CARTOGRAPH AI

WHITEPAPER

CARTOGRAPH SOLUTIONS

Contents

Version Control

Version	Author	Details	Date
1.0	Patrick Robak	Version 1	12/5/2024
1.1	Patrick Robak	Addition of graphics	7/16/2025
1.2	Patrick Robak	General revisions	8/13/2025

I. Introduction

A. Problem

LexisNexis Risk Solutions released its annual report detailing the "True Cost of Financial Crime Compliance". The 2023 study revealed that FCC costs companies $206.1 billion globally, and the driving force behind the surge is the rising price of expertise and technology. This problem is not unique to Anti-Money Laundering, or even Financial Institutions – it applies to all Enterprise Software and Services across all domains and businesses.

While combating rising costs, technology teams are also contending with common organizational challenges:

- People – Resource constraints and unclear data ownership
- Processes – Manual data wrangling and nascent data governance
- Technology – Use of redundant tools and limited architecture discipline
- Data – Lack of trust in siloed data and limited awareness of data assets

B. Solution

CartographAI is an automated data analysis, mapping, and transformation engine that supercharges subject matter expertise with artificial intelligence. Core functionalities include:

1. Profiling and Discovery

a. CartographAI analyzes, reviews, and summarizes source data
b. Understands structure, content, and interrelationships
c. Assigns meaning to tables (ex. Transactions, Accounts, Customers, etc.)
d. Assigns meaning to attributes (ex. Transaction ID, Account Number, Amount, etc.)
e. Flags quality, accuracy, completeness, consistency, timeliness, and accessibility issues

2. Mapping and Transformation

a. CartographAI extracts relevant data from the source
b. Transformations adhere to target data types, constraints, and dependencies
c. Data is loaded according to the specified timing and frequency
d. Detailed documentation is automatically generated for transparency and traceability
e. Mappings requiring manual intervention are flagged for review
f. Perpetual monitoring ensures data quality does not degrade over time

C. Use Cases

Compelling use cases for CartographAI continue to emerge as the product develops and is socialized with senior industry insiders facing unique challenges.

1. New Implementations

The implementation of a new software application represents perhaps the most natural use case for CartographAI. Unknown and poorly documented sources must be mapped to an entirely new target data model within the constraints of an aggressive project timeline. CartographAI accelerates this effort with a high degree of confidence, allowing team members to focus on other complex deliverables.

2. Additional Data Sources

A smaller effort but more frequent occurrence, the addition of new data sources is largely automated using CartographAI. Bring a new business line or product online faster while ensuring compliance with data security requirements.

3. Data Assessments

With data being the foundation upon which models are built, an increasing number of stakeholders are taking a microscope to the end-to-end data quality, completeness, governance, and lineage feeding those systems. CartographAI breaks down knowledge silos and provides unprecedented transparency to model owners, validators, auditors, and regulators.

II. Functionality

A. Modules

CartographAI is comprised of three core modules: Discover, Map, and Transform.

1. Discover

The Discover module of CartographAI rapidly accelerates the discovery, analysis, and profiling of unknown and poorly documented data sources.

It connects to and interprets data sources of many different types, including structured and unstructured data. These include, but are not limited to RDMS, NoSQL, Data Lakes, Lakehouses, CSV, XML, JSON, Parquet, and other file types. CartographAI does not store any source data, only the relevant metadata and the insights/outputs it produces.

After connecting to a source, Discover consumes various inputs to gain deeper insights into the nature of the data. The more inputs available to the Discover module, the more accurate the outputs are expected to be. However, input data quality is also an important consideration, as poor input quality could adversely affect results. Discover inputs include:

a. Source table names, schemas, and indexes
b. Source attribute names, data types, lengths, defaults, and constraints
c. Source attribute values
d. Source metadata
e. Source documentation

After evaluating the available inputs, the Discover module outputs several key artifacts that detail the data source. Discover outputs include:

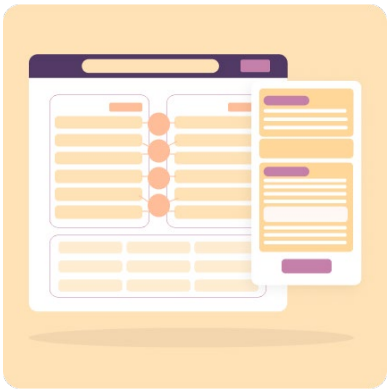
a. High-level functional descriptions of each table and each attribute
b. A detailed data dictionary
c. An Entity Relationship Diagram (ERD)
d. A data profiling report

In addition to the aforementioned outputs, the Discover module also offers the following features:

a. Confidence score: All insights generated by the Discover module are measured against a configurable confidence score. Any output that does not meet the predetermined threshold is flagged as such and either requires modified module inputs or manual verification.
b. Co-pilot: The Discover Co-pilot is an AI-powered virtual assistant. Users can interact with the co-pilot to clarify Discover insights and drill deeper into source databases, tables, attributes, and interrelationships.
c. Duplicate data: The Discover module will have the ability to identify duplicate data, either within the same source or across different sources. This helps ensure the same

transactions, customers, or accounts are not loaded into the target application more than once.

2. Map



The Map module of CartographAI leverages the Discover module output along with a knowledge base of mapping rules to propose detailed source-to-target mappings for any target FCC application.

CartographAI will include mapping rules for all significant Financial Crime Compliance applications, beginning with the most prevalent. The knowledge base will be expanded and improved over time.

After evaluating the available inputs, the Map module outputs detailed and exportable source-to-target mapping specifications for each critical target attribute.

In addition to the aforementioned outputs, the Map module also offers the following features:

a. Manual review/override: Each mapping produced by CartographAI can be manually reviewed in the Map Matrix UI. The mapping can either be approved or overwritten, if deemed necessary. All such actions would be fully audited.
b. Sample output: Each mapping is accompanied by a real-time output sample to visualize the effects of any applied transformations. This assists the user in validating the mappings and any necessary changes.
c. Output validation: Automated scripts that validate the mapping output to ensure it conforms to the necessary data types, constraints, and dependencies.
d. 4-eye review: A validation that requires at least 2 users to approve a mapping. These validations can be applied to specific target attributes, or across all mappings.
e. Confidence score: All mappings generated by the Map module are measured against a configurable confidence score. Any output that does not meet the predetermined threshold is flagged as such and either requires modified module inputs or manual verification.
f. Reverse mapper: Rather than use Discover insights and the rule library as inputs to produce source-to-target mappings, use existing ETL code and scripts as inputs to reverse engineer existing mappings. This is useful in cases where the current transformations and pipelines are poorly documented but need to be validated. The reverse mappings can also be compared to those CartographAI would typically generate to identify potential gaps for remediation.

3. Transform

After generating detailed data mapping specifications, Transform consumes those instructions as an input to autonomously generate ETL code/pipelines that can be executed manually or through a job scheduling tool.

ETL pipelines come in different forms, all of which can be accommodated by CartographAI. Including, but not limited to SQL, Python, Apache Airflow, Apache Spark, Google Composer and more.

