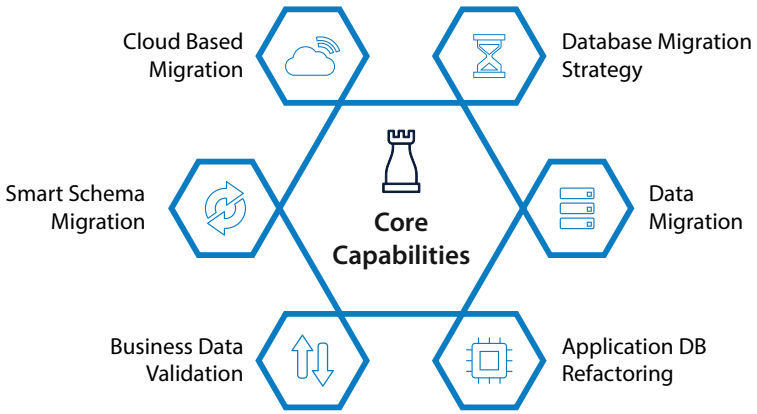


Figure 2: AI Levers and Capabilities for Data Exchange

External Document © 2024 Infosys Limited

Business Proposition

Time and effort to onboard new clients – Typically, new sponsor project setup can take anywhere between 4-8 months. This includes customizing plans, agreeing on data formats and frequencies, and programming data exchange processes. Several hundred man-hours are needed both on the provider and record keeper stakeholders or partners to align and set up the required processes for data.

- **Legacy Data Exchange –** Sponsors typically prefer to receive data in a standard file format from their partners. Programming & testing account for most of the time and effort required to automate payroll integration.

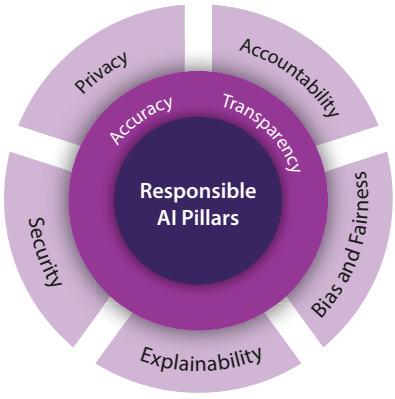
Data Exchange Solution – The Infosys data exchange solution, with its ability to configure rules (as opposed to programming), assists in faster adaption of different file formats across partners and vendors. This eliminates setup, programming time, and format translation time, leading to more transparent data access and use.

Key Solution Components in Autonomous Data Processing

- **Infosys Data Migration Solution:** It is a comprehensive end to end data migration management solution for legacy modernization and enterprise transformations. It streamlines data migration by optimizing pipelines, capturing and analyzing exceptions for improved data availability.
- **Infosys Database Migration Solution:** It provides seamless end-to-end database migration, including application database refactoring, business data validation, and AI-powered schema migration.
- **Data Privacy and Security Solution:** It protect enterprise data with AI-based Data Discovery, Data Protection, and Synthetic Data Generation Capabilities.
- **Infosys Data Workbench:** A comprehensive suite of tools for data cleansing, harmonization, and standardization using pre-built auto-cleansing libraries.

AI presents a large economic opportunity, and with increasing adoption rates, AI is set to automate many decision-making processes. Even though AI presents a large economic opportunity, it brings significant risks due to extensive data processing by systems, which is persistent, pervasive and is leveraged for training behavioral patterns.

AI will automate many decision-making processes. Hence, they require a comprehensive framework to Scan, Shield and Steer for Responsible AI design, deployment, and monitoring



Privacy Accountability Fairness

Explainability Bias Transparency

Accuracy Security Human Controlled

A holistic approach to developing, assessing, and deploying AI systems in a safe, trustworthy, and ethical way.

Figure 3: Infosys Responsible AI Components

External Document © 2024 Infosys Limited

This is especially important when automated systems deal with personal and even more highly sensitive personal data. This document lists down the essential outlines and principles in operationalizing AI for data tools across the lifecycle handling stages from ingestion, processing, privacy and finally transferring the processed data to downstream applications for further use.

