ikigai

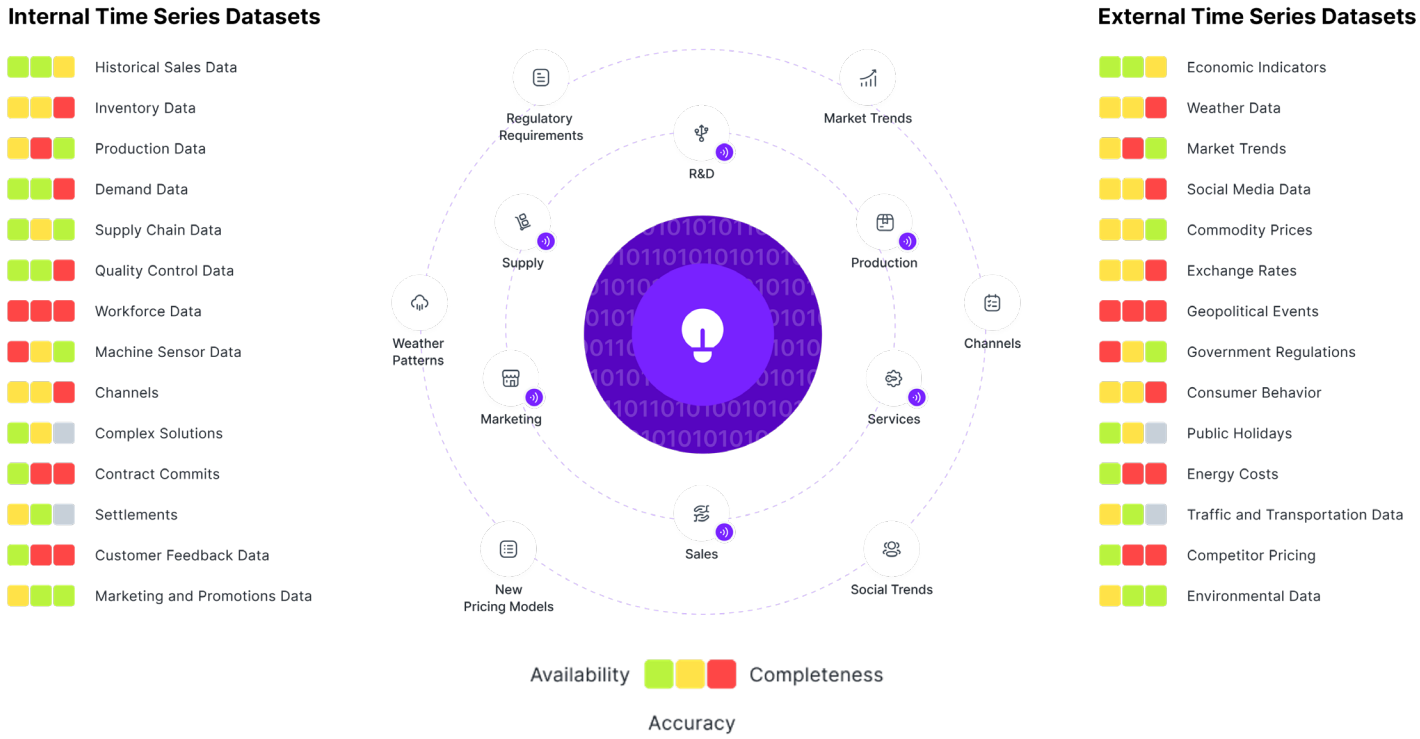
RECONCILING DISPARATE DATA SOURCES WITH AIMATCH AND EXPERT-IN-THE-LOOP

WHAT'S INSIDE

INTRODUCTION

The goal of applying AI/ML to enterprise data is to extract meaningful patterns and insights, enabling impactful decision making. To achieve this end, it is essential to understand how all the data fits together, so meaningful relationships can be surfaced. However, enterprise data is often siloed across multiple data sources, primarily due to different aspects of the business being managed within different processes and tools, resulting in many partial sources of truth. To gain a holistic understanding of the data, it is necessary to stitch the information together across these disparate sources.

Image 1. Disparate data sources needed for accurate enterprise decision making



In an ideal scenario, where data sources are designed carefully by a centralized system designer, a natural way to stitch them together would be through traditional database or data warehouse operations such as joins relying on shared identifiers or keys.

