fiddler

TECHNICAL BRIEF

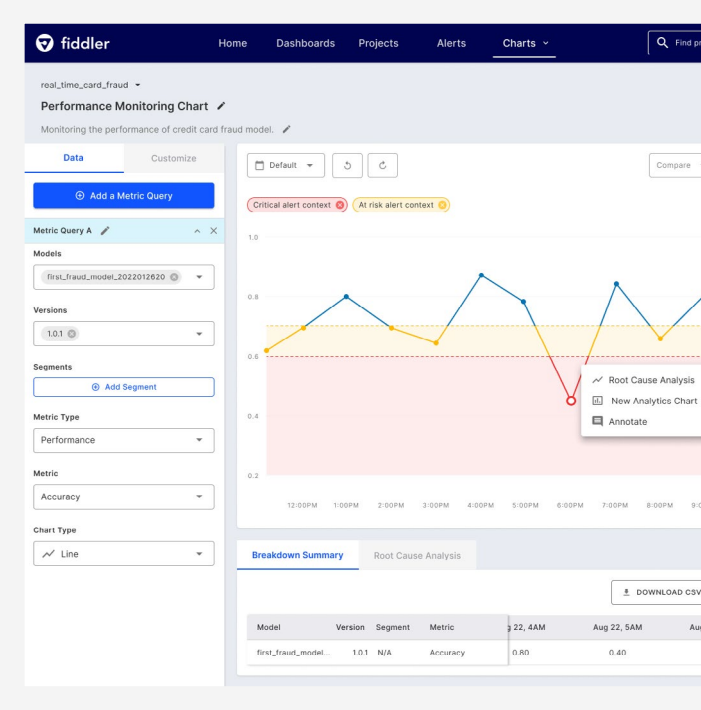
How Explainable AI Works in Fiddler

What Fiddler does

Fiddler is a pioneer in Model Performance Management (MPM) for responsible AI.

Fiddler empowers Data Science, MLOps, Risk, Compliance, Analytics, and LOB teams to validate, monitor, explain, analyze, and improve model performance.

The Fiddler Model Performance Management platform operationalizes ML/AI with trust by delivering a unified environment with centralized controls and actionable insights. As a result, teams can accelerate AI time-to-value and scale, build trusted AI models, and increase revenue by connecting predictions with context to business value.



What is MPM?

MPM serves as the centralized control system for ML workflows—tracking and monitoring model performance at every stage. Teams can go beyond metrics to explain machine learning results, thus creating a long-term framework for responsible AI.

Model Performance Management (MPM) standardizes MLOps practices throughout the lifecycle.

When powered by explainable AI (XAI), MPM becomes essential for model risk management, model governance, and optimizing MLOps.

Why explainable AI is important

ML models are increasingly complex opaque boxes making it hard to understand the how and why behind models' decisions, causing risks and impeding troubleshooting.

Explainable AI demonstrates the effect of a model's inputs on its predictions. Fiddler's proprietary XAI technology provides complete context and visibility into ML model behaviors and predictions, from training through production. XAI is critical for detecting bias in training data and skew within models.

Fiddler uses widely-adopted explainability techniques to increase model transparency, including Shapley Values (SHAP), Integrated Gradients, and Partial Dependence Plots—all of which are available as open source. Fiddler provides a proprietary, improved version of SHAP, which is backed by a peer-reviewed technical paper that received the best paper award.

Fiddler's explainability and fairness solutions work for structured (tabular) data, unstructured data such as natural language processing (NLP) and computer vision (CV), and for multimodal features; e.g: for both tabular and text features. Both solutions also work for all categories of classification, regression, and ranking models.

The Fiddler MPM Platform employs XAI to

Maximize Confidence

Provide explanations for all model predictions

Minimize Risk

Enable AI governance and model risk management processes

Increase Brand Loyalty

Delight customers with responsible AI

Explainable AI feature highlights

Shapley Values (SHAP) and Fiddler SHAP:
Increase model transparency and interpretability using SHAP values. Fiddler's proprietary SHAP helps interpret models quicker using approximations.

Integrated Gradients:
Comprehend how data features contribute to data skew and model predictions.

'What-If' Analysis:
Better understand model predictions by changing any values to see the impact on scenario outcomes.

NLP and CV Monitoring:
Increase prediction accuracy by monitoring complex and unstructured data, such as natural language processing and computer vision.

Global and Local Explanations:
Increase confidence in how selected features contribute to model predictions. Discover how each feature contributes to the model's predictions (global) and uncover the root cause of an individual issue (local).

Surrogate Models:
Improve the interpretability of models before they go into production by using surrogate models automatically built in Fiddler.

Custom Explanations:
Create customized explanations for specific use cases via APIs.

