White paper

Responsible AI in Practice

Teradata's ClearScape Analytics™ enables responsible AI at scale

Trish Lugtu
Principal Data Scientist

teradata.

Table of Contents

Executive overview

The concept of responsible artificial intelligence (AI) is gaining traction, with many organizations now recognizing its role in mitigating technology’s risks. However, the topic is largely misunderstood, and there are few resources explaining how to put it into practice.

A common misconception is that responsible AI is simply about implementing AI in an ethical way. This could be due to increased awareness around potentially discriminatory outcomes that can result from AI solutions with bias. In fact, ethics is only one component of responsible AI—albeit an important one.

Responsible AI is a comprehensive approach to practicing ethical AI while striving for accountability, compliance, and good stewardship to positively impact customers and empower organizations.

This white paper provides practical guidance on implementing AI solutions according to these principles.

It’s time for responsible AI

With the AI disruption caused by GenAI and large language models (LLMs) especially in the last year, the need to adopt a well-defined set of responsible AI practices has never been more pressing.From 2019 to 2020 alone, executive ownership of AI initiatives rose sharply from 39% to 71%, and IT initiatives more than doubled at companies with budgets over $5 million.[1] One poll found that 55% of companies in 2020 reported that the pandemic accelerated AI strategies, and another survey revealed that AI adoption generally doubled between 2017 to 2022.

Amid this surge in AI utilization, many companies rushed to implement solutions without establishing the responsible practices needed for governance. By late 2022, most organizations that had moved forward with their AI strategies continued to lag with immature responsible AI programs.[2]

Building a vision for responsible AI in practice

In a Nature Machine Intelligence article published in 2020, the authors called on organizations to address “ethics with urgency” to buttress the rapid increase in AI deployment.[3]

Today there is a growing consensus that the responsible practice of AI is necessary for implementing analytics at scale, yet many organizations still lack the governance framework and road map needed to embed it into their AI strategies.

So, what does responsible AI look like in practice? Let’s explore the key themes and practices that enable companies to operationalize responsible AI at scale.

Three key themes of responsible AI

All responsible AI practices center around three key themes: governance, ethics, and efficiency.

- **Governance** builds accountability and addresses risk and compliance.
- **Ethics** calls for continued improvements to model fairness and transparency.
- **Efficiency** addresses practical operations for growing analytics at scale.

Together these themes form the pillars of the responsible AI framework that support every stage of the AI lifecycle. Using this framework not only enables companies to establish a foundation for practicing responsible AI, but also to sustain it at scale. AI that is implemented responsibly builds trust among customers and within organizations.

Responsible AI Themes

Governance	Ethics	Efficiency
Accountability	Fairness	Operations
Risk Compliance	Transparency	Scale

Figure 1. Responsible AI themes

1 https://appen.com/whitepapers/the-state-of-ai-and-machine-learning-report/
2 https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review
3 https://www.nature.com/articles/s42256-020-0195-0

Governance:
The foundation of responsible AI

Governance is the foundational glue that binds and strengthens the framework. To better explain governance, we borrowed from one of Teradata's solution integration partners, Deloitte, which introduced the Trustworthy AITM framework. In this model, Trustworthy AI is the core of a hub that is wrapped with an inner layer of regulatory compliance and an outer layer of AI governance. Practices that support the responsible AI strategy extend from this hub like the spokes of a wheel.

In Deloitte's model, AI governance spans all the dimensions of responsible AI. Deloitte further explains this about AI governance:

"At its foundation, AI governance encompasses [all stages throughout the AI lifecycle], and is embedded across technology, processes and employee trainings [to develop and ensure ethical safeguards across all dimensions]. This includes adhering to applicable regulations, as it prompts risk evaluation, control mechanisms, and overall compliance. Together, governance and compliance are the means by which an organization and its stakeholders ensure AI deployments are ethical and can be trusted."[4]

In organizations without AI governance, analytics projects tend to grow autonomously, with practices varying from project to project. When there are no standards to guide implementation, auditing for risk throughout the AI lifecycle is overlooked. By contrast, with proper AI governance, decisions about technology, people, and processes are guided by the regulations and policies imposed over the complete AI lifecycle from inception to retirement. AI solutions and their applied machine learning (ML) models should be audited for how they are approved, improved and versioned, deployed, evaluated, retrained, and ultimately retired. The feature sets or datasets used to train the ML models should also be versioned and tracked.

