

DAVA

The New AI Data Engineer

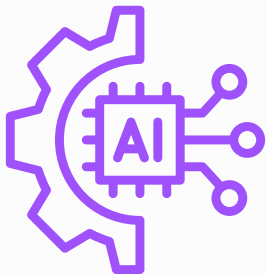
A AI-powered data normalization platform that turns chaotic, multi-format datasets into unified, actionable intelligence automatically.

Built for enterprise scale. Designed for human simplicity.

www.avaresearch.ai

Why DAVA?

Most organizations drown in disjointed data – CSVs, SQL dumps, APIs, and logs that never align. DAVA changes that. It uses multi-agent LLM intelligence to generate custom parsers on the fly, process files securely inside containerized sandboxes, and deliver normalized datasets with 90 %+ accuracy.



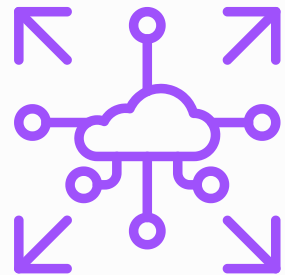
Automation



Accuracy



Security



Scalability

The result:

instant interoperability, zero manual mapping, and full audit-ready visibility.

Core Capabilities

Automated Data Parsing

Generates custom parsers for CSV, JSON, SQL, TXT using large language models.

Intelligent Format Detection

Identifies schema & structure with 90%+ parsing accuracy through dual AI evaluation.

Secure Docker Execution

Sandboxed environments for safe code execution and data privacy.

Enterprise Deduplication

Sandboxed environments for safe code execution and data privacy.

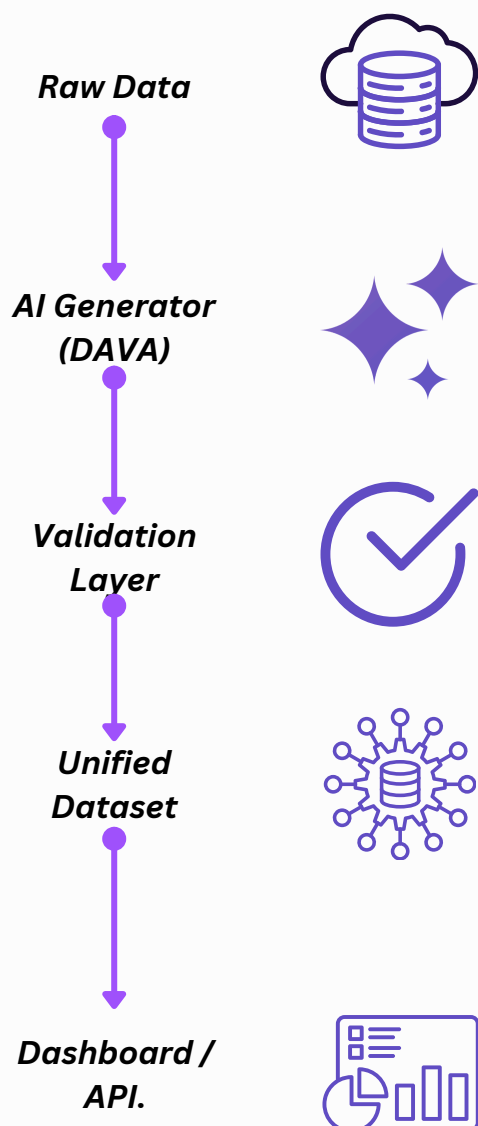
Real Time Monitoring

Web dashboards & Grafana integration for observability.

System Limits

DAVA has no enforced file-size caps. Throughput depends on: Machine memory and CPU, Selected LLM model, Retry steps required for complex schemas.

