

Minsky™ DataCleanse

AI-Driven Dynamic Data Cleansing & Transformation Platform

Overview

Minsky™DataCleanse is an enterprise-grade, AI-driven dynamic data cleansing and transformation platform engineered to convert raw, inconsistent enterprise datasets into target-system-ready, validated, and auditable data. It combines a two-stage AI orchestration model with deterministic, rule-based validation to automatically understand, cleanse, normalize, and transform data at enterprise scale — replacing slow, error-prone manual mapping with a governed, repeatable pipeline.

Built for the AI-first, data-sovereign, and migration-intensive enterprise — offering LLM provider-agnostic architecture, deterministic governance, and full audit traceability from raw source file to target-compliant record.

Problem Statement

Enterprises struggle with legacy and third-party data that is inconsistent, poorly structured, and not compliant with modern ERP or analytics systems. Manual data cleansing and mapping are slow, error-prone, expensive, and difficult to audit — especially during large-scale migrations to systems such as SAP S/4HANA or Oracle Fusion, where every field mapping, transformation, and exception must be traceable for governance sign-off.

Platform Navigation & Modules

A single, unified workspace covering every stage of the data transformation lifecycle:

Dashboard	Upload / Ingest	Schema Mapping	Validation Rules
Transformation Jobs	Exception Queue	Audit & Lineage	Reports

Every dataset — legacy source files, third-party extracts, or live ERP exports — flows through the same governed pipeline: AI-driven understanding, executable transformation, and rule-based validation, with full audit visibility at every stage.

Our Solution — How It Works

Minsky™DataCleanse uses a two-stage AI orchestration model combined with deterministic validation to automatically understand, cleanse, normalize, and transform enterprise data.

Stage 1 – Intelligent Data Understanding: An AI-driven analysis engine examines the uploaded dataset to infer schema, field semantics, data quality issues, and target-system expectations.

Stage 2 – Executable Data Transformation: The platform generates fully transformed, target-compliant records using AI, followed by strict rule-based validation to ensure accuracy and governance.

AI Capabilities

AI is embedded natively across Minsky™DataCleanse — not a bolted-on add-on tier.

AI Schema Inference – automatically detects field types, relationships, and target-system schema from raw, unstructured, or semi-structured source data

Field Semantic Mapping – AI-driven mapping of source fields to target-system fields (SAP S/4HANA, Oracle Fusion, and other ERP/CRM schemas)

Anomaly & Outlier Detection – flags malformed, duplicate, missing, or out-of-range values before they reach the target system

Auto-Generated Transformation Rules – the AI layer proposes cleansing, normalization, and mapping rules, reviewable before execution

Confidence Scoring & Exception Routing – every transformed record carries a confidence score; low-confidence records route to a human-in-the-loop exception queue

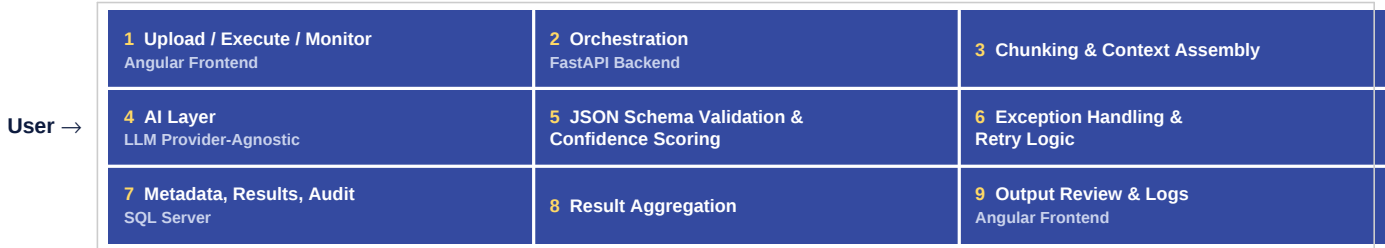
Ask Minsky™ – conversational AI assistant for natural-language queries against job status, data quality metrics, and audit history

AI in Action

Schema Auto-Mapping	Ask Minsky™	Exception Resolution Assist
Upload a raw dataset and the AI layer infers schema and proposes field-level mappings to the target system — reviewable before any record is transformed.	Ask in plain English — 'How many records failed validation in the last SAP migration batch?' — and get instant job status, data quality, and audit answers.	Low-confidence and failed records are queued with AI-suggested corrections, so data stewards resolve exceptions in minutes, not hours.

Architecture

A modern, LLM provider-agnostic pipeline: an Angular frontend for upload, execution, and monitoring; a FastAPI orchestration backend; a provider-agnostic AI layer with chunking and context assembly; JSON schema validation with confidence scoring; and full exception handling, audit, and result aggregation logged to SQL Server before output review.



User Journeys

Data Engineers – dataset upload, schema mapping review, transformation job execution, error triage
IT / Migration Leads – ERP migration planning, target-schema configuration, go-live readiness tracking

Data Governance & Compliance – audit and lineage review, validation rule sign-off, exception oversight
Business Analysts – data quality reports, dashboards, and Ask Minsky™ natural-language queries

Security & Governance

Enterprise-grade security and governance are built into every layer of Minsky™DataCleanse.

Role-Based Access Control – field-level and job-level permissions by role and team
Deterministic Validation – strict rule-based checks applied after every AI transformation, ensuring accuracy and repeatability
Encryption – end-to-end, at rest and in transit

Full Audit Logs – every upload, mapping decision, transformation, and exception tracked and retained
Data Residency Control – via On-Prem and Sovereign Cloud deployment
Tenant Isolation – strict row-level data separation between clients

Integration Ecosystem

Minsky™DataCleanse connects directly into target ERP systems — including SAP S/4HANA and Oracle Fusion — plus SQL Server for metadata and audit storage, and open REST APIs and webhooks for the broader migration toolchain.

How a Migration Runs



Why Enterprises Choose Minsky™DataCleanse

2-Stage	100%	Full	Any LLM
AI understanding + executable transformation, not a single black-box pass	Records validated by deterministic, rule-based checks before output	Audit and lineage traceability from source file to target record	Provider-agnostic AI layer — no single-vendor lock-in

Minsky™DataCleanse vs. Manual / Legacy ETL Tools

Capability	Minsky™DataCleanse	Manual / Legacy ETL Tools
Schema & Field Mapping	AI-inferred, reviewable before execution	Manually authored, project-by-project
Data Understanding	AI-driven semantic analysis of source data	Relies on tribal knowledge and manual review
Transformation Rules	AI-generated, human-in-the-loop approval	Hand-coded scripts and mappings
Validation	Deterministic, rule-based, applied to every record	Inconsistent, often spot-checked only
Exception Handling	Confidence-scored routing to review queue	Manual discovery, often post-migration
Auditability	Full metadata, results, and audit log by default	Limited or bolted-on after the fact
Deployment Model	Cloud, On-Prem, or Sovereign Cloud	Typically fixed to a single environment

Deployment Options: Subscription Multitenant Cloud | White-Label on Client Server | On-Prem Servers | Sovereign Cloud Servers | Full Solution with Source Code

See Minsky™DataCleanse in Action

Schedule a live walkthrough and discover AI-driven data migration governance at a fraction of the cost of manual ETL.

Info25@ailabsinc.com | www.ailabsinc.com