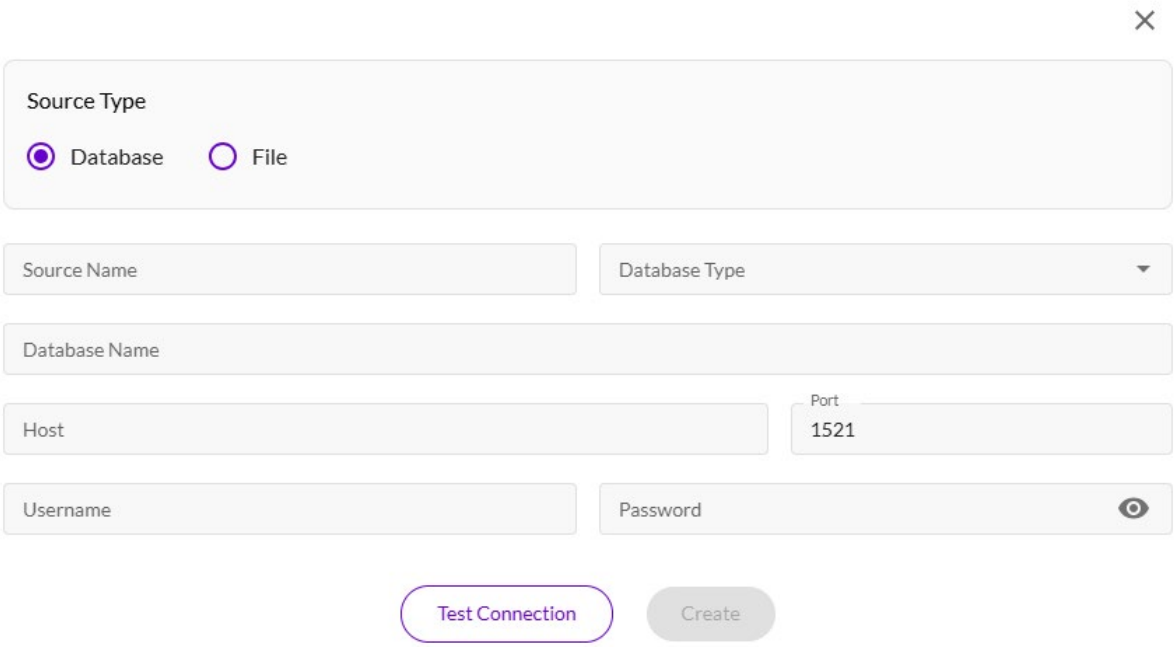
In order to productionize ETL pipelines, CartographAI seamlessly integrates with key systems in the existing ecosystem.

B. User Interface

CartographAI features a specially designed user interface to configure and interact with each core module. The main components of the UI include:

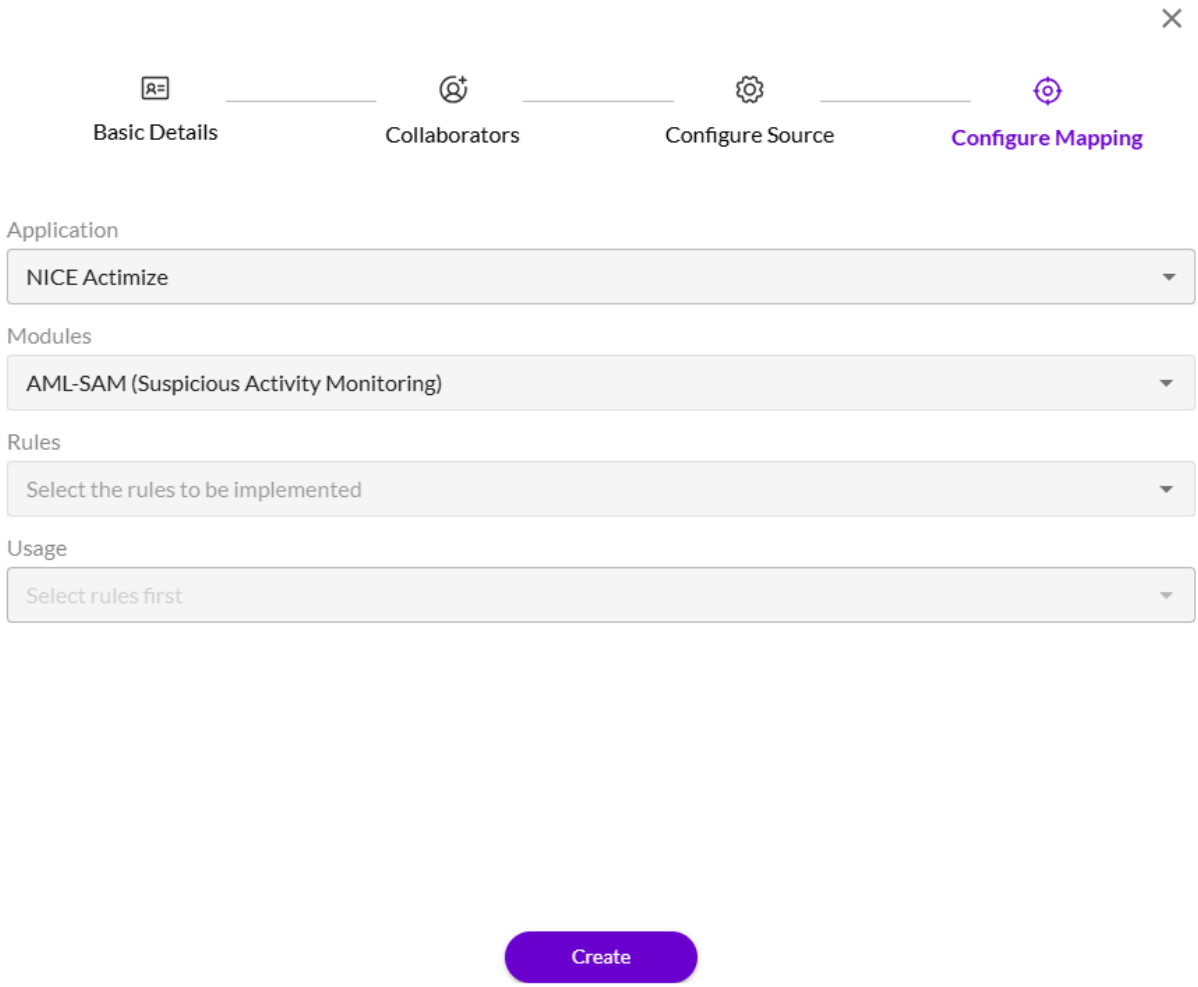
1. Data Source Configuration

The data source interface is a way to connect CartographAI to data sources of various types, typically by entering credentialed information like username, password, hostname, port, etc. The interface saves connection details for future use.



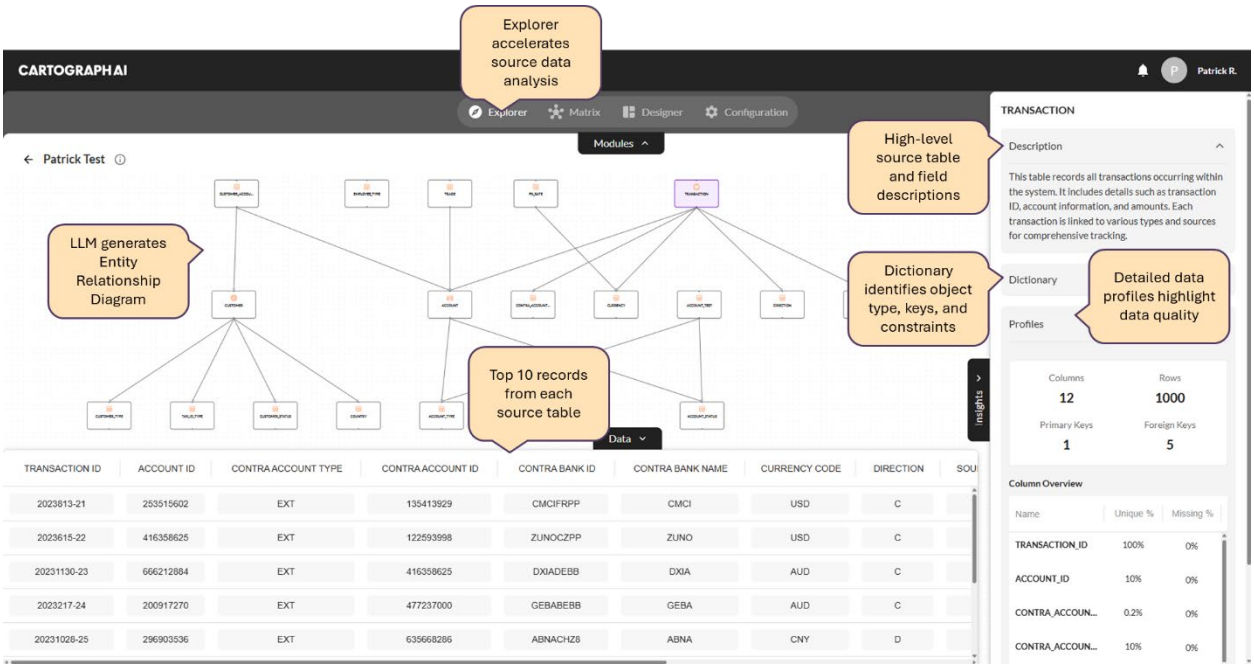
1. Target Application Configuration

The target application interface is a way to filter the mapping rule library to only map those fields within the scope of each specific project. The user specifies the target application, modules, rules, and even field usage.