How can AI governance be implemented in practice?

The ability to audit ML models for governance within the ML model lifecycle is the first way in which ClearScape Analytics™ helps to implement responsible AI. ModelOps, a component of ClearScape Analytics, is a Teradata Vantage™ extension for automating the management and governance of the ML model lifecycle. Teradata was one of the conceiving members that developed the cross-industry standard practice for data mining (CRISP-DM) in 1996.[5] CRISP-DM helped to define a standard methodology for governing ML model lifecycles. Figure 2 shows an example of the ModelOps interface that indicates which stage of the ML model lifecycle is current for the ML model version.

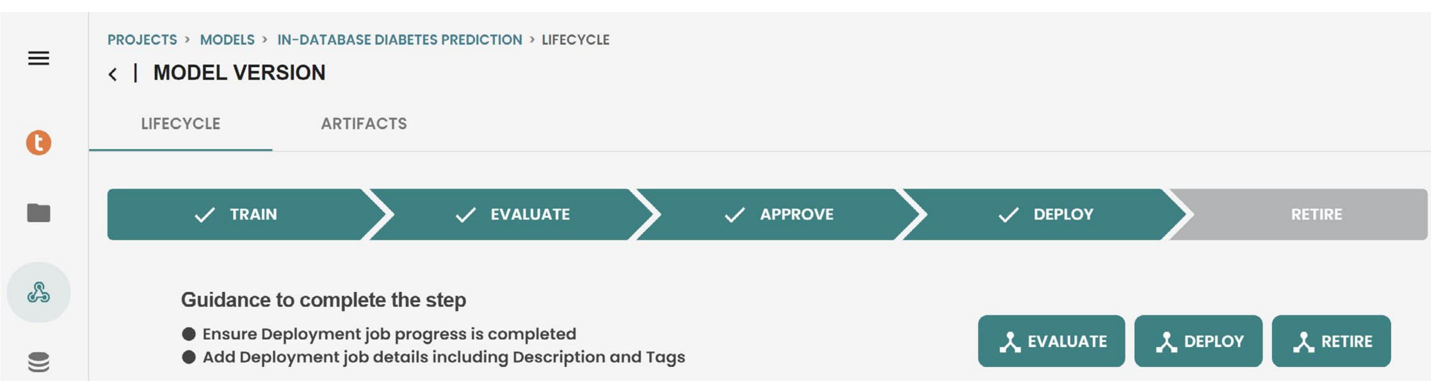


Figure 2. Lifecycle stages of the ML model within ModelOps

4 https://www2.deloitte.com/us/en/pages/deloitte-analytics/solutions/ethics-of-ai-framework.html
5 https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

The ML models within defined, secured projects are tracked for approvals, dataset definitions, evaluation metrics, model versioning, monitoring of feature and model drift, and more. The auditing capabilities readily support regulatory compliance, and the visibility boosts organizational trust. The following sections include additional details about how ModelOps enables responsible AI.

Enabling accountability

Governance requires accountability for success; that is, someone to take ownership of the framework, enable people with processes within the framework, and provide oversight to enforce compliance of the policies and processes adopted.

Another Teradata partner, Accenture, agrees that accountability builds internal trust in an organization's AI technologies, and that without accountability, AI will suffer false starts. According to Accenture, "insufficient clarity on governance and accountability, unnecessary conflicts, and competing incentives across groups ultimately led to responsible AI inertia and a reactive mindset."[6]

This reactivity leads to AI solutions that are ad hoc or tactical rather than strategic, which in turn breeds skepticism of AI. This also illustrates why the communication of governance and accountability is crucial. Organizations that deliberately clarify roles and accountabilities for their AI practices are perceived by employees as working transparently and responsibly.

While accountability of actions can be technically tracked within the ML model lifecycle, accountability is one dimension of responsible AI where a technology solution alone will not suffice. Leadership must ensure the proper mechanisms (strategic, legal, compliance, ethics, and security) are in place to support the successful implementation of responsible AI across the organization.

Organizations that deliberately clarify roles and accountabilities for their AI practices are perceived by employees as working transparently and responsibly.