Supported frameworks & data types

Fiddler supports all major frameworks, such as PyTorch, TensorFlow, and scikit-learn, and provides deep, seamless integrations with different cloud-based ML platforms, including Amazon SageMaker, Microsoft Azure ML, and Google Vertex AI.

With Fiddler, data scientists and ML practitioners can set up explainability, fairness, and monitoring capabilities without writing a significant amount of code or tuning configuration parameters.

Model validation & model inference

Fiddler integrates explainability and fairness frameworks during both model validation (post training/testing) and model inference (post deployment).

The integration is performed by giving Fiddler access to the training data (treated as the baseline data), the trained model artifact, and the production logs, which consist of model inputs and outputs.

Model Validation

Fiddler helps:

1. Measure different types of biases exhibited by the model across different demographic groups and intersectional subgroups
2. Obtain the global importance of different features.

Model Inference

Fiddler monitors for:

1. Data integrity violations
2. Data drift (changes in the distribution of input features)
3. Prediction drift (changes in the distribution of model prediction scores)
4. Model performance degradation

Model fairness

Fiddler ensures predictions are fair by providing global and local explanations, counterfactual analysis, and bias measures with respect to demographic groups and intersectional subgroups.

Fiddler supports model-agnostic and model-specific explanations via integration with PyTorch Captum. In addition, Fiddler provides different bias metrics corresponding to legal and model governance requirements.

Counterfactual analysis

Understanding causal relationships is an important aspect of the root-cause analysis associated with ML model predictions.

Fiddler supports counterfactual analysis with “what-if” functionality, whereby individual features can be changed to understand the exact effect on the model’s predictions.

As a result, a deeper understanding is developed around how the (often opaque) model makes predictions based on the input features.

Actionable recommendations

Fiddler provides actionable insights and recommendations around model explainability and fairness, guiding ML practitioners to take specific actions such as:

1. Decide to retrain the model after obtaining a specific amount of additional training data from a specific sub-population.
2. Design new features that are most similar (or dissimilar) to the important features.
3. Only rely on the model predictions in certain specific settings but defer to human domain experts when the model is not sufficiently confident.

Explainable AI use cases

Enterprises can use XAI for a plethora of rich and hybrid use cases including but not limited to:

Healthcare:
Medical and healthcare practitioners can gain greater accuracy in diagnosing illnesses and provide timely customized patient care by recognizing patterns in medical images

Automobile:
The automobile industry can dramatically improve precision in autonomous vehicles by identifying the relative distance between moving and stationary objects to increase passenger safety

Government:
Defense agencies improve the precision in simulated missions by identifying concealed weapons and vehicles with NLP and CV monitoring

Retail:
eCommerce platforms can optimize customer experiences with an improved product recommendation engine with deeper personalizations related to colors, textures, and skin tones

Financial Services:
Financial institutions gain increased confidence and transparency in their model predictions by understanding why only a subset of applicants are granted loans

Manufacturing:
Manufacturing companies can minimize disruptions in assembly lines when they can determine why a process is defective on the manufacturing floor

Fiddler aligns stakeholders across the organization

The Fiddler MPM platform is designed to allow a variety of roles to understand how ML models make predictions and ensure predictions are fair. In a single platform, Fiddler enables collaboration and alignment between various teams throughout the model lifecycle.

Data Scientists:
Determine which features are most important for the model as a whole (global explanation) and for a specific prediction (local explanation). These insights help craft better features to improve the AI model and debug issues. Any biases exhibited by the model can be identified across different demographic groups (as well as intersections of groups across more than one attribute), and action can be taken to mitigate the biases

Customer Support Representatives:
Address customer complaints and inform product improvements by determining the features most important for the corresponding model prediction and performing counterfactual analysis using Fiddler's "what-if" functionality

Business Owners:
Decide if they trust the AI decisions by validating whether the most important features agree with the business domain knowledge

IT and MLOps Teams:
Monitor and debug ML models

Auditors, Compliance, and Risk Management Teams:
Validate whether ML model predictions are fair and compliant with model governance requirements

fiddler

Fiddler is a pioneer in **Model Performance Management** for responsible AI. The unified environment provides a common language, centralized controls, and actionable insights to operationalize ML/AI with trust. Model monitoring, explainable AI, analytics, and fairness capabilities address the unique challenges of building in-house stable and secure MLOps systems at scale.

Unlike observability solutions, Fiddler integrates deep XAI and analytics to help you grow into advanced capabilities over time and build a framework for responsible AI practices. Fortune 500 organizations use Fiddler across training and production models to accelerate AI time-to-value and scale, build trusted AI solutions, and increase revenue by improving predictions with context to business value.

fiddler.ai

sales@fiddler.ai

Request Demo