The Engine Behind DAVA

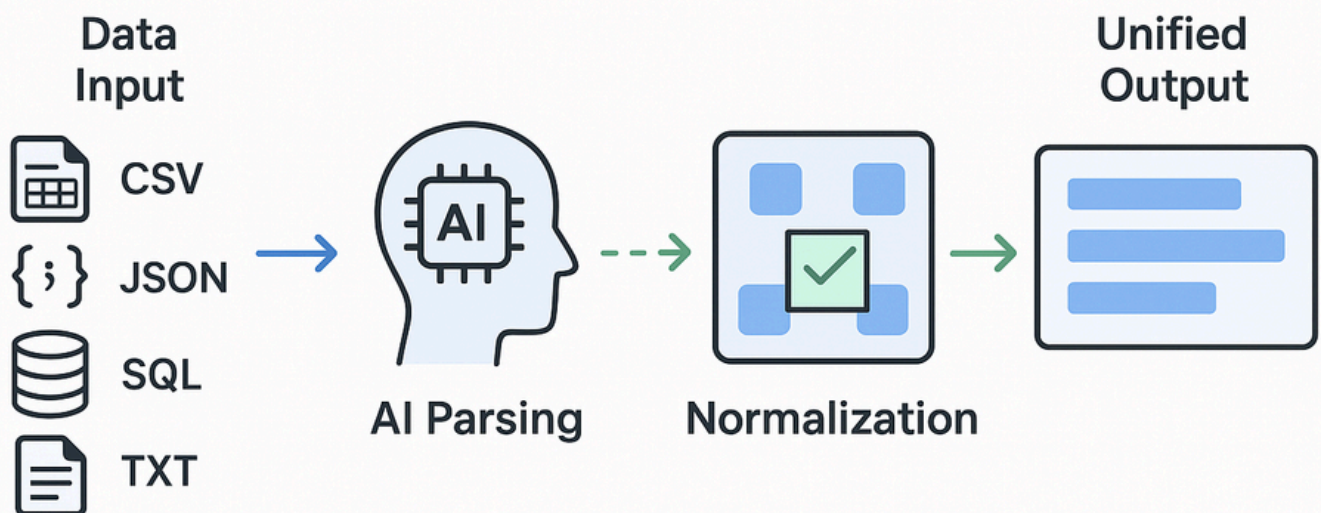


- **AI-Generated Parsers:**
 - DAVA's large-language-model agents write, test, and execute format-specific parsers in real time – from CSVs to JSON, SQL, or proprietary logs.
- **Intelligent Format Detection:**
 - Dual-layer validation ensures schema recognition with 90 %+ parsing accuracy.
- **Secure Execution:**
 - All processes run in isolated Docker containers with zero data leakage.
- **Live Monitoring:**
 - Grafana dashboards visualize every transformation.
- **LLM Redundancy:**
 - Multi-provider architecture (OpenRouter, Anthropic, Ollama) guarantees uptime and fallback continuity.

We empower organizations to bridge complexity, scalability and speed into their data engineering.

Architecture Highlights

AI-Powered Data Normalization Platform



Transforms diverse, messy data sources into clean, unified, and actionable intelligence

- Containerized & cloud-ready deployment
- Multi-provider LLM support (OpenRouter, Anthropic, Ollama)
- Automatic file watching & live ingestion
- Distributed processing (scalable pipeline)

Installation & Setup

DAVA installs in minutes and operates fully on your local environment.

Requirements:

Python 3.11+

Install dependencies via `uv sync` or `pip install -e .`

Configure a `.env` file with your preferred LLM provider key (OpenRouter, Anthropic, or Ollama)

Docker optional; recommended for sandboxed execution

Automatic fallback to a secure local sandbox when Docker is unavailable

USE CASES / INDUSTRIES

± 15 Days

Finance & Compliance:

Convert scattered audit logs into unified reporting tables.

Telecom & Retail:

Clean, deduplicate, and merge multi-source customer data.