Enhancing risk and compliance

Along with accountability, governance also requires risk management of AI/ML and collaboration with an organization's compliance leaders. Among the organizations surveyed for Appen's State of AI and Machine Learning 2021, risk management was the primary concern for 60-67% of all enterprises regardless of size.[7] Fortunately, risk management is not new to organizations, and the following standard practices still apply, even to AI risks:

1. Identify the risk.
2. Assess the risk.
3. Prioritize risk mitigation efforts.
4. Assign owners for risk mitigation.
5. Implement mitigation practices.
6. Monitor progress.

6 https://www.accenture.com/us-en/insights/artificial-intelligence/responsible-ai-principles-practice?src=SOMS&
7 https://appen.com/whitepapers/the-2021-state-of-ai-and-machine-learning-report/

However, with responsible AI, it's critical to mitigate AI/ ML risks throughout the entire lifecycle—from ideation to data sourcing, model development, model evaluation, deployment, monitoring, and ongoing maintenance.[8] Table 1 includes some of the potential risk categories for sample AI/ML scenarios.

One way that technology can support risk mitigation is through ongoing monitoring of active ML model deployments and the operations of the serving platforms. ModelOps enables the monitoring of models, evaluation of new data, model metrics, and more with proactive alerting.

Risk Category	Examples
Organizational risk	Operational disruption as machine learning code is lost with employee turnover; lack of transparency; inaccuracy of data
Operational risk	Discontinuity during turnover; no code sharing or versioning
Ethical risk	Discrimination at scale in an ML model so that bias becomes exponential
Regulatory or Compliance risk	Privacy issues, such as unnecessary disclosure of personally identifiable information (PII) or protected health information (PHI) in healthcare during model development
Technical risk	Security issues, such as unaccounted-for PHI on a data science cloud platform outside of the organization's accounted systems
Reputational risk	Bias is found in a commercially sold AI solution
Financial risk	Bias found in a commercially sold AI solution leads to recall of solution

Table 1. Examples of AI risk

8 https://www.mckinsey.com/business-functions/quantumblack/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development

Ethics:
How to responsibly combat bias in AI

Ethics is the component of responsible AI that has gained the most attention, yet it is the least prioritized by enterprises according to Appen's State of AI 2020 Report.[9] Ethics in AI encompasses the concepts of fairness and transparency, which can be impacted by inherent biases in data. Bias can be introduced at various layers of AI, whether through the dataset, the model algorithm, or the AI objectives themselves.

For illustrative purposes, consider a hypothetical ML model that predicts the demographic profile of the next president of the United States. The algorithm learned from historical percentages of each characteristic shown in Table 2.[10,11] Because the algorithm was not aware of current efforts to bring gender equality and diversity to the presidency, it failed to account for the potential influence of these efforts. As a result, it predicted, with a high level of confidence, that the next president will be a Caucasian, Christian male between the ages of 49 and 63. According to this model, there is 0% chance that voters will elect a female, and only a 2-3% chance they will elect a non-Caucasian or non-Christian.

Bias occurred at two levels in this scenario: in the dataset and in the AI objective. In the dataset, bias was introduced because the data did not represent the broader pool of candidates that today's social discourse on diversity would suggest. The dataset was further biased because a political subject matter expert was not engaged to choose the common characteristics or values that would be ideal given the sociopolitical climate. Instead, the selected features included the easiest characteristics to collect. Lastly, the objective itself—predicting the demographic profile of the next president—was biased and arguably unethical. Rather, a model predicting the character values and politics of the next president would have been more appropriate.

The above example is instructive about the consequences of introducing bias to real-world predictive models, such as credit application approvals, health risk assessments for insurance premiums, or financial aid awards for college students. Because machines do what they are programmed to do and perform their tasks inhumanly well, bias in machine learning creates a vicious cycle of discrimination caused by incorrect conclusions drawn from inaccurate predictions.

Characteristic	U.S. Presidents
Male	100%
Caucasian	97.8%
Christian	97.8%
Average age	56.7 years

Table 2. Demographic characteristics of U.S. presidents

In another example, a study published in Science described the discovery of racial bias in a commercial prediction algorithm used by health systems for population health management.[12] The algorithm inadvertently encapsulated racial bias by using healthcare cost as a proxy for illness. Because Black patients had lower healthcare costs, the predictive model incorrectly intimated that they were generally healthier than white patients. In fact, they were equally ill. The algorithm failed to account for socioeconomic disparities that limit access to healthcare for disadvantaged Black patients.