However, the reality is that such data sources are rarely designed with easy integration in mind. To make matters worse, the data often contains missing values as well as errors and anomalies, making automation of data unification a challenging task.

The goal of aiMatch, a generative AI solution from Ikigai, is to precisely address this challenge. With a computationally efficient approach via Large Graphical Models (LGMs), Ikigai AI capabilities – such as aiMatch – can solve large and critical problems for the enterprise.

For the problem of data reconciliation, aiMatch brings together previously disparate datasets by matching data across tables with AI and human oversight, known as expert in the loop (XitL). In the following sections, we walk through an example of a typical business scenario that is supported by aiMatch.

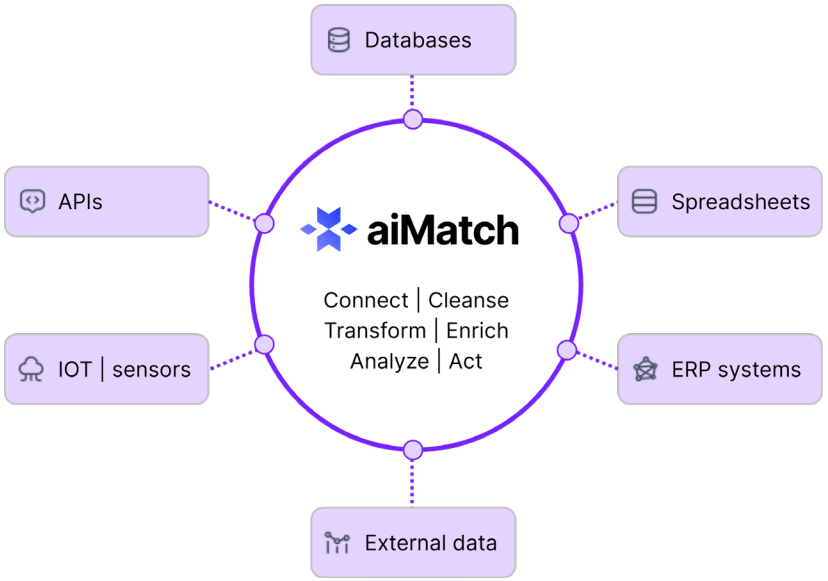


Image 2. aiMatch and eXpert-in-the-Loop unify enterprise data

DATA INPUT: TWO MISMATCHED TABLES

Consider the following business scenario: a large business with a complex network of suppliers and vendors relies on its General Ledger (GL) system to maintain its overall financial records, while using a separate cash management system to manage banking transactions and cash flows across multiple banking partners and payment processors.

When the company makes a sale, pays a vendor, or incurs an expense, the transaction is recorded in the GL system using the appropriate GL account names, debit amounts, and credit amounts. These transactions are summarized and batched before being entered into the GL system, with each batch assigned a unique Batch ID. The company's bank transactions, such as customer payments, vendor payments, and bank charges, are recorded in the cash management system, where each transaction is assigned a unique statement item (SI) number, Reference ID, and other relevant details like value date, posting date, and transaction reference.

While each system serves its purpose, reconciling the data from any two systems is no easy task, with the resulting output of these systems being two mismatched tables that need to be reconciled by the business. In the next several paragraphs, we will show how aiMatch can be used in this example to reconcile data between two disparate sources.

The following two sample datasets will be used in the following sections to illustrate how aiMatch and eXpert-in-the-Loop are used to connect, harmonize, and validate disparate data sources.