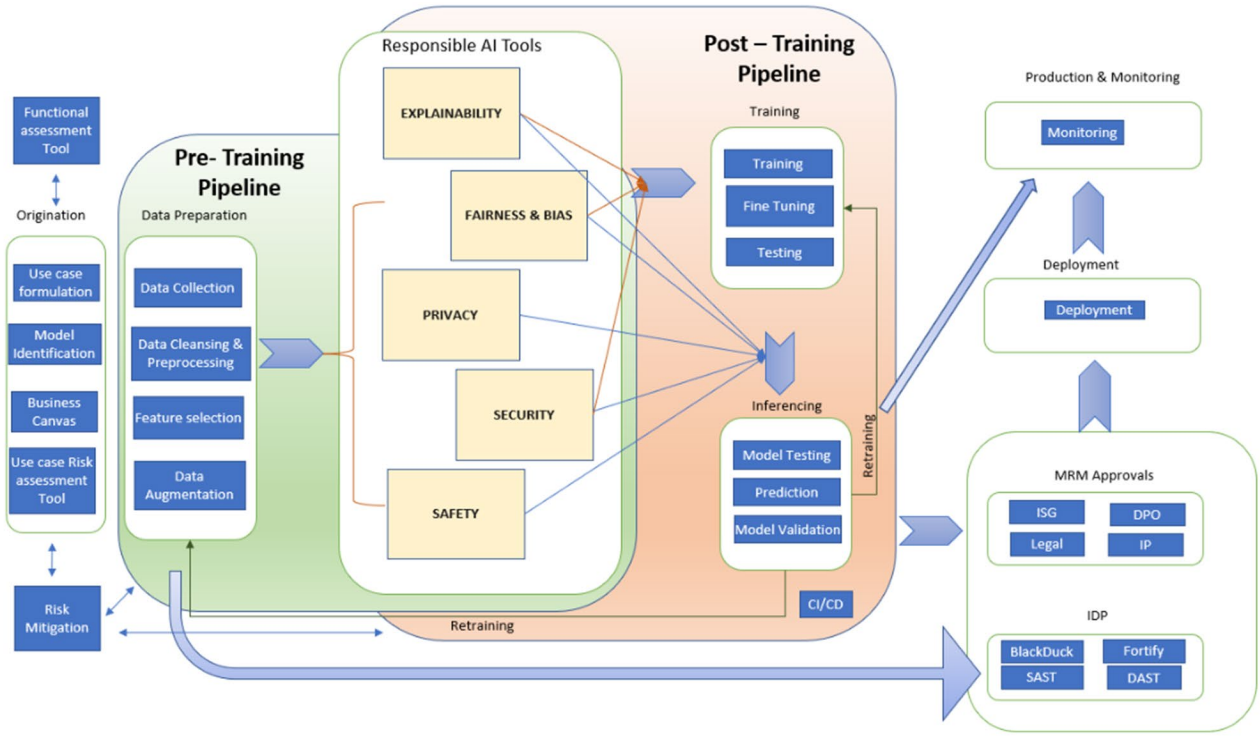
Whether you are a consumer, developer, or in any part of AI lifecycle, one should be prepared to manage the risks that AI poses at both social and governance components. As applications consume, maintain, manage, and expand embedded data and implement new algorithms for generating insights, governance is a critical guardrail ensuring ethical end objective mandated usage. This becomes even more important as AI systems are utilized across various solutions and modules.

Responsible Artificial Intelligence (Responsible AI) is an approach to building and deploying AI systems according to the pillars of privacy, accountability, fairness, explainability, security, accuracy, and transparency. The key objective of Responsible AI is to keep people and their decisions (stated or unstated) at the center of system design decisions while respecting values of fairness, reliability, explainability and transparency.

We will now explore some key pillars and their corresponding algorithms within **Infosys Responsible AI Policy** framework, highlighting their connection to ICETS IPs.