Inadvertent bias also occurs in AI-powered hiring practices, and some companies have discovered their recruiting engines are biased against women.

ML models are trained on past data to find patterns, so if most recruits have been historically male, the algorithm will apply this pattern of gender imbalance to future hiring practices. To eliminate bias from the algorithm, gender would have to be removed from the model.

9 https://appen.com/whitepapers/the-2021-state-of-ai-and-machine-learning-report/
10 https://en.wikipedia.org/wiki/Religious_affiliations_of_presidents_of_the_United_States
11 https://www.statista.com/topics/6272/us-presidents-1789-2020
12 https://www.science.org/doi/full/10.1126/science.aax2342

The Role of Fairness in Ethics

Fairness in AI is accomplished by reducing inadvertent discrimination resulting from bias introduced into the different layers of AI. Fairness must be assessed not just by a data scientist but by a team that includes domain experts and others with diverse perspectives.

Developing a ML approval process that includes review by an ethical AI committee is one way in which companies are incorporating fairness into the governance of AI. Another method is to collect fairness measures and monitor the fairness of ML models and data over time. There are many real-world examples of inadvertent discrimination in ML models. In one study, a deep learning algorithm called a convolutional neural network (cNN) seemed to detect malignant lesions in images of skin more accurately than its human counterparts.[13] However, another study[14] and the National Cancer Research Institute[15] noted a significant lack of images of darker skin in the datasets used to train the algorithm. This AI/ML solution, had it been widely accepted, would have likely resulted in missed skin cancer diagnoses of darker-skinned patients.

Other examples of breach of fairness in AI can be found in the financial sector, where AI/ML-based decisions to grant loans have often been racially biased due to skewed socioeconomic factors. Although this was not intentional, organizations have a responsibility to avoid relying on ML models trained on imbalanced data—especially those that skew unfairly against protected classes of people. To achieve fairness in AI/ML, it's necessary to maintain ongoing efforts to monitor data for bias in models. AI fairness measures have been developed and incorporated into open source packages such as AI Fairness 360, Project Veritas, and fairML.

Figure 3. Fairness metrics from AI Fairness 360 demo[16]

13 https://www.sciencedirect.com/science/article/pii/S2352914819302047
14 https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00252-1/fulltext
15 https://www.ncri.org.uk/ai-to-spot-skin-cancer-lacking-pictures-of-darker-skin/
16 https://aif360.mybluemix.net/check

Building transparency

Transparency refers to the explainability of AI/ML; that is, the ability for people to understand how models arrive at their decisions.[17] Black box AI models, such as deep learning models, are the opposite of explainable: their operations are not visible to humans because the mathematics behind them are too complex to explain. Yet building transparency into AI/ML solutions is essential to breeding trust in the reliability—and fairness—of a model's predictions.

Visual tools used to explain ML models, such as Shapley plots, can be used to monitor risks and thereby increase transparency. Shapley values represent the impact of features on an ML model. These plots can be integrated with ModelOps and incorporated into a model evaluation or the ongoing monitoring of ML models.

Figure 4 shows a sample plot that was created using a dataset representing 10 years of clinical care data.

In addition, ModelOps provides the ability to report, monitor, and notify changes in the datasets and models, allowing for more efficient governance of ML models to include:

- Tracking model metrics over time for every ML model evaluation with new data
- Tracking data-level statistics for feature drift over time
- Configuring proactive alerts for feature and model drift

