



Your AI Powered Engine For Trusted, High-Quality & AI Ready Data

Transforming Data into
Intelligence

- Presented by VE3



THE DATA DILEMMA



AI Can't Work Without Clean, Connected Data

Poor data is still the silent killer of digital transformation. While enterprises race to adopt AI, automation, and analytics, the real bottleneck is broken, mismatched, and incomplete data.

- Spreadsheets, CRMs, ERPs, and documents all speak different languages.
- Critical records remain duplicated, unmatched, and misaligned.
- Data teams spend more time fixing data than using it.

The Outcome? Delays, wrong decisions, wasted investments, and AI that fails to deliver.

You can't build intelligence on chaos. You need foundational trust in your data — clean, matched, enriched, and understood.

The Stats

84%

of digital transformation initiatives fail due to poor data quality.

40-60%

of data teams time is spent cleaning and preparing data.

\$15M/year

is the average cost of poor data to enterprises (Gartner).

80%

of AI project failures stem from poor quality data, not model performance.

Manual data wrangling is not scalable. Quality, matching, and enrichment need to be automated — with intelligence built-in.

You're not alone. 80% of AI project time is spent fixing data. But it doesn't have to be this way.

MEET MATCHX?

Your End-to-End AI-Powered Data Matching & Quality Platform

MatchX simplifies, accelerates, and elevates every stage of your data pipeline. It's not just another data tool. MatchX is an intelligent, no-code platform that automates your entire data workflow - from ingestion to matching, from quality scores to document reconciliation — everything is AI-driven, explainable, and configurable. Whether you're working with CRM exports, finance records, PDFs, or millions of supplier rows — MatchX transforms your data into trusted, AI-ready assets.

Designed for:

- ▶ Enterprises drowning in fragmented records
- ▶ Data teams burdened by cleanup and matching
- ▶ Analysts and business users needing trusted data

Built for:

- ▶ Multi-source data ingestion
- ▶ AI-led matching, scoring, and profiling
- ▶ End-to-end lineage, governance, and dashboards

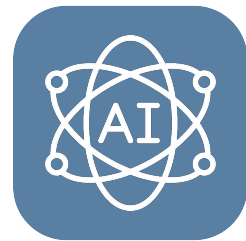
MatchX is how your data gets better faster, smarter, and continuously.

Finally, a data engine that doesn't just store or move data it understands it.

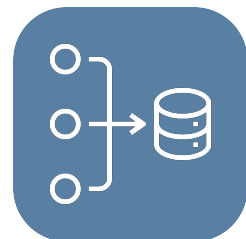
VE3

WHAT MAKES MATCHX DIFFERENT?

AI at Every Step: From rule creation to matching logic to scoring



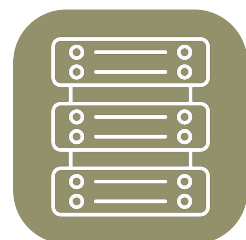
AI at Every Step: From rule creation to matching logic to scoring



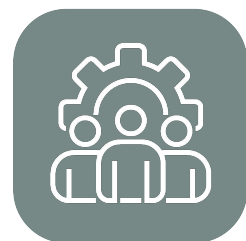
Works Across Sources: Spreadsheets, CRMs, ERPs, documents — even messy ones



No Code Needed: Just describe what you want. Let the engine do the rest



Built for Scale: Handles 1,000 records or 100 million — same ease



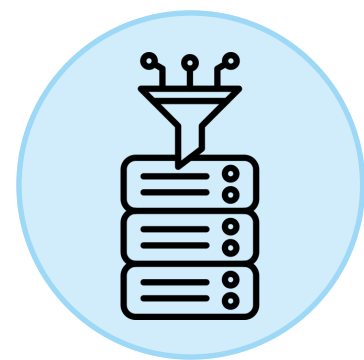
Designed for Your Team: Data, Ops, Compliance, Product, or AI — everyone benefits

Just upload your data.
Tell MatchX what you
want. It handles the rest
fast, accurate, and
beautifully simple.

WORKFLOW

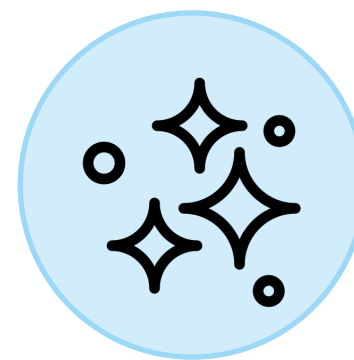


MatchX automates the entire data transformation process from ingestion to high-quality, AI-ready data ensuring accuracy, consistency, and compliance. With our AI-powered workflow, anyone can effortlessly manage data, as our intelligent automation simplifies every step, making data quality, matching, and governance seamless and intuitive.



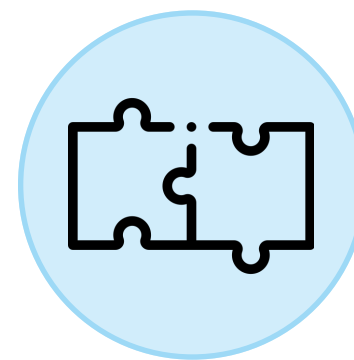
Ingest & Understand

- Multi-Source Data Ingestion
- AI-Based Profiling
- AI-Based Data Quality & Confidence Scores



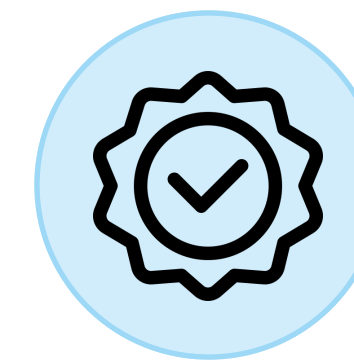
Prepare & Clean

- AI-Based Rule Library
- Automated Cleaning, Correction & Validation Engine



Match & Link

- AI-Driven Matching Engine
- Multi-Matching Type Support
- AI-Based Document & Paragraph Matching



