TAZI

TECHNICAL BRIEF · DATA SHEET · 2026

TAZI AI Platform

Business-in-the-Loop AI, Generative AI and Grounded Agents

SUMMARY

Agents and machine learning on one platform, sold and deployed together. The agents are grounded in your own predictive models and your own data — the platform trains and explains those models, then wraps them in agents that propose and validate actions. Underneath sits the full lifecycle: connect and profile data, engineer features in real time, train, explain, serve, govern and monitor in production. Installed on-premises or in the cloud with identical functionality, and able to run completely isolated from the internet. Three paths onto it — bring your own predictive models written in Python or R, deploy pre-built use cases, or build new ones — all landing on the same serving, monitoring and documentation infrastructure. 7 granted US patents on continuously learning, human-understandable online machine learning · 33 Gartner mentions, 2019–2025.

GROUNDED AGENTS · THE VALIDATED-ACTION LAYER

The agent layer sits on the models and data described in the rest of this sheet. Nothing an agent proposes reaches a queue until it has been grounded in your data, explained in your language and screened against your policy.

Grounding	Every agent reads the predictive model, the companion explanation model and your dynamic population of data — not a general-purpose corpus. Continuous learning and GEPA keep the models, explanations and agent prompts current without full retrains.
TAZI Embedded Agents	Generative models running inside a use case as pipeline steps — classifying, extracting, summarizing or drafting from your structured and unstructured data, with their output feeding the predictive models and the downstream agents. They run on open-weight models on your own intranet or on a commercial provider, under the same roles, groups, logging and data connections as the rest of the platform.
TAZI Decision Models	Deterministic decision models — JEV, LAYA and others — run the proposed-action and validated-action steps with no large language model involved: they read the prediction probability, choose the appropriate action from your action inventory and assign the owner it is routed to, under the same policy and thresholds. Where a language model is unavailable, not permitted or simply not wanted, they are the alternative to TAZI Embedded Agents for those two steps; the conversational, focus-group and expert-panel agents still use one.
Proposed-Action Agent	Drafts one action per case from the prediction, its reason code and the action inventory agreed for the use case.
Expert-Panel Agent	Applies your policy — pricing, regulatory, brand, operational constraint and the ROI threshold an action must clear — and discards proposals that fail it.
Focus-Group Agent	Stress-tests the surviving action against each persona or segment in the focus population before it is released.
Conversational Agent	Business users interrogate every agent in the pipeline in plain language; answers cite the model output and the named pattern behind it. Questions are tailored per use case, and limited-access field users see only their own population.
Solution Handoff Agent	Produces model, performance, fairness verdict and data-readiness score at go-live, as a signed-off handoff document.
Monitoring Agent	Reports health score, drift, live performance and fairness every month the agent is in production.
Agent interfaces	Model Context Protocol tools, an SDK, the workflow engine and ETL pipelines expose the platform to your own agents; a no-code user interface and AI Ops serve the humans. Approval gates can stop any action, and every step is logged.

DATA MANAGEMENT · CONNECTIONS

Every connection below is available for both model input and output.

Databases	Oracle · MS-SQL · PostgreSQL · BigQuery · SAP HANA · Presto · Incorta · Snowflake · Databricks · Cloudera Impala · Elasticsearch · Data iO
Files & object stores	File system · Hadoop HDFS · Amazon S3 · Google Cloud Storage · Microsoft Azure Blob Storage · web services and APIs over HTTP
Streaming	Kafka
Applications	Salesforce
File formats	CSV · Excel · JSON · Parquet

DATA MANAGEMENT · DATA ANALYSIS

Profiler — numeric	Relevance to the target based on mutual information · entropy · empty count · malformed count · unique values · minimum · maximum · mean · median · first and third quartile · outliers · mode · standard deviation · skewness · kurtosis · graphical histogram, per category where the data is labelled
Profiler — categorical	Relevance based on mutual information · entropy · empty count · malformed count · unique values · graphical distribution of values
Recommendations	One-click transformations proposed by the profiler: category weights for unbalanced data · features to ignore for low relevance · one-hot and label encoding · numerical transforms (logarithm, square root, square) where they raise relevance · numerical combinations of feature pairs (sum, difference, product, quotient, ratio) · concatenations of categorical pairs

© 2026, TAZI AI Systems Inc. All rights reserved. Do not copy or share without TAZI's approval.