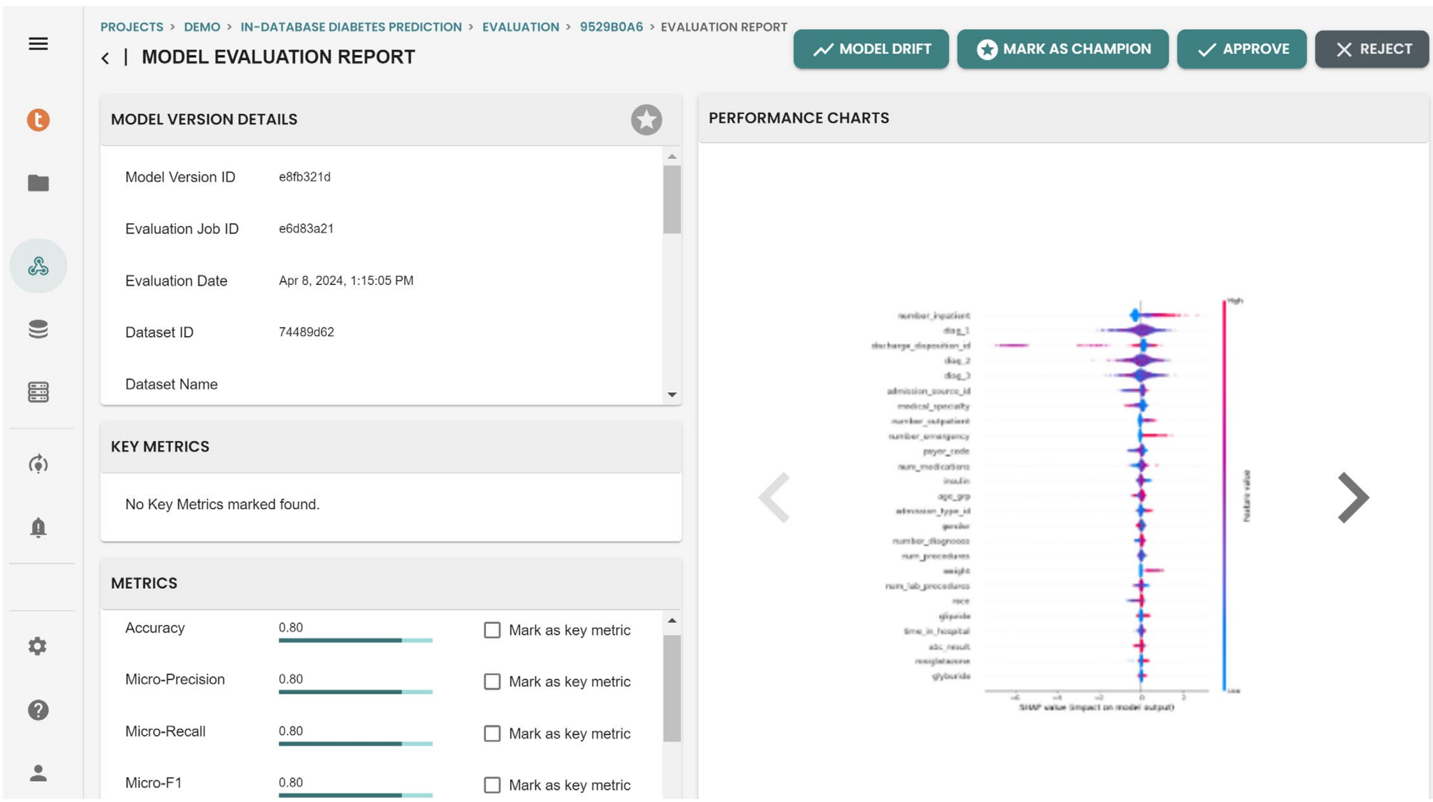


Figure 4. ModelOps with a SHAP value plot showing feature importance

17 https://www.analyticsvidhya.com/blog/2021/11/model-explainability/

Efficiency:
Optimizing responsible AI at scale

Efficiency addresses an organization's ability to operationalize and sustain ML models with horizontal scale; that is, the ability to manage many different models in production. Efficiency in AI is often hampered by a lack of continuity within and across data science teams, which often stems from the high turnover rates among data scientists. In organizations that lack governance, vacating data scientists leave room for their successors to adopt new practices. This inevitably creates additional incongruous workflows, model versioning, and uncoordinated processes, leaving AI in a fractured, unsustainable state. For example, progress on a specific ML model might stop when the author of its source code leaves the company. The lost code leaves orphaned ML models in production that can no longer be retrained or improved. This is not only an operational setback but a loss of value to the organization; ultimately, efficiencies are lost.

Stories like these can be found in most organizations whose data science teams are just getting started and even in organizations that have been using AI for years but lack a governance framework.

Teradata's ClearScape Analytics enables vertical (model size) scale through its in-database analytics and massively parallel processing (MPP). It also enables horizontal (many models) scale through the ModelOps extension. AI/ML sustainability is accomplished through operational practices supported by technology, while scale is accomplished through superior enabling technologies like ClearScape Analytics.