Govern, Approve & Refresh

- Lineage Tracking & Approval Workflows
- Data Refresh & Linking



Deliver Insights

- Comparison Engine
- Live Dashboards & Data Health Insights

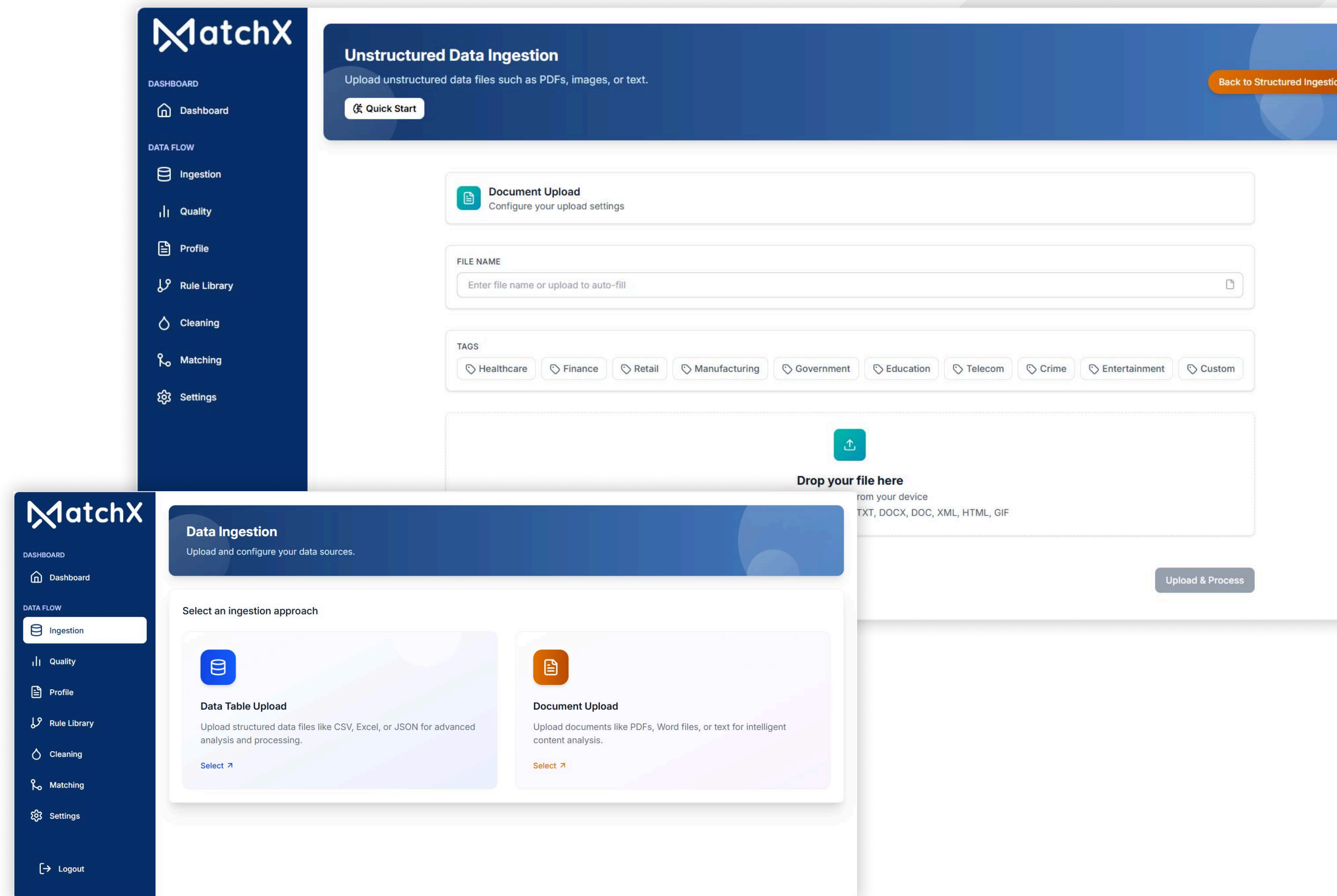


01. MULTI-SOURCE DATA INGESTION

Bring Every Data Type Together, Seamlessly

- Upload from spreadsheets, databases, PDFs, images, APIs
- Support for CSV, XLSX, JSON, XML, and document formats
- Connectors for S3, MongoDB, Postgres, and more
- Real-time streaming or scheduled batch uploads
- Schema mapping and auto-validation at ingestion
- Upload structured tables and unstructured documents — side by side

MatchX is built for multi-format, multi-system data, brought together into a single intelligent pipeline.



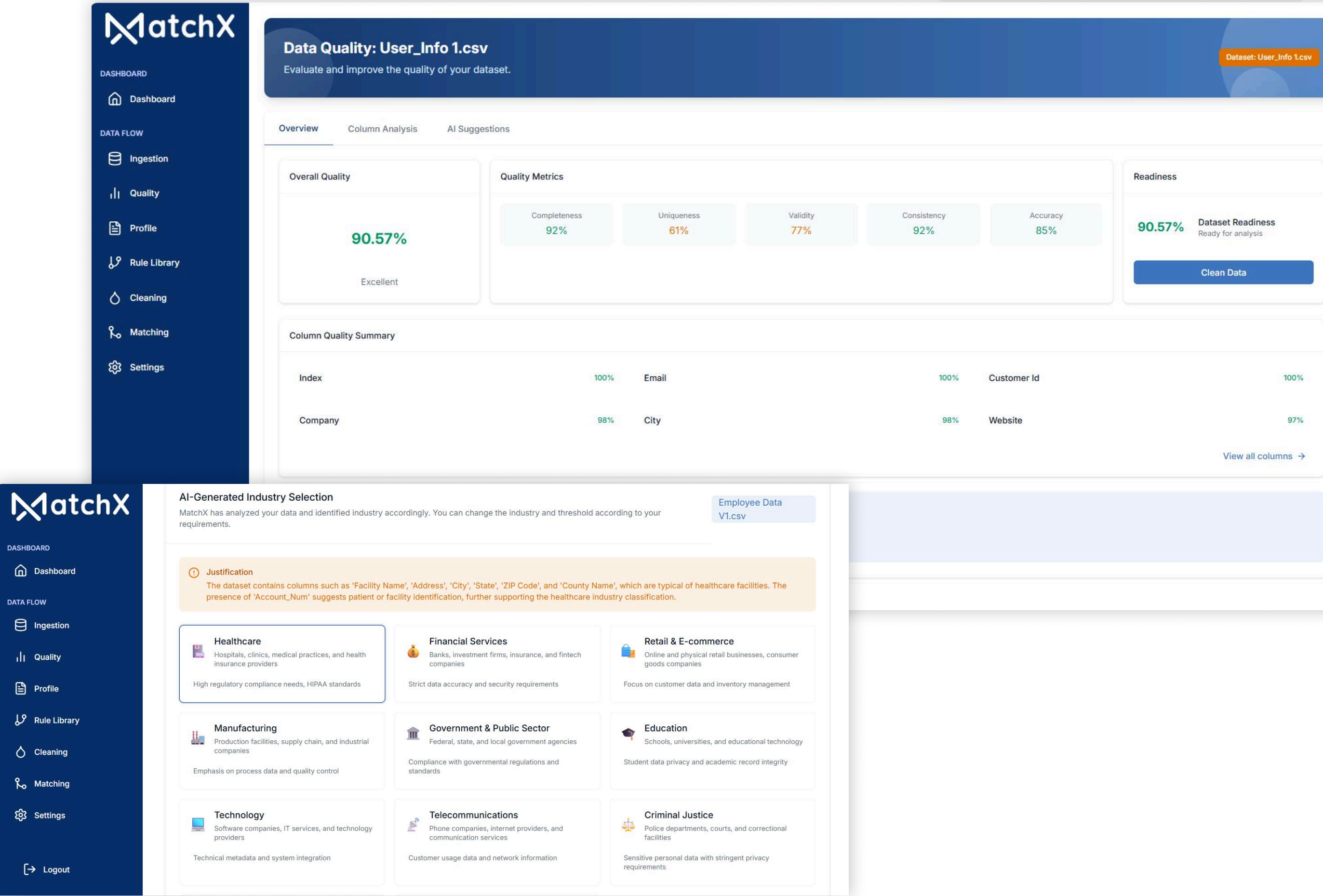


02. AI-BASED DATA QUALITY & CONFIDENCE SCORES

Know Your Data Quality — Before You Use It

- AI calculates quality scores at dataset and column level
- Evaluates completeness, accuracy, consistency, uniqueness, and timeliness
- Auto-detects domain-specific thresholds (e.g., finance vs healthcare)
- Generates confidence scores to guide approvals and reviews
- Visual highlights of low-quality records and high-risk attributes

Use MatchX scores to accept, reject, or fix records — all backed by explainable AI.