Pillar	Key Components & Use	Details
1. Privacy	• Data Protection • Data Minimization • Fair & Transparent Data Use • User Control • Data Security - Data Collection and Usage	
2. Accountability	Governance and accountability for actions • Transparency • Explainability • Responsibility • Audits • Complaints Handling • Legal / Regulatory framework	
3. Bias and Fairness	Causes unfair outcomes • Representation Bias • Aggregation Bias	• Blind Taste Tests Fairness & Bias - AI First Live Enterprise | External Share (external-share.com)
4. Explainability	Difficult to understand complex AI algorithms / decisions. Understand with Regression, Classification & Clustering	• Local Interpretable Model-Agnostic Explanations (LIME) • DeepLIFT (Deep Learning Important FeaTures)
5. Security	Security risks as cyber-attacks due to system vulnerabilities and AI driven autonomous weapons due to lack of human control	
6. Accuracy		
7. Human Controlled		
8. Transparency		

External Document © 2024 Infosys Limited



1. Privacy

Privacy ensures that every individual has knowledge and ability to exercise discretion on the personal information collected and its intended usage Since responsible data collection, reusability, utilization, and security are fundamental for any AI system, and because most datasets inevitably contain personal information, robust and secure data privacy practices are essential. To ensure responsible stewardship of personal information, including both Personal Identifiable Information (PII) and Sensitive Personal Information (SPI), we must implement controls and safeguards that comply with all relevant privacy laws and Infosys Privacy Principles.

IEDPS has the enabled safeguards to secure personal information PII & SPI with the following modules and capabilities –

- Sensitive Data Discovery and Tagging
- Data Anonymization
- Privacy Risk Assessment
- API Data Generation

2. Accountable

The aim is to enhance accountability of data use in AI systems. The key guidelines can be outlined as:

Transparency and Interpretability:

- Prioritize transparency, visibility, traceability, and interpretability in AI systems.
- Ensure that decision-making processes are understandable not only to developers but also to users, consumers, and external parties.
- Transparency builds trust and confidence.

Explanations for Decisions:

- Develop AI systems that not only provide transparent decisions but also offer comprehensible explanations.
- Users should understand the rationale behind AI-based decisions.
- Regularly review these explanations to ensure accuracy.

Process Documentation:

- Establish mechanisms for transparency and documentation.
- Document AI system processes, algorithms, and decision logic.
- Clear documentation enhances interpretability.

Interdisciplinary Collaboration:

- Collaborate with experts from various domains (e.g., AI, ethics, design).
- Diverse perspectives facilitate the development of more interpretable AI systems.
- Interdisciplinary teams can address complex challenges effectively.

Human-Centered Design:

- Incorporate human-centered design practices.
- Design AI systems with the user in mind.
- Encourage AI model outputs and decisions that are comprehensible and effectively communicated.

Explainability Techniques:

- Leverage explainability techniques when applicable.

External Document © 2024 Infosys Limited

- Examples include feature importance analysis, model visualization, and rule-based models.
- These techniques contribute to overall explainability and interpretability.

3. Bias and Fairness

As AI-enabled systems consume relevant historic data for training in order to understand and embed patterns and logic that were leveraged for making crucial decisions, there is always a chance of in-built bias in the data being consumed, e.g., computational, and societal bias in data. It is important to realize that Bias isn't a technical issue only. It is not practical to identify or avoid data with any prevalent bias issue. The aim should be to document and mitigate bias issues in raw data and focus should be on documenting the inherent bias in the data and features while building processes & methods to identify features and inference results so the right procedures can be put in place to reduce potential risks subsequently.

Bias in AI starts with the data generation and collection. Various stages in data development and data labeling contribute to the overall bias of an AI-ML model.

When these biased data sets train an algorithm, it results in a **Biased algorithm/model.** There should be a fair evaluation of the intent, inputs, and evaluation criteria of an algorithm. Models developed on biased data result in inaccurate and discriminatory outputs. This drives the concepts of responsible data and training models on datasets that represent the target population accurately. Establishing policies and practices to monitor and mitigate bias in every stage of model evaluation, ensuring all algorithms and models are labeled for their biases and frequent, time-bound evaluation to ensure elements of the model and data do not discriminate any particular community or an ethnic group are essential to a Governance process or framework for removing Bias.