Data science and software engineering best practices in concert

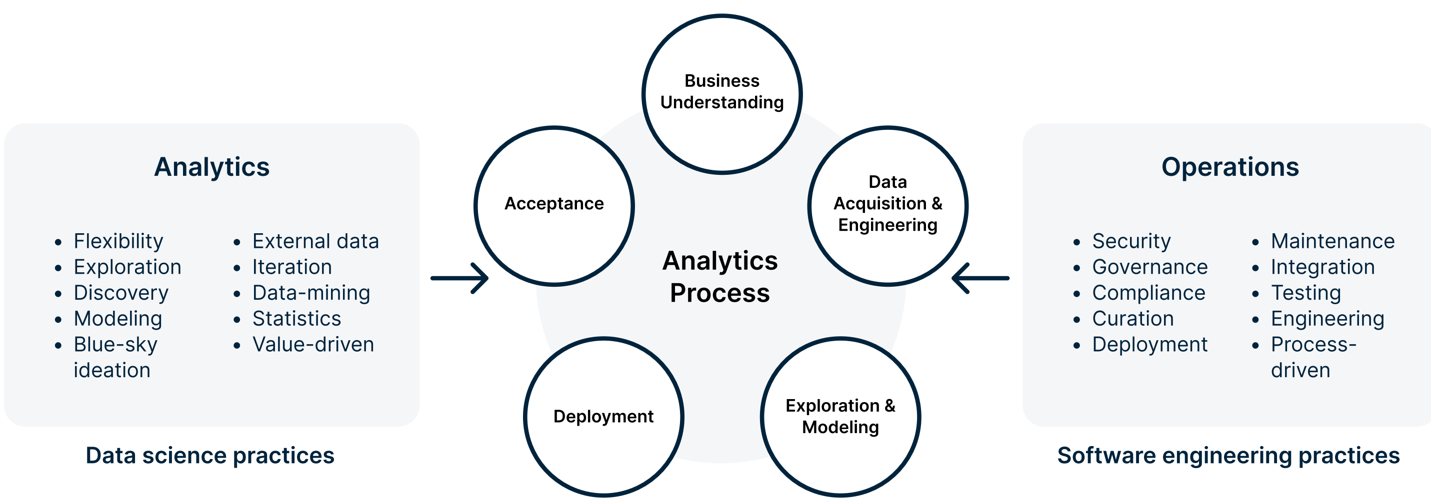


Figure 5. Data science and software engineering practices

Improving Operations

Data science is a creative, scientific discipline that leverages experimentation. Naturally, its practice does not lend itself to rigor in operations. Fortunately, the type of rigor required by organizations to achieve ML sustainability is not new. Organizations can implement practices similar to those leveraged by software engineering teams.

For example, they can implement more process-focused practices throughout the ML lifecycle. And they can use git repositories for code versioning and sharing. ModelOps enforces the use of git or git-like repositories. Models deployed through ModelOps must be sourced from a git repository that is set up during the project configuration, and transitions through the stages of the ML lifecycle are captured and logged.

To improve AI/ML operations, data scientists can borrow from practices used by software engineering teams and adapt them to their own needs.