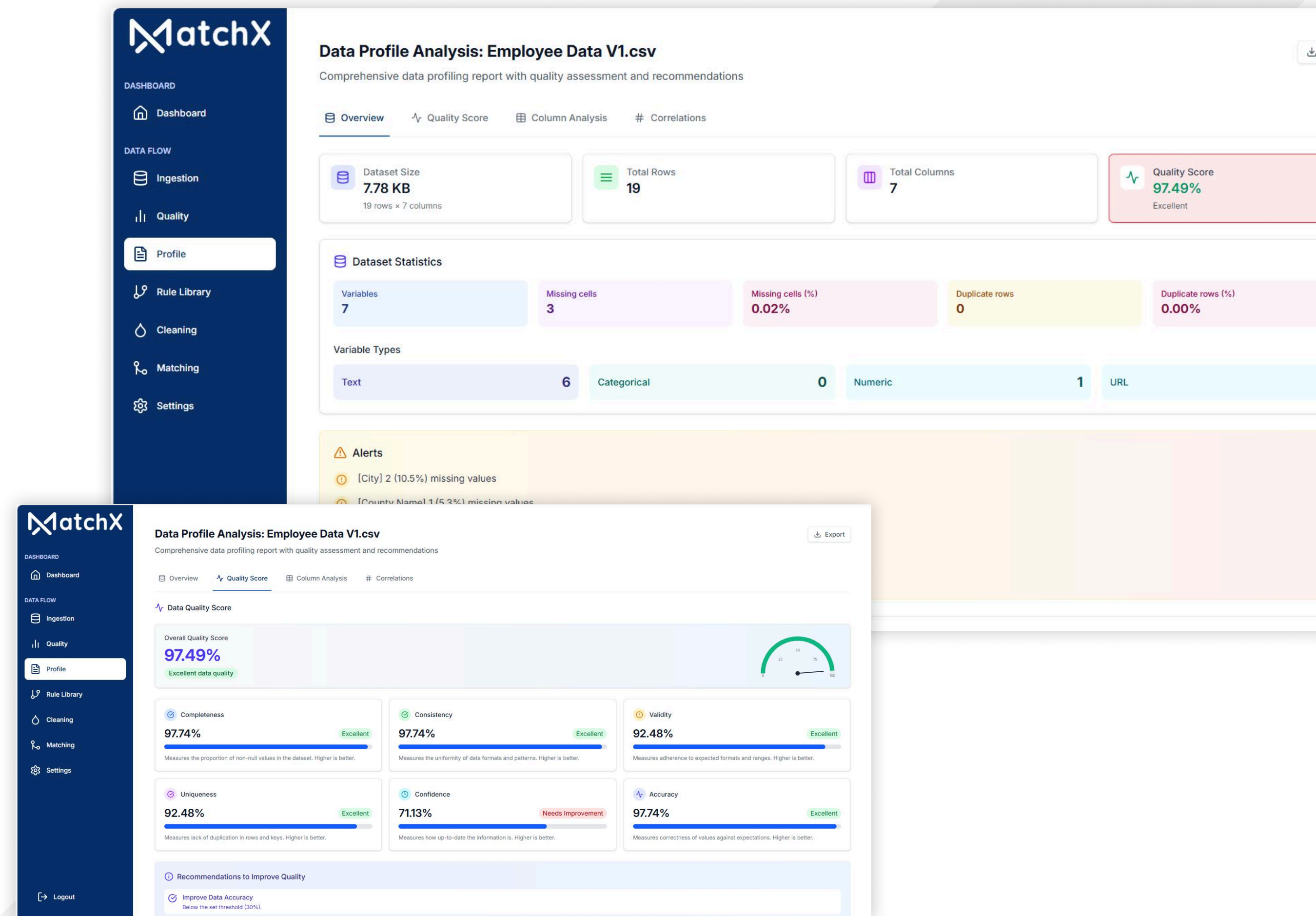
03. AI-BASED SMART PROFILING



Reveal the Story Behind the Data

- Profile data fields to uncover anomalies, outliers, and patterns
- Understand data distribution, ranges, missing values, and duplicates
- OCR and NLP-enabled document metadata profiling
- Filter and sort documents or fields by confidence, size, type, or quality
- AI models highlight trends and issues automatically

Turn unknown datasets into known entities — with clear profiling, instantly.



04. AI-BASED RULE LIBRARY & PROMPT BASED GENERATION



Rules That Write Themselves (Almost)

- Prebuilt rules for cleaning, validation, and transformation
- Natural language prompt-based custom rule generation
- Apply column-level rules dynamically across workflows
- Validate with live previews and rollback capability
- Role-based access to rule editing and publishing

Anyone — not just coders — can now define how their data should behave.

The MatchX interface displays a 'Data Rule Library' with a sidebar navigation menu. The main area shows 'Rule Categories' including All Rules (41), Validation (1), Profiling (12), and Transformation (4). The 'AI Rule Generator' section allows users to describe rules in natural language, with an example: 'e.g., Validate email formats in the customer_data table...'. Below this, several prebuilt rules are listed, such as 'Trim Leading ...', 'Trim Trailing ...', 'Remove Comma ...', 'Remove Apos...', and 'Split by Separators'. A 'Rule Preview' window is open, showing a table of data and the details of a 'Remove Special Characters' rule.

MatchX

DASHBOARD

- Dashboard

DATA FLOW

- Ingestion
- Quality
- Profile
- Rule Library
- Cleaning
- Matching

Data Rule Library

Create, manage, and apply data quality rules to ensure data integrity

Rules Workflows

Rule Categories

- All Rules (41) View all rule types
- Validation (1) Check data quality and business constraints
- Profiling (12) Understand data characteristics
- Transformation (4)

AI Rule Generator

Describe your rule in natural language

e.g., Validate email formats in the customer_data table...

Generate Rule

Trim Leading ... Published
Trims spaces from the beginning of text values.

Trim Trailing ... Published
Trims spaces from the end of text values.

Remove Comma ... Published
Identifies and removes unnecessary commas in text.

Remove Apos... Published
Identifies and removes unnecessary apostrophes.