GL Account ID	Entry ID	Booking Date (Entry Date)	Transaction ID	GL Account Name	Debit Amount	Credit Amount	Notes
100104	20051383	2023-08-04 10:55:26	LN13081794844575	Nostro/ Clearing Acc CB	N/A	571	Cleared transaction from CBUAE & LN13081794844575 .
100104	20104417	2023-08-04 17:49:26	LN13330092087947	Nostro/ Clearing Acc CB	N/A	1890	Cleared transaction from CBUAE & LN13330092087947 .
100104	20095336	2023-08-04 16:31:26	LN13336869693964	Nostro/ Clearing Acc CB	N/A	4158	Cleared transaction from CBUAE & LN13336869693964 .
100104	20066164	2023-08-04 12:30:30	LN13189688496732	Nostro/ Clearing Acc CB	N/A	17579	Cleared transaction from CBUAE & LN13189688496732 .
100104	20023273	2023-08-04 06:33:25	LN12977733928092	Nostro/ Clearing Acc CB	N/A	9525	Cleared transaction from CBUAE & LN12977733928092 .
100104	20065210	2023-08-04 12:20:30	LN13130860940167	Nostro/ Clearing Acc CB	N/A	1754	Cleared transaction from CBUAE & LN13130860940167 .
100104	20093272	2023-08-04 16:12:31	LN13269877626355	Nostro/ Clearing Acc CB	N/A	154938	Cleared transaction from CBUAE & LN13269877626355 .
100104	20065084	2023-08-04 12:19:30	LN13183736957743	Nostro/ Clearing Acc CB	N/A	260847	Cleared transaction from CBUAE & LN13183736957743 .
100104	20082169	2023-08-04 14:41:31	LN13268692210564	Nostro/ Clearing Acc CB	N/A	44700	Cleared transaction from CBUAE & LN13268692210564 .
100104	20097658	2023-08-04 16:49:31	LN13292357136032	Nostro/ Clearing Acc CB	N/A	11970	Cleared transaction from CBUAE & LN13292357136032 .
100104	20073586	2023-08-04 13:28:31	LN13171587568398	Nostro/ Clearing Acc CB	N/A	64000	Cleared transaction from CBUAE & LN13171587568398 .

Table 1. The General Ledger system records and organizes all financial transactions of a business. When the company makes a sale, pays a vendor, or incurs an expense, the transaction is recorded in the GL system using the appropriate GL account names, debit amounts, and credit amounts.

Sl No.	Reference ID	Value Date	Posting Date	Tran Type	Batch ID	Narrative 1	Narrative 2	Narrative 3	Narrative 4	Transaction Reference	Debit/Credit	Transaction Amount
1974	I950050823000001	2023-04-08	2023-04-08	IPS	IQ5N	IPI23080403N6P4Q	LN13081794844575	REMIBANNUMB-AE920860000009752447672	RQST-230804105525.RESP-230804105531	AUHIQ5N232166380	Debit	571
4447	I950050823000001	2023-04-08	2023-04-08	IPS	IQ6H	IPI23080403NALF9	LN13330092087947	REMIBANNUMB-AE670860000009292489351	RQST-230804174912.RESP-230804174922	AUHIQ6H232166533	Debit	1890
4133	I950050823000001	2023-04-08	2023-04-08	IPS	IQ6C	IPI23080403N9VKD	LN13336869693964	REMIBANNUMB-AE270860000009020546400	RQST-230804163020.RESP-230804163035	AUHIQ6C232166509	Debit	4158
2555	I950050823000001	2023-04-08	2023-04-08	FTS	RLCC	July invoice	LN13189688496732	LN13189688496732		AUHRLCC232160001	Debit	17579
499	I950050823000001	2023-04-08	2023-04-08	IPS	IQ58	IPI23080403N56SK	LN12977733928092	REMIBANNUMB-AE460860000009277897049	RQST-230804063153.RESP-230804063154	AUHIQ58232166320	Debit	9525
2482	I950050823000001	2023-04-08	2023-04-08	FTS	RJXY	Hotel booking	LN13130860940167	LN13130860940167		AUHRJXY232160001	Debit	1754
3917	I950050823000001	2023-04-08	2023-04-08	FTS	SCEE	Ply Wood Marine 18mm4FT8FT 5to7 tim	LN13269877626355	LN13269877626355		AUHSCEE232160001	Debit	154938
2475	I950050823000001	2023-04-08	2023-04-08	FTS	RJS0	bill payment	LN13183736957743	LN13183736957743		AUHRJS0232160001	Debit	260847
3393	I950050823000001	2023-04-08	2023-04-08	FTS	S1PD	to pay venders	LN13268692210564	LN13268692210564		AUHS1PD232160001	Debit	44700
4087	I950050823000001	2023-04-08	2023-04-08	FTS	SG80	Invoice no 1560	LN13292357136032	LN13292357136032		AUHSG80232160001	Debit	11970
2945	I950050823000001	2023-04-08	2023-04-08	FTS	RT23	pay	LN13171587568398	LN13171587568398		AUHRT23232160001	Debit	64000
4733	I950050823000001	2023-04-08	2023-04-08	IPS	IQ6N	IPI23080403NBHM9	LN1344702881414	REMIBANNUMB-AE470860000009694844644	RQST-230804192816.RESP-230804192818	AUHIQ6N232166096	Debit	82.95

Table 2. The company's bank transactions, such as customer payments, vendor payments, and bank charges, are recorded in the cash management system, where each transaction is assigned a unique statement item (SI) number, Reference ID, and other relevant details like value date, posting date, and transaction reference.