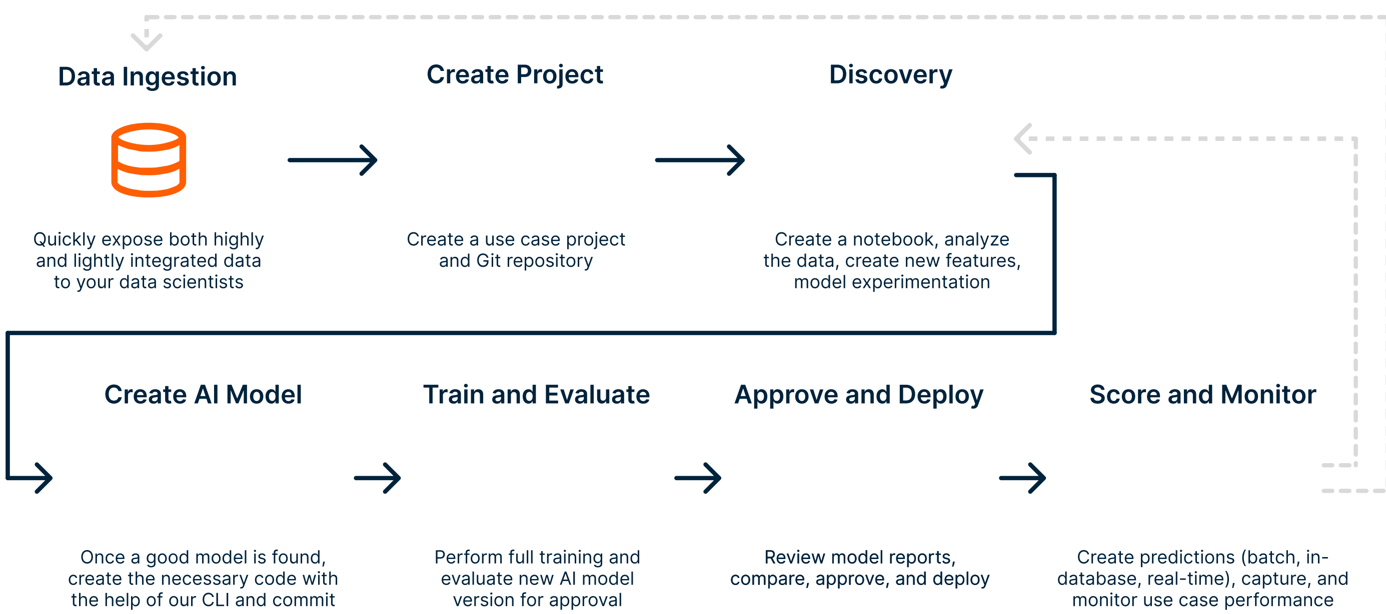


Figure 6. Project code workflow through ClearSpace Analytics and ModelOps

Creating Scale

For a platform to scale AI/ML well, it must handle the massive amounts of data and processing required by the ML algorithm. ClearScape Analytics provides the necessary features, as listed in Figure 7. The in-database functions and data pipelines of ClearScape Analytics bring the power of the MPP architecture to ML models to create vertical scale. Further, the expanded strategic partner integrations bring their innovative ML algorithms and engines to the ClearScape Analytics ecosystem. And the ModelOps extension brings horizontal scale.

ModelOps also offers the Feature Catalog, where you can define features and datasets within the ModelOps user interface. Through this, data scientists can enable data statistics for each job to get a snapshot of the data for every training, evaluation, and scoring at the feature level.

Native in-database function library	• 150+ advanced analytics functions to be run in-database, in parallel across industry use cases • Includes unbounded analytics for high-volume time-series and digital- signal processing
End-to-end in-database pipelines	• An expanded machine learning library that covers all stages of the workflow • Minimizing data movement through a new capabilities wrapper
Partner integrations	• Native analytic execution integrations with Amazon SageMaker, Azure Machine Learning, and Google Cloud Vertex AI • Bring-your-own-model (BYOM) support for PMML, ONNX, and H20.ai
ModelOps extension	• Enterprise-class model lifecycle management with Feature Catalog model deployment, model monitoring and more

Figure 7. Key ClearScape Analytics capabilities

In addition, as shown in Figure 8, the Vantage Enterprise Feature Store (EFS) creates efficiencies by democratizing ML model features for other data scientists and analysts across the organization, as features are often reused across ML models. With an EFS, production-worthy features are automated via ETL/ELT or DataOps processes for reuse among many models to gain “build once, use many times” efficiencies. The EFS enables time travel. This allows the database to be queried to see what data was used to train and score models at any point in time. In this scenario, Teradata Vantage enables the EFS.