DATA MANAGEMENT · DATA ENGINEERING

Multiple sources Join a secondary data source through primary keys to create rich features on the fly, with no intermediate data storage.

Cascading models The output of one model can be the data source for a downstream model; event-driven operation keeps solutions of several connected models running robustly.

ETL pipelines Extract, transform and load pipelines move data between the connected sources and the platform on a schedule or on an event, applying the same cleaning, joining and feature logic used at training time — so what is served is exactly what the model was trained on. Pipelines are built without code and reused across the models in a solution.

Real-time features Engineered features are created in real time without intermediate storage and without changes to the original data sources.

Feature type	Available feature-engineering functions
Any	history (time lag)
String · Boolean · Char	count_same, count_unique (over multiple windows)
Byte · Short · Int · Long · Float · Double	mean, stddev, min, max, skewness, kurtosis, normalized (over multiple windows)
Date	DayOfMonth, DayOfWeek, DayOfYear, DaysToEndOfMonth, MonthOfYear, WeekOfYear
DateTime · Timestamp	the Date functions plus HourOfDay, MinuteOfHour, AmPm
Time	HourOfDay, MinuteOfHour, AmPm
IPv4 · URI	IPv4: mask, count_same, count_unique (over multiple windows). URI: domain, host, path, suffix

Beyond the built-in functions, custom engineered features are created with complex mathematical and logical expressions through a drag-and-drop interface.

AutoML · ALGORITHMS

Built-in algorithms, plus mechanisms to bring your own algorithm or trained model written in Python or R. The online decision tree is a patented continuous-learning algorithm; the deep neural network is also configured for continuous learning. Ensembles are encouraged for robust production performance.

Instruments DSPy & GEPA · Hermes · deep neural network · bring your own Python and R

Classification Online decision tree · deep neural network · XGBoost · logistic regression · linear support vector classification · random forest · LightGBM

Regression Gaussian regression · deep neural network · XGBoost · generalized linear model · linear regression · random forest · LightGBM · linear support vector regression

Forecasting FB Prophet · ARIMA · exponential smoothing · Croston · Croston TSB

Clustering & anomaly K-Means for clustering · Isolation Forest for unsupervised anomaly detection

AutoML · METRICS, EXPLANATIONS AND THE HUMAN IN THE LOOP

Classification metrics Confusion matrix · F1 score · accuracy · weighted accuracy · equally-weighted accuracy · precision · recall · Top-K · ROC curve · lift chart

Regression metrics Mean squared error · normalized mean squared error · mean absolute error · normalized mean absolute error · mean absolute percentage error · symmetric mean absolute percentage error · R-squared · adjusted R-squared

Business value Where a business KPI feature is specified in the data schema, the business value corresponding to each machine-learning metric is calculated alongside it.

Explainable AI Classification — an online decision-tree surrogate model explains every output instance, with a sunburst visualization of the surrogate for overall model explanation and a confusion matrix of the surrogate against the underlying model. Regression — LIME provides localized explanations. SHAP values give per-feature attribution for an individual prediction and for the model as a whole, across both problem types.

Human in the loop Subject-matter experts give feedback on a segment basis, approving or changing the model prediction on a simple graphical interface; that feedback is used to tune the model.

Dashboards & tooling A business-user dashboard is created automatically for every model and updated as output is generated — configurable, scriptable and interactive, with default layouts per problem type. Hypo tunes hyperparameters with an evolutionary algorithm. TAZI Lab provides private or shared Jupyter notebooks with access to platform data sources, and notebook code deploys as part of a model. Shareable Python environments keep package versions consistent across the model lifecycle.

GENERATIVE AI

Supported models Open-weight models deployed on your own intranet, and commercial providers such as OpenAI, Azure AI, Google and Anthropic — allowing complete on-premises operation. Access is controlled by the platform's roles and groups, and the uniform interface changes the underlying model without affecting business-model settings.

Data connections Generative models use the same data sources and destinations as the rest of the platform. Prompts are built from selected fields of your structured sources, and the output of other business models can be used inside the prompts.

© 2026, TAZI AI Systems Inc. All rights reserved. Do not copy or share without TAZI's approval.