Split by Separ... Published
Splits strings based on delimiters like @

Rule Preview

View the Rule details and modify the rules

Back to Rule Library

Preview Results

0	1	2	3	4	5
httpsrpyclubg9wvWF	None	None	None	None	None
None	None	None	None	None	None
Paid Telegram Chan...	Sr. No	Name	Email	Company	CTC
Click Here	1	Devendra Vilas Kolhe	Hidden	_urus	9.8LPA
Click Here	2	Ubale Ajay Ambadas	Hidden	_bixcash	6.4LPA
Click Here	3	Adesh Prashant Vairag	Hidden	_bixcash	3LPA
Click Here	4	Dudhnath M Shivank	Hidden	_hinkitive	3LPA
Click Here	5	Kalpesh Sanjay Patil	Hidden	_ockwell Automation	6.25LP
Click Here	6	Kadam Ganesh Digar	Hidden	_onata	4LPA
Click Here	7	Pooja Chandrakant Sa	Hidden	_onata	4LPA

Preview Settings

Dataset: Placement list

Columns: 2, 3, 0

Rule Information

Name: Remove Special Characters

Category: Cleaning

Status: Published

Description: Detects and removes unwanted special characters.

05. AUTOMATED CLEANSING, CORRECTION & VALIDATION



Fix What's Broken. Automatically.

- Auto-detect and correct errors like format mismatches, nulls, or typos
- Preview corrections, accept/reject at field level
- Apply standardization, enrichment, and cleanup at scale
- Correct via AI-driven rule execution or manual review
- Confidence-scored suggestions for every correction

With MatchX, data hygiene is no longer a manual chore.

The MatchX DataCleaner interface is a comprehensive tool for data cleansing and quality improvement. It features a sidebar with navigation options: Dashboard, Ingestion, Quality, Profile, Rule Library, Cleaning (selected), Matching, and Settings. The main dashboard displays a 'Data Quality Dashboard' for the 'Employee Data V1.csv' dataset, showing a 'Good quality data ready for analysis' status with a 97.49% score. Key metrics include Uniqueness (92.48%), Validity (92.48%), Accuracy (97.74%), Completeness (97.74%), and Consistency (97.74%). The 'Column Analysis' section provides a detailed view of data quality for specific columns, such as 'Account_Num', 'Facility Name', 'Address', and 'City', each with a 'Good' status and a 100% quality score. The 'Column Details' section further breaks down these metrics, showing 'Quality Score' at 100%, 'Confidence' at 70%, 'Completeness' at 100%, 'Consistency' at 100%, and 'Validity' at 100%. The 'Data Preview' section displays a table of data rows, including columns for Account_Num, Facility Name, Address, City, State, ZIP Code, and County. The 'Applied Rules' section lists various rules for data cleansing, such as 'Duplicates', 'Missing Values', 'Whitespace', 'Categorical Values', 'Email Format', and 'Data Types', each with a 'Fix All' button.

Account_Num	Facility Name	Address	City	State	ZIP Code	County
10605	SAGE MEMORIAL HOSPITAL	STATE ROUTE 264 SOUTH 191		AZ	0.8505	
24250	WOODRIDGE BEHAVIORAL CENTER	600 NORTH 7TH street	WEST MEMPHIS	AR	72301	CR
10341	DOUGLAS, HOSPITAL	5200 NE 2ND AVE	MIA~MI	FL	33137	MIA
81101	SUNCOAST BEHAVIORAL HEALTH CENTER	4480 S1ST ST W	BRADENTON	FL	3421	MAN
39835	TREASURE VALLEY HOSPITAL	8800 WEST EMERALD STREET	BOISE	ID	83704	ADA
48460	INTERMOUNTAIN, ...	303 NORTH ALUMBAUGH STREET	BOISE	ID	83704	ADA
30098	CHOATE MENTAL HEALTH & DEVELOPMENT CTR	1000 N MAIN ST	ANNA	IL	62906	UNH
92639	GOODLAND REGIONAL MEDICAL CENTER	220 WEST SECOND STREET	GOODLAND	KS	67735	SHE
86448	SPRING H~ARBOR HOSPITAL	123 ANDOVER ROAD		ME	4092	CUK
28455	EASTERN SHORE, HOSPITAL CENTER	POST OFFICE BOX 800	CAMBRIDGE	MD	21613	DOF
79639	CHOCTAW REGIONAL MEDICAL CENTER	8613 MS HWY 12	ACKERMAN	MS	39735	CHC
58893	ST LOUIS CHILDRENS HOSPITAL	ONE CHILDRENS PLACE	SAINT LOUIS	MO	63110	ST. L

06. AI-POWERED MATCHING ENGINE



One Platform. All Matching Types.

- AI-assisted configuration suggests best matching approach
- Auto-selects matching columns and algorithms (fuzzy, phonetic, etc.)
- Threshold-based confidence scoring with explainability
- Smart matching for non-Roman scripts and alphanumerics

MatchX delivers high precision — from people names to vendor IDs.

07. MULTI-MATCHING TYPES & TECHNIQUES



Because One Match Logic Doesn't Fit All

- Exact match, fuzzy match, Jaro-Winkler, TF-IDF, Phonetic matching
- Probabilistic record linkage with Splink integration
- Hybrid algorithms tuned by AI based on dataset structure
- Adjustable thresholds for match, review, or rejection
- Document and metadata-based matching for unstructured files

Every matching scenario — covered.

The screenshot displays the MatchX web application interface. On the left is a dark blue sidebar with the MatchX logo and navigation links: Dashboard, Ingestion, Quality, Profile, Rule Library, Cleaning, Matching (highlighted), and Settings. A Logout link is at the bottom. The main panel is titled 'Column Matching' and includes a help link. It features a light blue instruction box and a yellow warning box stating 'No columns found in source dataset. Please check your data.' Below this is a table titled 'Added Column Pairs' with columns for SOURCE, TARGET, ALGORITHM, and THRESHOLD. The table lists six column pairs (Account_Num, Facility Name, Address, City, State, ZIP Code) all using the 'Jaro-Winkler (best for names)' algorithm. Each row has a slider and a text input for the threshold, all set to 70%.