Typical types of Bias are -

- **Representation Bias**
- **Aggregation Bias**

Some of the main ways Bias is identified in data are outlined below -

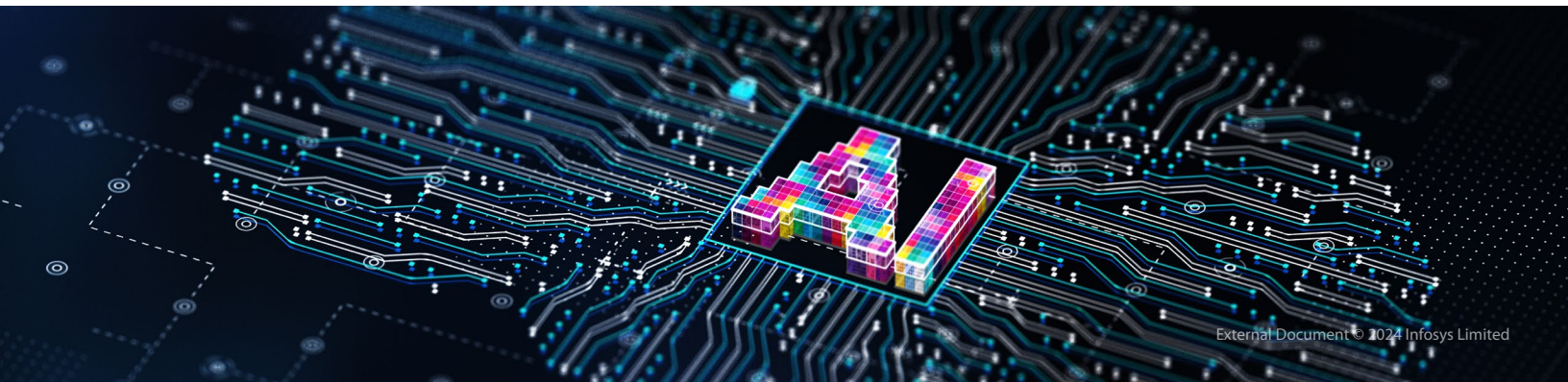
Performance Disparity: These metrics identify potential bias in model performance across different subgroups within the data and include:

- Difference in Accuracy (rate)
- Difference in Error (rate)
- Difference in Precision
- Difference in Recall or Reproducibility
- Difference in Mean Absolute Error (MAE)

Selection Rate Disparity: This metric specifically focuses on the difference in the rate of favorable predictions (e.g., loan approvals) across subgroups. It compares the fraction of data points classified positively (for binary classification) or the distribution of predicted values (for regression) across different groups.

- Some common ways for Removal of bias are:
  - **a.** Setting goals around fairness objectives for system, considering different end users.
  - **b.** **Measure & discover** disparities in potential outcomes and sources of bias across various users or groups.
  - **c.** **Mitigate** any unintended consequences using proposed remediation strategies.
  - **d.** **Monitor & control** systems with processes that flag and resolve future disparities as the AI system evolves.

Bias Removal Objective	Bias Removal Objective	Details
Achieve Demographic Parity	To mitigate and optimize allocation harms	Binary classification, Regression
Equalized Odds	Diagnose allocation and quality-of-service harms	Binary Classification
Equal Opportunity	Diagnose allocation and quality-of-service harms	Binary Classification
Bounded Group Loss	To mitigate quality-of-service harms	Regression



External Document © 2024 Infosys Limited

The following are leveraged to mitigate Bias.

Bias Mitigation Algorithms: Bias mitigation algorithms are categorized into three categories as data considered for AI models: **pre-processing, in-processing** and **post-processing** algorithms.

Parity constraint	Purpose & Applicability
Adversarial Debiasing	In-processing fairness mitigation using adversarial training and involves simultaneous training of a predictor and a discriminator
Reweighing	preprocessing Weights in each (group, label) combination to ensure fairness before classification
Model Documentation	
Optimized Pre-processing and rejection of options -based on classification	Edit features and labels in the data with group fairness, individual distortion and data fidelity constraints and objectives through a probabilistic transformation

Fairness Identification Parameters: The following parameters are used to indicate bias optimized data for AI use

Fairness Indicator	Purpose
Positive Rate / Negative Rate	Percentage of data points that are classified as positive or negative, independent of ground truth
True Positive Rate / False Negative Rate	Percentage of positive data points (as labeled in the ground truth) that are correctly classified as positive, or the percentage of positive data points that are incorrectly classified as negative
True Negative Rate / False Positive Rate	Percentage of negative data points (as labeled in the ground truth) that are correctly classified as negative, or the percentage of negative data points that are incorrectly classified as positive
Accuracy & AUC	Predictive Parity
False Discovery Rate	Percentage of negative data points (as labeled in the ground truth) that are incorrectly classified as positive out of all data points classified as positive
False Omission Rate	Percentage of positive data points (as labeled in the ground truth) that are incorrectly classified as negative out of all data points classified as negative

4. Explainability

AI systems should be continuously improved to explain the predicted results based on features and / or models chosen. While the accuracy may decrease, the resultant increase in transparency and explainability in processes helps increase trust in AI systems and in the embedded decision-making process. Understanding decisions is a key human need. Reproducibility of machine learning systems, running simulations, and comparing the output with the training data set helps finetune and understand logic and builds trust.