MLOps

Workflow engine	Multi-step jobs are composed and orchestrated as one runnable pipeline — ingestion, ETL, profiling, training, scoring, business rules and agent calls chained together, started on a schedule or by an event, with every step's state visible and notifiable. Agents and the SDK invoke workflows as tools.
Business rules	Applied in real time to model output, created in a no-code environment, chained together and shared within a solution.
Export & import	Models export from one TAZI installation and import into another, or export as self-contained container images that run anywhere.
Labels & tags	Three predefined deployment labels — DEVELOPMENT, STAGING, PRODUCTION — renamable per organization, plus user-created tags for filtering which models appear in the interface.
Notifications	Subscribe to events such as a failed model run, completed training or disk-space pressure. Delivered by browser notification, Slack channel, Microsoft Teams or email over SMTP, to an individual user, a group, an organization or every user on the installation.
Custom services	Any Python application that complies with the prescribed input/output specification deploys on the platform as a RESTful service, so complete business applications can run with AI models as components.
Resource management	An administrator can cap the number of active models on a server; models beyond the cap queue and wait for running models to stop or sleep. A modestly sized server can then work through dozens of overnight model runs unattended.
Event-triggered operation	With a database or streaming input the model runs continuously, detecting and processing new instances. After a configurable idle interval it sleeps, unloading state to disk to reduce its memory footprint, and wakes when new data arrives — more robust than schedulers when several business systems must work in tandem.
Scheduled operation	A model can run once or repeatedly, down to minute granularity.
REST API for models	Enabled for any model by selecting one of its saved states as the serving model. One installation URL serves every RESTful model, each reached by a unique key; key management swaps models behind the API without changing callers. Heavy traffic is served by multiple load-balanced instances.

GOVERNANCE & SECURITY

Authentication	Configurable password complexity and two-factor authentication, or your existing identity-management system such as Active Directory via SAML 2.0, OIDC and LDAP.
Authorization	Controlled by role assignments and group memberships.
Roles	A user holds exactly one role. **SystemAdmin** — system administration only, no data-science tasks; does not count against license restrictions. **Admin** — SystemAdmin plus full data-science access including solution creation. **Data Scientist** — full data-science access within their groups; cannot create solutions. **BusinessUser** — as Data Scientist but cannot run Hypo. **User** — views the machine-learning and business dashboards in their groups. **Profiler User** — runs the profiler only.
Groups	Created by Admin users; a user can belong to several. Solutions and business models belong to a group and are visible only to its members.
Reports	Business model reports — a downloadable PDF with creation date, data source and training performance metrics. Data profiler reports — a PDF of all feature statistics, tables and graphs. Audit reports — platform activity logged to an audit table that Admin users filter by date and time, user name and activity type.

INSTALLATION

Deployment	Single server or Kubernetes, on-premises or in the cloud, with identical functionality; the platform can be completely isolated from the internet.
Server requirements	RedHat variants or Ubuntu. CPU, RAM and disk sizing follow your workload, with a minimum of 64 GB RAM and 30 GB disk for the software; a separate disk for operations data is recommended. No GPU required. The server can be resized later without loss of data, and the account team provides a detailed requirements document.
Admin panel	A web interface for SystemAdmin users to monitor and restart platform services, observe resource utilization of running model processes and set environment variables. Model internals — input, output and logs — are not accessible here, for privacy and security.
Updates	Online from the TAZI distribution servers, or offline packages downloaded to an intermediate computer and uploaded through the Admin Panel for isolated installations. Running models are not impacted by platform updates.
Self-test	A module that runs a fixed set of scenarios emulating user behaviour, extensible per installation, intended to run after each update.

WHAT ONE USE CASE CONTAINS

Up to 5 models — prediction, explanation, segmentation — and up to 5 agents: proposed action, validated action, focus group, expert panel and conversational. **You provide** your data, your large-language-model API keys and their usage, and the server, storage and network infrastructure. **TAZI provides** the platform, onboarding, configuration, monitoring and updates.

EASY TO TRY, EASY TO BUY

Engagements start with a scoped **Proof of Value** on your anonymized data, run on the TAZI server against what to predict and the value created by actions, agreed before it begins. Onboarding, configuration, monitoring, drift detection, solution updates, support and business reviews sit inside a single license you control — Start, Grow and Scale packages sized to your use cases and users.

TAZI

Technical deep-dive or demo: info@tazi.ai
San Francisco (HQ) · Istanbul (R&D) · Charlotte (East Coast) · Boston (Northeast)

© 2026, TAZI AI Systems Inc. All rights reserved. Do not copy or share without TAZI's approval.