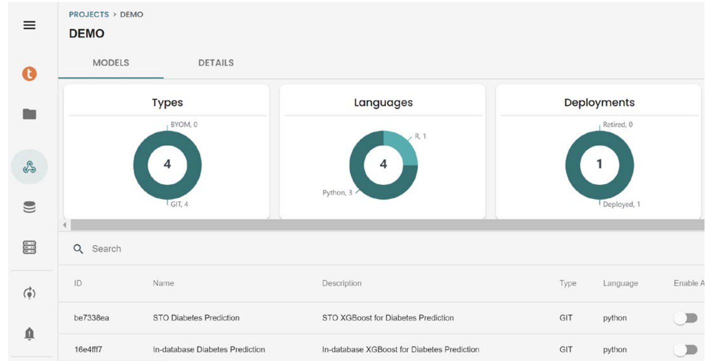
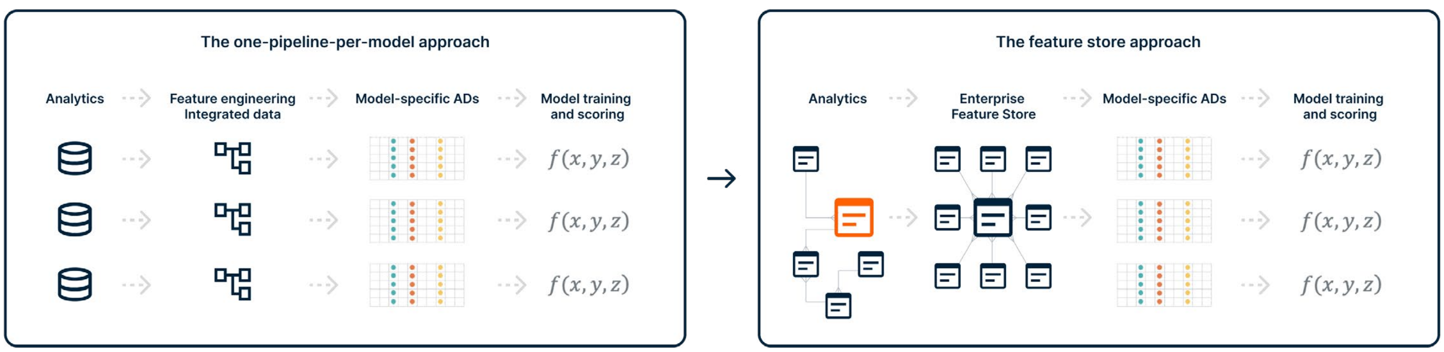


Figure 9. ML model status snapshots on ModelOps dashboard

ModelOps provides automated lifecycle management through its dashboard, enabling ML model deployment at tremendous horizontal scale. Data scientists can reduce the time they spend on operational tasks, such as evaluating and retraining models, freeing them to devote their energies to increasing business value through new AI/ML use cases.



One pipeline per model → Long data prep cycles and slow time to market

Redundant infrastructure, processing, and effort → High TCO

Limited reuse of pipeline or features → Poor productivity and data silos

Data scientists maintaining data silos → Inefficient allocation of resources

“Off the peg” features dramatically improve analytic cycle time and time to market.

Extensive reuse reduces TCO and improves analytic data quality and predictive model accuracy.

ADS layer enables model-specific customization while eliminating analytic data silos.

Separation of duties and improved productivity.

Figure 8. The Enterprise Feature Store enables “build once, use many times” for valuable ML model features

Conclusion

Enabling Responsible AI at Scale with Teradata VantageCloud and ClearScape Analytics.

Organizations can drive more growth, value, and performance with Teradata VantageCloud, the complete cloud analytics and data platform. With the holistic, end-to-end advanced analytic capabilities of ClearScape Analytics, VantageCloud provides the technology and tools necessary to support the implementation of responsible AI practices across the business. To support governance, the ModelOps extension provides auditing capabilities throughout the ML lifecycle. It further supports the software engineering best practices for code sharing and versioning to create efficiencies and avoid disruption of operations or workforce turnover.

The automation of model and data monitoring frees up data scientists' valuable time to focus on creating new AI/ML models, and it provides transparency and visibility of fairness metrics. Lastly, ClearScape Analytics empowers scale—both vertically through the powerful in-database analytics and MPP architecture of Vantage; and horizontally through ModelOps automation to increase efficiency of operations.

About Teradata

At Teradata, we believe that people thrive when empowered with trusted information. We offer the most complete cloud analytics and data platform for AI. By delivering harmonized data and trusted AI, we enable more confident decision-making, unlock faster innovation, and drive the impactful business results organizations need most. See how at **Teradata.com**.

About the author

Trish Lugtu is a Minneapolis-based principal data scientist on the advanced analytics team at Teradata. She has more than 20 years of experience working with data and analytics in the healthcare and life sciences industries.

17095 Via Del Campo, San Diego, CA 92127 Teradata.com