Transparency Builds Trust:

- When AI systems are transparent, users and stakeholders can better understand how decisions are made. This transparency fosters trust because people feel more comfortable when they can see the inner workings of a system.
- In financial services, where decisions can have significant consequences, transparency is even more critical. Users want to know why a loan application was approved or denied, how investment recommendations were generated, and what factors influenced those decisions.

Technical Soundness and Understandability:

- Transparent AI systems are technically sound. This means that their algorithms and models are well-designed, thoroughly tested, and free from biases.
- Understandability is equally important. Stakeholders should be able to comprehend how AI system arrived at a particular outcome. If the process is too complex or opaque, it erodes trust.

Accountability and Responsibility:

- Transparent AI systems are accountable. There should be clear ownership and responsibility for their development, deployment, and maintenance.
- When something goes wrong, stakeholders need to know who is responsible. Accountability ensures that errors are addressed promptly and that corrective actions are taken.

Collaboration and Empowerment:

- Transparency encourages collaboration. When stakeholders understand how AI works, they can provide valuable feedback and contribute to its improvement.

Empowering users with information about AI-generated insights allows them to make informed decisions. Whether it's an investment recommendation or a credit score, transparency enables users to assess the reliability of the system and provides **visibility** into every application's hidden factors.

Traceability:

It is necessary to documenting design and decision-making processes to the point where if a mistake occurs, it can be reverse-engineered to determine why and what transpired and do a root cause analysis.

Explainable:

Necessary to ensure data and results are explainable in a way that humans can interpret.

External Document © 2024 Infosys Limited

Prediction-accuracy and traceability address **technology requirements**, while decision-understanding addresses **human needs.** The following steps detail the process of unravelling the back box with relevant control experiments:

How To -> (Interpretation and Gaining Understanding) –

- Document design and decision-making processes to the point where if a mistake occurs, it can be reverse-engineered to determine what transpired.
- Ensure data is explainable in a way that a human can interpret.
- Visibility into every application's hidden factors.

What–if analysis and Algorithms to understand (XAI Process)---

Indicator	Machine learning Algorithm Indicated
Proxy Modeling	Causal analysis help in estimating the effect of a feature on an outcome of interest on average, across a population or a cohort, and on an individual level
Interpretability by Design	
Algorithmic Tools	LRP Layer wise Relevance Propagation
	LIME (Local Interpretable Model-Agnostic Explanations)
	SAP (Shapely Additive Explanations)
Counterfactuals	This helps in disentangling the impact of correlated features in isolation. This is useful to examine fairness and reliability criteria
Integrated Gradients	

What-If Counterfactuals and their significance in understanding and debugging machine learning models -

- Address the question of what the model would predict if the inputs were changed by revealing how a model reacts to input changes.
- Counterfactual analysis helps disentangle the impact of correlated features and provides nuanced insights.

Importance -

- Unlike standard interpretability techniques that approximate a model or rank features, counterfactuals directly interrogate the model.
- They help identify feature changes needed to flip a classification decision or alter a regression output.

Application -

- Useful for model debugging, fairness assessment, and improving trust.
- By exploring counterfactuals, we gain insights into model behavior and decision boundaries.

5. Security

- All systems leveraging AI need to be mitigated against a multitude of security threats including Evasion, Poisoning, extraction, and Inference attacks.
- Strong data governance framework that includes traceability, data lineage and access control are vital components that need to be built in to ensure data security.

Various Security features built in include:

- Secure SSL for at-rest data processing tasks
- WS-Security, Content is Base 64 encoded for in-flight data
- Spring security (uses bearer/JW Token) to access different services
- HTTPS enabled with SSL/TLS encryption
- Encryption by AES256 algorithm

6. Accuracy

Model training, Data quality and Error analysis are key parameters driving accuracy parameter of Responsible AI. A continuous improvement program requires insights into model behavior and output in stress test conditions.