SOURCE	TARGET	ALGORITHM	THRESHOLD
Account_Num	Account_Num	Jaro-Winkler (best for names)	Threshold for Account_Num: 70% 70 %
Facility Name	Facility Name	Jaro-Winkler (best for names)	Threshold for Facility Name: 70% 70 %
Address	Address	Jaro-Winkler (best for names)	Threshold for Address: 70% 70 %
City	City	Jaro-Winkler (best for names)	Threshold for City: 70% 70 %
State	State	Jaro-Winkler (best for names)	Threshold for State: 70% 70 %
ZIP Code	ZIP Code	Jaro-Winkler (best for names)	Threshold for ZIP Code: 70% 70 %

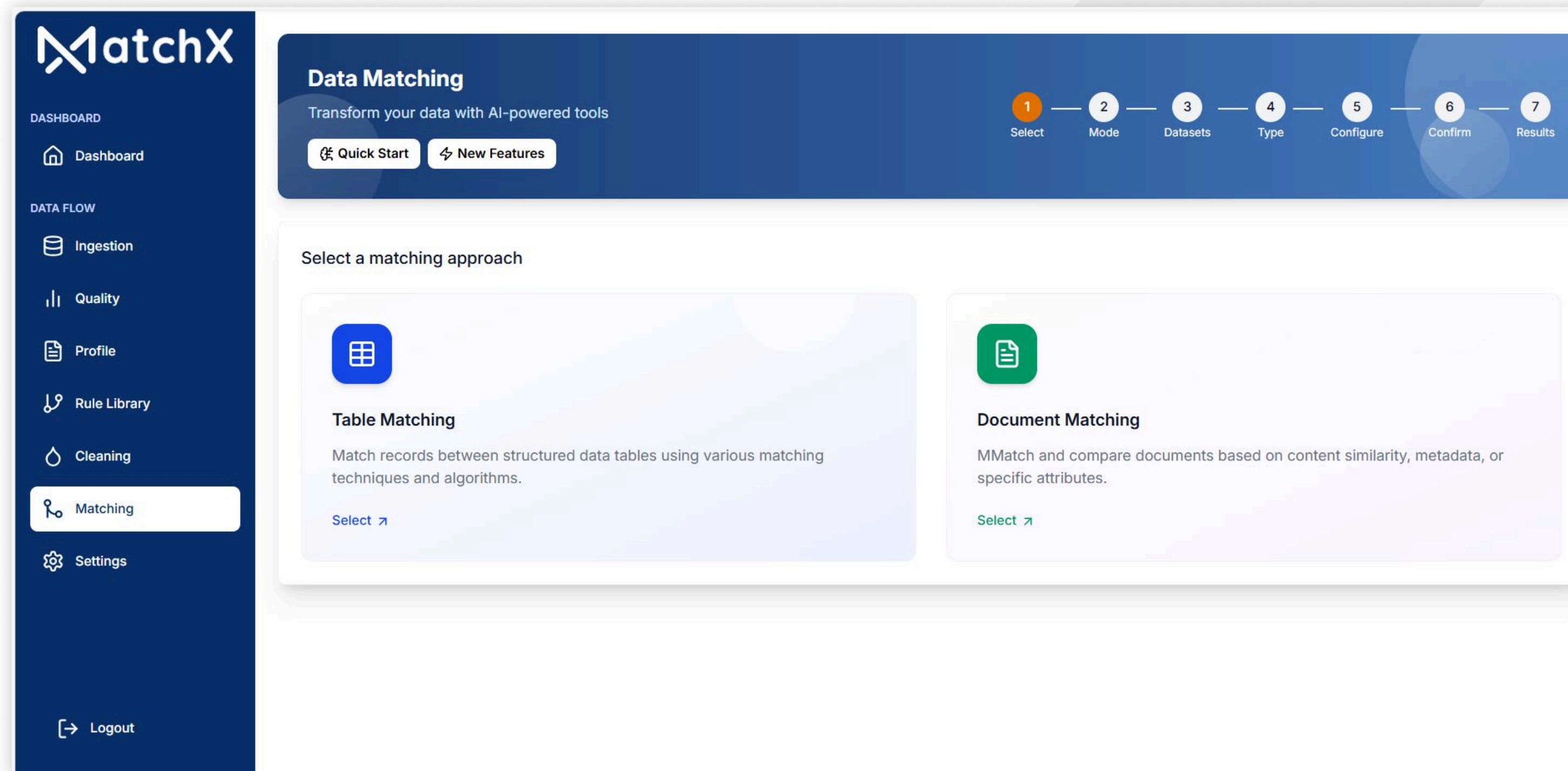
08. DOCUMENT & PARAGRAPH MATCHING



Not Just Rows. Paragraphs. Pages. PDFs.

- Compare invoices, contracts, or scanned PDFs line by line
- Use NLP and cosine similarity to detect near-matches in content
- Detect paragraph-level differences, overlaps, and duplications
- AI-powered OCR extracts content from scanned documents
- Match structured data with associated documents seamlessly

No document mismatch goes undetected.



09. DATA REFRESH, LINKAGE & RELATIONSHIP DETECTION



Your Data Is Always Changing. So Is MatchX.

- ▶ Auto-refresh datasets and re-trigger workflows
- ▶ Detect new, removed, or modified records
- ▶ Track linkage changes and log every action
- ▶ Discover relationships between datasets through smart linkage
- ▶ Manual + auto linking based on thresholds

Data isn't static. MatchX ensures your relationships stay relevant.

The screenshot displays the MatchX web application interface. On the left is a dark blue sidebar with the MatchX logo at the top. Below the logo, the sidebar is divided into two sections: 'DASHBOARD' with a 'Dashboard' link, and 'DATA FLOW' with links for 'Ingestion', 'Quality', 'Profile', 'Rule Library', 'Cleaning', 'Matching', and 'Settings'. At the bottom of the sidebar is a 'Logout' link. The main content area is titled 'Dataset Refresh' with a subtitle 'Review and apply dataset changes' and a 'Back' button. It features a card for 'Employee Data V1.csv' showing it was uploaded on 18/6/2025 at 5:12:57 pm, with tags for 'CSV' and 'Local Connection'. Below this card is a warning box titled 'Local Dataset Limitation' with the message: 'Local datasets cannot be refreshed. Upload a new version or connect to an external source.'