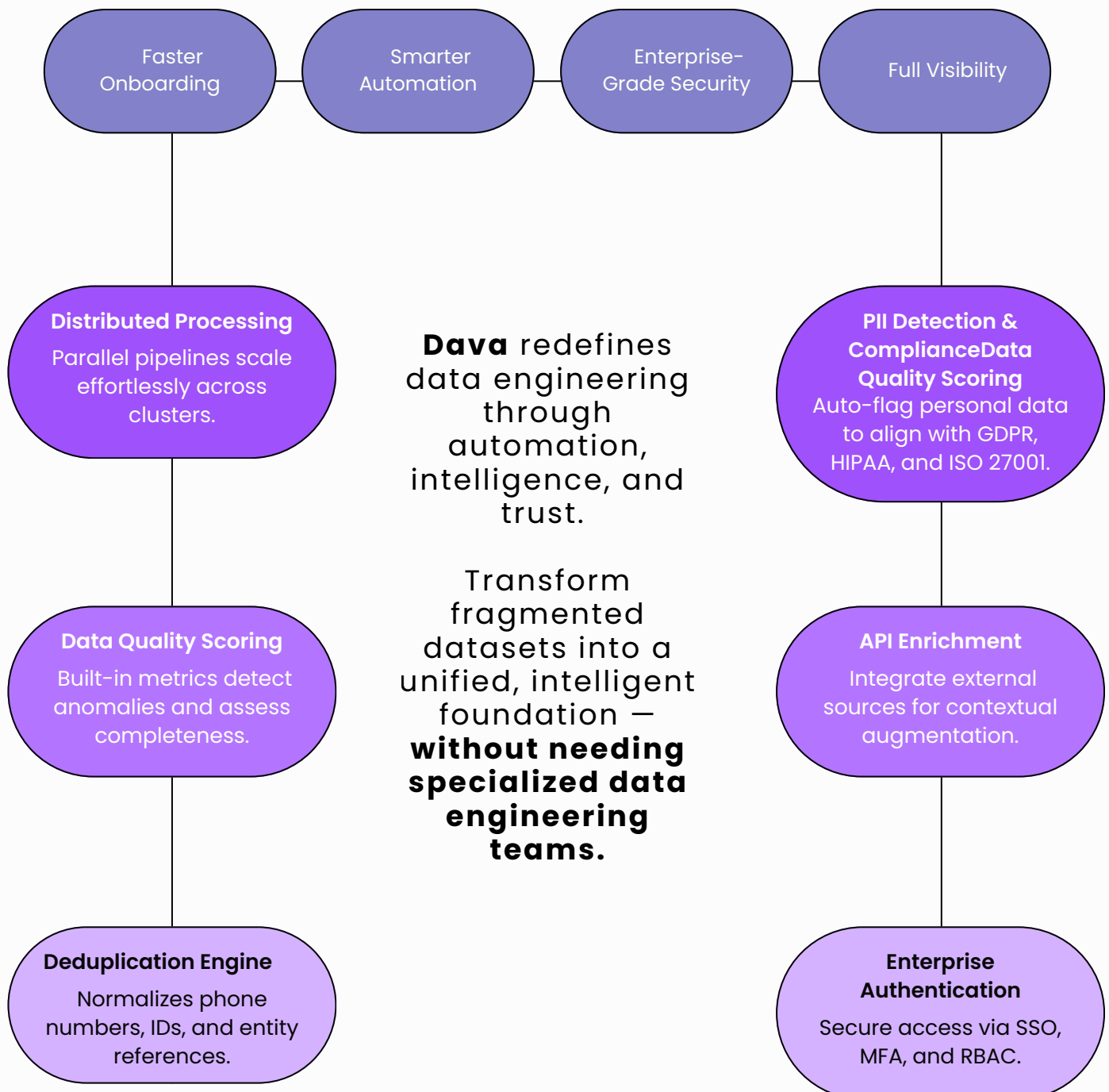
Public Sector:

Standardize incoming CSVs and XLSX forms across departments with zero engineering overhead.

Deploy in 1 day. Scale to billions of records.

—
“DAVA replaced weeks of manual data cleaning with a single upload. It became our invisible engineer — fast, precise, and auditable.”
— CTO, Confidential Beta Partner

Enterprise Features



Enterprise Features

Security Posture

DAVA is designed with a strict local-first model to protect sensitive data.

- All processing occurs locally; no data leaves your environment unless explicitly configured
- LLM-generated parser code executes in a secure sandbox (Docker recommended)
- Simple API key authentication
- All logs and invalid records remain local
- No automatic deletion; retention policies are fully user-controlled

Data Storage

- DAVA provides full transparency and control over where your data lives.
- Raw files are stored in uploads/
- Normalized outputs are stored in a SQLite database (normalized_data.db)
- Smart Tables mode creates tailored SQLite table structures automatically
- All exports are saved as CSV files inside output/jobs/{job_id}/

Upcoming Enhancements

- Human-in-the-Loop Approval
- Before final ingestion, DAVA presents a proposed schema for review, allowing teams to edit or approve mappings.
- Automatic Mapping to Internal Schemas
- DAVA aligns incoming fields with internal taxonomies for seamless integration into existing systems.



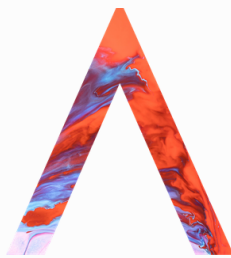
DAVA by AVA Research



AI-driven normalization. Enterprise-grade trust.
Transform your data chaos into clarity.

-

Contact
avaresearch.ai/dava
or contact anca.goron@avaresearch.ai



AVA RESEARCH

www.avaresearch.ai