Model validation includes testing features or the and involves the following –

Transparent Predictions: The model should be able to explain its reasoning behind predictions. This is achieved by generating feature importance scores, highlighting which features have the strongest influence on the model's output. These scores can be calculated for the entire model (providing a global explanation) or for individual data points (offering local explanations).

Real-World Interpretability: The model should be understandable when applied

External Document © 2024 Infosys Limited

to real-world datasets, both during training and when making predictions (inference). This means humans can grasp the model's decision-making process and how it works with real-world data, not just the training data used to build it.

The accuracy test includes various component parameters –

- **Target feature:** The feature that the model was trained to predict.
- **Error analysis:** An error tree and heatmap visually represent errors in the model's predictions.
- **Error heatmap feature:** Select up to two features to pre-generate an error heatmap.
- **Counterfactual explanations:** This component generates "what-if" scenarios that illustrate how small changes to a data point could have resulted in different model predictions.
- **Number of counterfactuals:** This specifies the number of counterfactual examples generated per data point, allowing for a more granular analysis.
- **Generate explanations for Causal analysis:** Document model explanations for AI dashboard. This will help in understanding the causal effects of various features.

7. Transparency

Transparency is essential to be able to explain both the overall decision-making process and the individual predictions generated by your AI decisions, especially when those decisions impact people's lives. Explainability is key for developing trust and increased adoption of AI systems.

8. Accountability through human oversight

It is important that identified humans and actors are accountable and have clear ownership and responsibility for their development, deployment, and maintenance, and most crucially over outcomes of AI systems. Human oversight ensures that errors are addressed promptly and that corrective actions are applied throughout the lifecycle of AI driven solutions as per established regulations.

Governance needs to be appropriate and current with SLAs and escalation hierarchy should be clearly established.

AI is often a black box rather than a transparent system, meaning it's not that easy to explain how things work from the inside. After all, while people sometimes can't provide satisfactory explanations of their own decisions, understanding complex AI systems such as neural networks can be difficult even for machine learning experts.

To prevent potential negative outcomes, companies must regulate their artificial intelligence programming and set a clear governance framework in advance with pre-defined principles, ethics, and rules to govern AI.

Responsible AI will continue to be an active area of research and development with the aim of creating standardized guidelines for organizations operating in different industries.

External Document © 2024 Infosys Limited

References

- Model References and Fairness - Machine learning fairness - Azure Machine Learning | Microsoft Learn
- Responsible by Design - Responsible by Design - AI First Live Enterprise | External (external-share.com)
- Fairness and Bias - Fairness & Bias - AI First Live Enterprise | External Share (external-share.com)

About the Authors

Eggonu Vengal Reddy is a Principal Product Architect with over 20 years of experience in Data Management, specifically Data Warehousing, Data Modeling, Big Data, and Data Science. He has provided architecture and design to develop tools and solutions to handle enterprise-wide database migrations, master data management, data quality and wrangling, explorative analysis, and feature engineering in the Machine Learning life cycle.

Tushar Subhra Das is a Senior Business Data Analyst with over 10 years of experience in Data Migration and Governance. He has worked with Europe and Australia-based clients for Data & Database Migrations, MDM and Data Quality, and Process governance. In his current role, Tushar is responsible for APAC and EMEA data migration deployments and enhancements, including product developments for iDSS as the next-generation industry-standard data management platform.

Gopinadh Bapatla is a Senior Technology Architect with over 20 years of experience in Application Development, Data Analysis, Data Science, Machine Learning, and Big Data. He has provided architecture and design to develop tools and solutions to handle data quality by data wrangling, explorative analysis, feature engineering, and building Machine Learning models.

External Document © 2024 Infosys Limited

Infosys Topaz is an AI-first set of services, solutions and platforms using generative AI technologies. It amplifies the potential of humans, enterprises and communities to create value. With 12,000+ AI assets, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale and build connected ecosystems. Connect with us at infosystopaz@infosys.com.

For more information, contact askus@infosys.com

Infosys®
Navigate your next

© 2024 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.