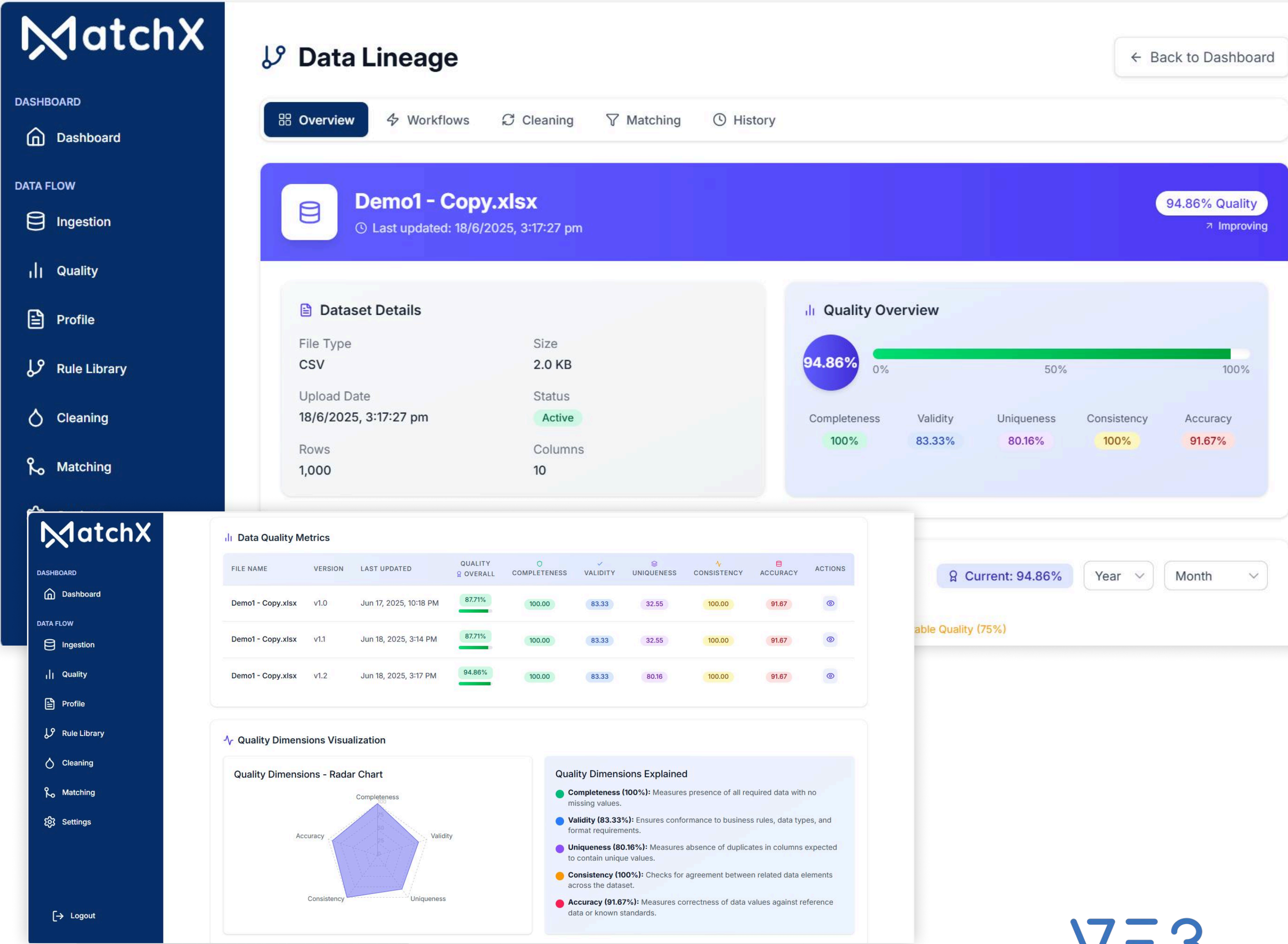
10. LINEAGE TRACKING & APPROVAL WORKFLOWS



Trust Requires Transparency

- Full audit trails of every change and decision
- Role-based approvals for rule execution or record acceptance
- View data lineage from source to current state
- Cell-level, rule-by-rule rollback
- Approval triggers for threshold breaches or match confidence drops

Every decision is explainable. Every action, traceable.



11. DASHBOARDS & DATA HEALTH INSIGHTS



From Metrics to Meaning

- Role-specific dashboards for data teams, compliance, or execs
- Data health scores, match accuracy, issue hotspots
- Filter by domain, project, status, or confidence
- Real-time alerts on anomalies or threshold breaches

You'll know the moment your data quality drops — and where.

MatchX

DASHBOARD

Dashboard

DATA FLOW

Ingestion

Quality

Profile

Rule Library

Cleaning

Matching

Settings

Logout

Data Lineage

Overview

Workflows

Cleaning

Matching

History

Back to Dashboard

Data Cleaning Operations

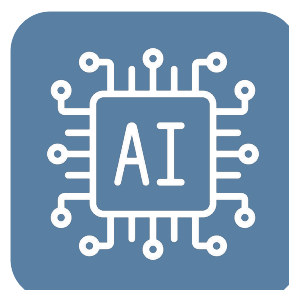
New Cleaning Job

FILE NAME	CLEANING TYPE	MODE	RULES APPLIED	PREVIOUS SCORE	CURRENT SCORE	QUALITY CHANGE	TIMESTAMP
Demo1 - Copy.xlsx	Autoquality Cleansing	Full Dataset	duplicateFixed missingFixed whitespaceFixed categoricalFixed	87.71	94.86	7.15	18/6/2025, 3:17:28 pm
Demo1 - Copy.xlsx	Autoquality Cleansing	Full Dataset	No Rules	87.71	0	-87.71	Invalid Date
Demo1 - Copy.xlsx	Manual Cleaning	Selected Columns	No Rules	87.71	0	-87.71	Invalid Date
Demo1 - Copy.xlsx	Manual Cleaning	Selected Columns	No Rules	87.71	0	-87.71	Invalid Date
Demo1 - Copy.xlsx	Manual Cleaning	Selected Columns	No Rules	87.71	0	-87.71	Invalid Date

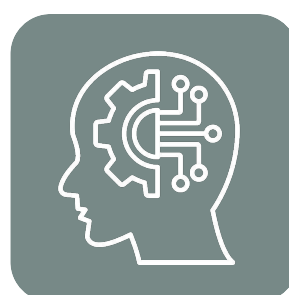
Next Steps

Run data profiling to identify additional issues

WHAT MAKES MATCHX AI-FIRST



AI drives profiling, matching, rule generation, and scoring



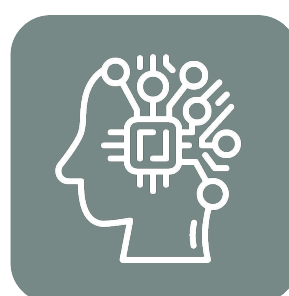
Models are trained across domains for contextual accuracy



No-code prompt interface for rule creation



Adaptive thresholds tuned by dataset and quality trends



Explainable AI at every step — human in the loop, always



HOW MATCHX TRANSFORMED DATA FOR LEADING ORGANIZATIONS



National Health Service (NHS):
Eliminating Duplicate Patient Records For Better Healthcare



Challenge:
Duplicate patient records, inconsistent medical history, and incorrect details across hospitals led to delays in care and compliance risks



MatchX Impact:

- AI-powered matching & deduplication: Eliminated duplicate patient records for accurate medical history.
- Automated data validation & cleansing: Standardized patient information to improve reliability.
- Approval workflows for compliance: Ensured patient records met NHS data governance standards.
- Data lineage tracking: Provided a transparent audit trail of all updates.

Results:

- 20% fewer duplicate records: Faster access to accurate patient history.
- Improved compliance: Enhanced patient data governance.
- Reduced manual effort: More time for patient care.