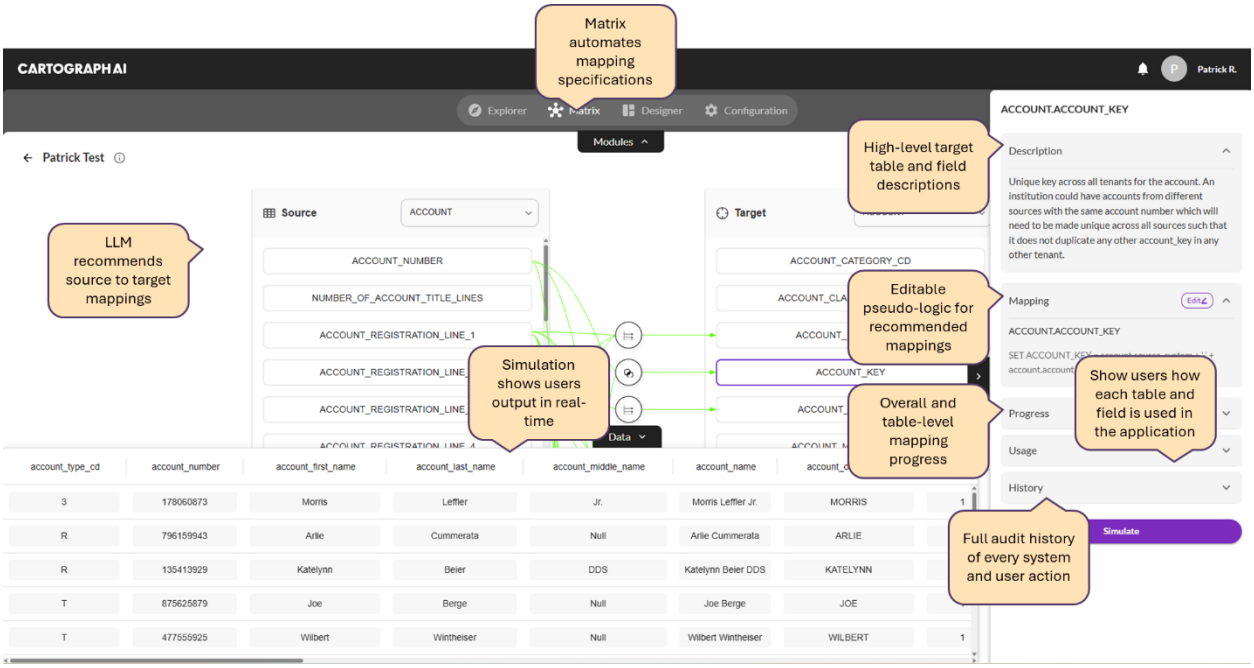
2. Discover Explorer

The Explorer is the primary screen of the Discover module, displaying all the relevant outputs in a user-friendly and interactive manner.