HOW MATCHX TRANSFORMED DATA FOR LEADING ORGANIZATIONS



UK Department For Work & Pensions (DWP):
Ensuring Accurate Welfare Payments



Challenge:

Duplicate and mismatched citizen records caused incorrect payments, fraud risks, and inefficiencies in benefit distribution



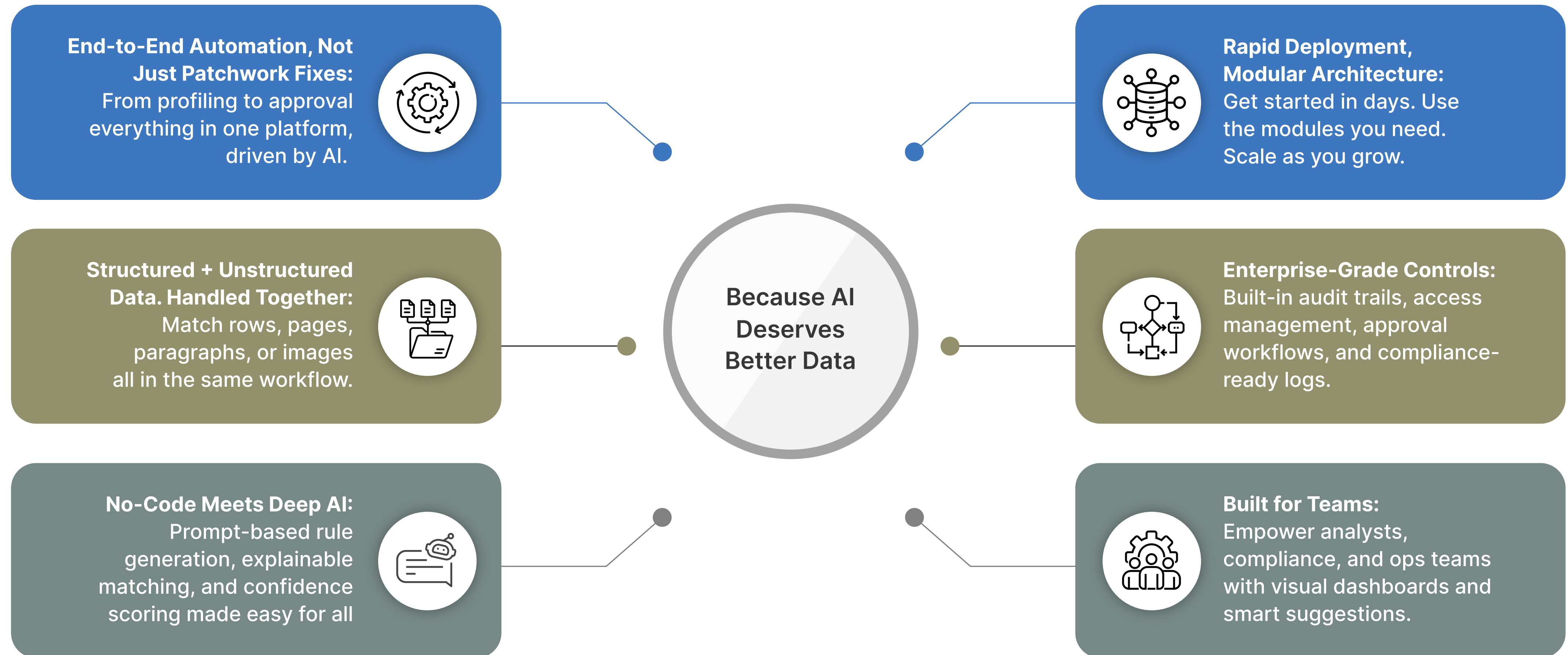
MatchX Impact:

- **AI-driven fraud detection:** Identified suspicious claims and anomalies.
- **Cross-dataset matching & deduplication:** Linked citizen records for accurate benefits processing.
- **Instant data quality insights:** Real-time dashboards showed data health.
- **Automated ingestion & validation:** Standardized data from government sources.

Results:

- **30% reduction in duplicate claims** Prevented fraud & overpayments.
- **Faster eligibility verification** – Reduced processing time for benefits.
- **Significant cost savings** – Improved efficiency in public services.

WHY MATCHX?





SECURITY & COMPLIANCE

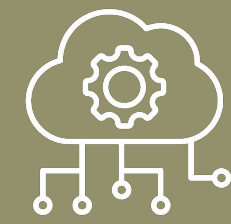
Built for trust, connected for flexibility, and ready to scale.



Security: MatchX ensures enterprise-grade data protection with end-to-end encryption (AES-256, TLS 1.2+), role-based access (RBAC), multi-factor authentication, and full audit trails. Compliant with GDPR, HIPAA, and SOC 2 standards.



Security: API-first and extensible, MatchX connects effortlessly with your ecosystem. Use REST or GraphQL APIs, pre-built connectors (Salesforce, Snowflake, Workday), or event-driven integrations via Kafka and webhooks.



Security: Cloud-native and performance-optimized, MatchX supports both batch and real-time processing at scale. Deploy across AWS, Azure, or GCP with built-in auto-scaling and distributed architecture.

SUBSCRIPTION TIERS & DEPLOYMENT OPTIONS

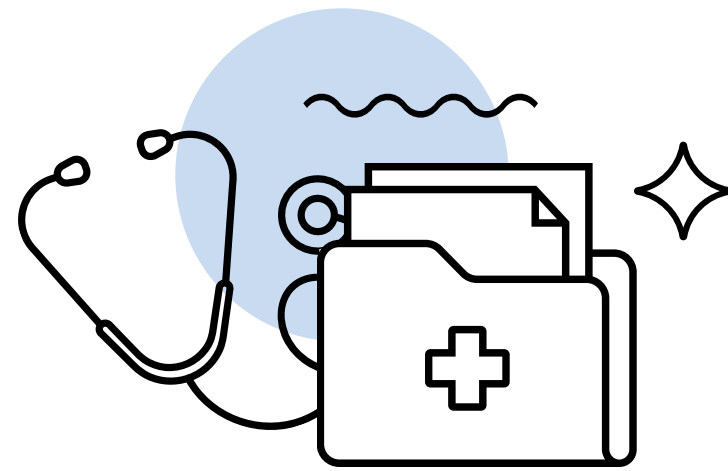


Tier	Features	Supportive File Types	Max File Size	Max Data Size
Basic	Core ingestion, AI-assisted profiling, and essential rule-based validation	CSV, JSON, Excel	50 MB	5 GB
Pro	Advanced AI-driven profiling, document matching, and dynamic rule application	CSV, JSON, Excel, PDF	100 MB	20 GB
Team	AI-powered cleansing, intelligent linking, API access, and interactive dashboards	CSV, JSON, Excel, PDF, XML	200 MB	100 GB
Enterprise	AI-driven threshold management, automated auditing, compliance monitoring, and seamless integrations	CSV, JSON, Excel, PDF, XML, Images	500 MB	1 TB
Premium Enterprise	Fully customizable AI-workflows, predictive insights, and enterprise-grade SLA support	All formats (CSV, JSON, PDF, XML, Images, Video)	Unlimited	Unlimited

INDUSTRY COVERAGE



Built for Your Domain. Trusted Across Industries.



Healthcare & Life Sciences

- Match patient records across hospitals, labs, and insurance systems
- Validate medical documents and reconcile scanned forms
- Maintain compliance with full audit trails and lineage



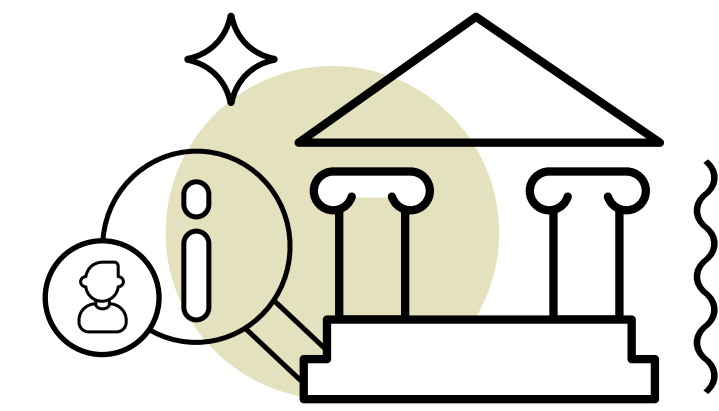
BFSI (Banking, Financial Services, Insurance)

- Match KYC records, transactions, and customer profiles
- Identify duplicates and prevent fraud in real-time
- Reconcile statements, contracts, and application forms



Retail & e-Commerce

- Create a single view of customer across POS, CRM, and online platforms
- Deduplicate supplier and product master data
- Match invoices and shipping documents to avoid leakage



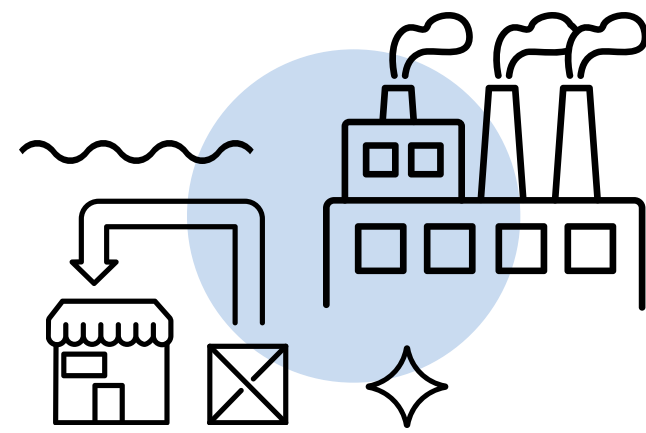
Public Sector & Government

- Link citizen identities across departments
- Match scanned forms, affidavits, and legacy documents
- Support national-level data standardization and transparency

INDUSTRY COVERAGE



Built for Your Domain. Trusted Across Industries.



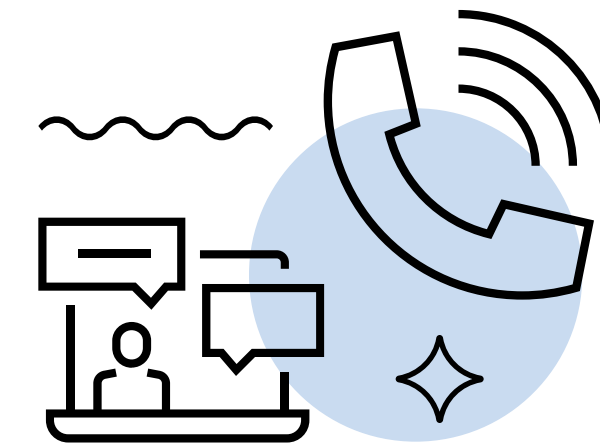
Logistics & Manufacturing

- Match supplier, vendor, and product records
- Reconcile delivery documents and inventory logs
- Link POs, GRNs, invoices across siloed systems



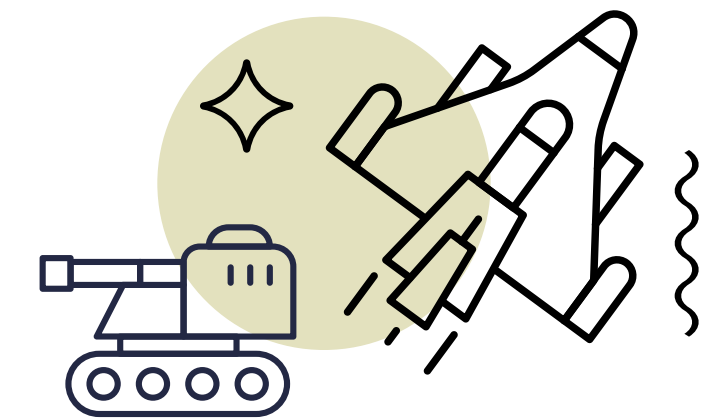
Legal Services

- Contract Comparison to detect discrepancies.
- Party/Entity Record Reconciliation for legal accuracy.
- Document Metadata Alignment to standardize case files.



Telecommunications

- Subscriber Record Validation for customer accuracy.
- Contact Detail Deduplication for better outreach.
- SIM/Device Linking to prevent fraud.



Aerospace & Defense

- Parts & Vendor Matching for procurement accuracy.
- Technical Spec Comparison for engineering consistency.
- Equipment Metadata Alignment to enhance maintenance.

USE CASE



One Platform. Many High-Impact Use Cases.

See how MatchX transforms data chaos into confident decisions across business functions.

- **1. Entity Resolution Across Systems**
Unify fragmented records by identifying and merging duplicates across CRMs, ERPs, and databases.
- **2. Contract & Document Comparison**
Detect paragraph-level changes between contracts, scanned forms, and legal documents using AI-powered NLP.
- **3. Supplier & Vendor Deduplication**
Remove duplicates and standardize supplier/vendor data across procurement systems and legacy platforms.
- **4. Healthcare Claims Matching**
Match patient records, scanned claim forms, and provider documents to accelerate claim approvals and reduce errors.
- **5. Research Data Harmonization**
Link clinical trial data, patient logs, and research notes for unified analysis in pharma and life sciences.
- **6. Customer 360 & Master Record Creation**
Create a single, trusted view of each customer by matching and linking records across multiple data sources.
- **7. Product & Catalog Matching**
Match product entries across supplier catalogs, inventory systems, and e-commerce platforms to remove redundancy.
- **8. Invoice & PO Reconciliation**
Match purchase orders, invoices, and receipts to flag discrepancies and improve accounts payable accuracy.
- **9. Citizen Data Linking for Government**
Match citizen identities across departments and records to enable unified service delivery and e-governance.
- **10. KYC & Identity Matching**
Automate customer identity verification by matching KYC documents and application data across systems.



READY TO MATCHX YOUR DATA?

You Don't Just Need Perfect Data. You Need MatchX.
Say goodbye to duplicates, delays, and dirty data. Say
hello to AI-ready clarity.

MatchX It. Because Your Data Deserves the Best

Contact Us



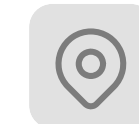
team@ve3.global



+44 (0) 20 4552 0840 | +1 (332) 2322334



www.ve3.global/matchx/



86-90 Paul Street, London EC2A 4NE
228 Park Ave S, New York, NY, 10003-1502, US