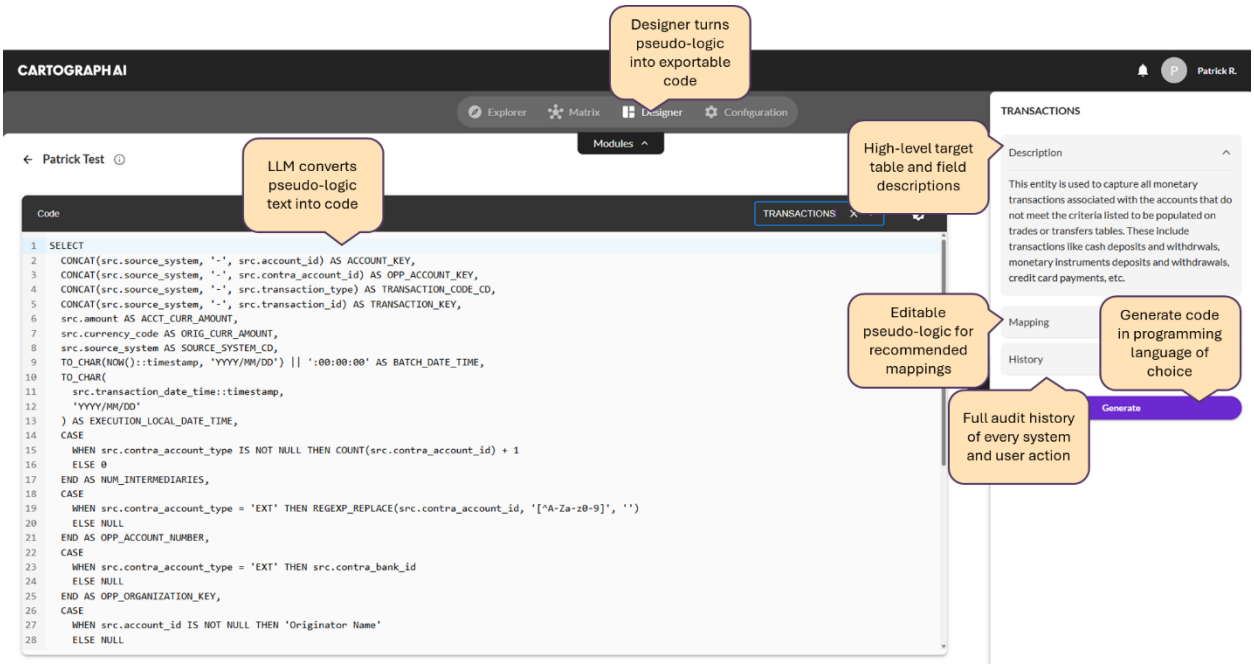
3. Map Matrix

The Matrix is the primary screen of the Map module, displaying all the relevant mappings in a user-friendly and interactive manner.



4. Transform Designer

The Designer is the primary screen of the Transform module. It is an easy-to-use web interface where users can create, modify, and manage processes.



C. Governance

The CartographAI data governance framework is a set of rules and processes that manage data within the application. The purpose of the framework is to ensure the integrity, security, and compliance of data, and to establish a standard for how data is collected, organized, stored, and used. This makes it easier to streamline and scale data governance, maintain policy and regulatory compliance, democratize data, support collaboration, and build trust.

1. Quality

Data quality involves processes to ensure data is accurate, complete, and reliable. CartographAI measures and reports the below data quality elements at both the source and target levels on a perpetual basis.

a. Accuracy
b. Completeness
c. Consistency
d. Timeliness
e. Accessibility

2. Auditability

The framework includes robust auditing functionalities that provide detailed records of every action taken within CartographAI, whether conducted by a user or the system. This audit history helps ensure data security and regulatory compliance by monitoring when, how, and by whom changes are made.

3. Lineage

Data lineage process within CartographAI tracks the life cycle of data, including how it's generated, transformed, moved, and used across the system. It provides a detailed record of the data's origins, transformations, and movements. This lineage is presented to users as both a report and as a visualization within the CartographAI user interface.

D. Security

The data security framework protects digital information from unauthorized access, theft, or corruption.

1. IAM

Identity Access Management grants secure access to CartographAI to verified entities, ideally with a bare minimum of interference. The goal is to manage access so that the right people can do their jobs and the wrong people, like hackers, are denied entry.

There are two parts to granting secure access to an organization’s resources:

a. Identity management
b. Access management (Role-based access control)

Granting the correct level of access after a user’s identity is authenticated is called authorization. The goal of the IAM system is to make sure that authentication and authorization happen correctly and securely at every access attempt.

2. Logging

The framework records detailed information about data access, modifications, and movement within the system by tracking the "trace" of a data request as it travels through different components, allowing security teams to identify potential threats and suspicious activity by monitoring how data is being accessed and manipulated across the system, essentially providing a comprehensive view of data flow for enhanced security analysis.

3. Compliance

CartographAI is working towards SOC 2 and ISO 27001 certified status.

E. Infrastructure

CartographAI follows cloud-native design and is vendor agnostic. It works both in public cloud infrastructure and private cloud infrastructure.

III. Licensing

Access to CartographAI is granted via subscription-based license. Every instance requires a separate license to run the application. CartographAI administrators have the provision to revoke the license as governed by the terms and conditions.