STEP 1: SCHEMA MAPPING

To reconcile these datasets together, we must identify which rows in Table 1 match to which rows in Table 2.

In order to see matches or more general similarities between any pair of rows, we must first understand what columns in Table 1 have affinity with columns in Table 2. To accomplish this, aiMatch generates Schema Mapping out-of-the-box. For the example tables shown above, the following is output generated by aiMatch.

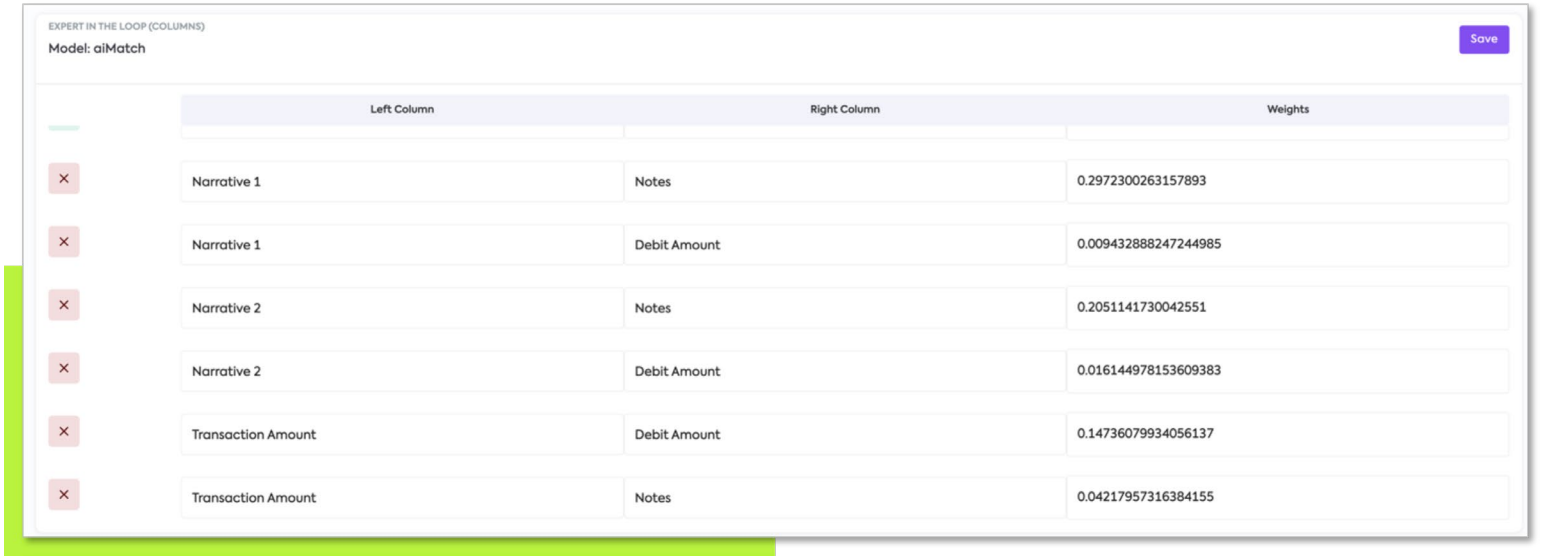


Image 3. Suggested matches are accompanied by the strength of the relationship, with a higher weight suggesting a stronger potential match. For example, with weights of 0.2972 and 0.2051 respectively, we can see that Narrative 1 and Narrative 2 are more likely to be related to Notes than to Debit Amount, which has been assigned much lower weights (0.0094 and 0.0161 respectively).

When a user runs aiMatch, the system suggests the top matches out-of-the-box, as shown in the example on the previous page. It shows that certain columns in Table 1 (under Left Column) are similar to the columns in Table 2 (under Right Column) with a similarity score (under Weights). For example, if two columns match perfectly then their similarity score will be 1, and if they do not match at all then their similarity score will be 0.

STEP 2: EXPERT FEEDBACK TO IMPROVE SCHEMA MAPPING

With the first attempt, the out-of-the-box column mapping is likely to have some errors. This can be corrected quickly with the help of an expert in the loop (XitL), who can remove suggested matches and add new matches to improve aiMatch's output.

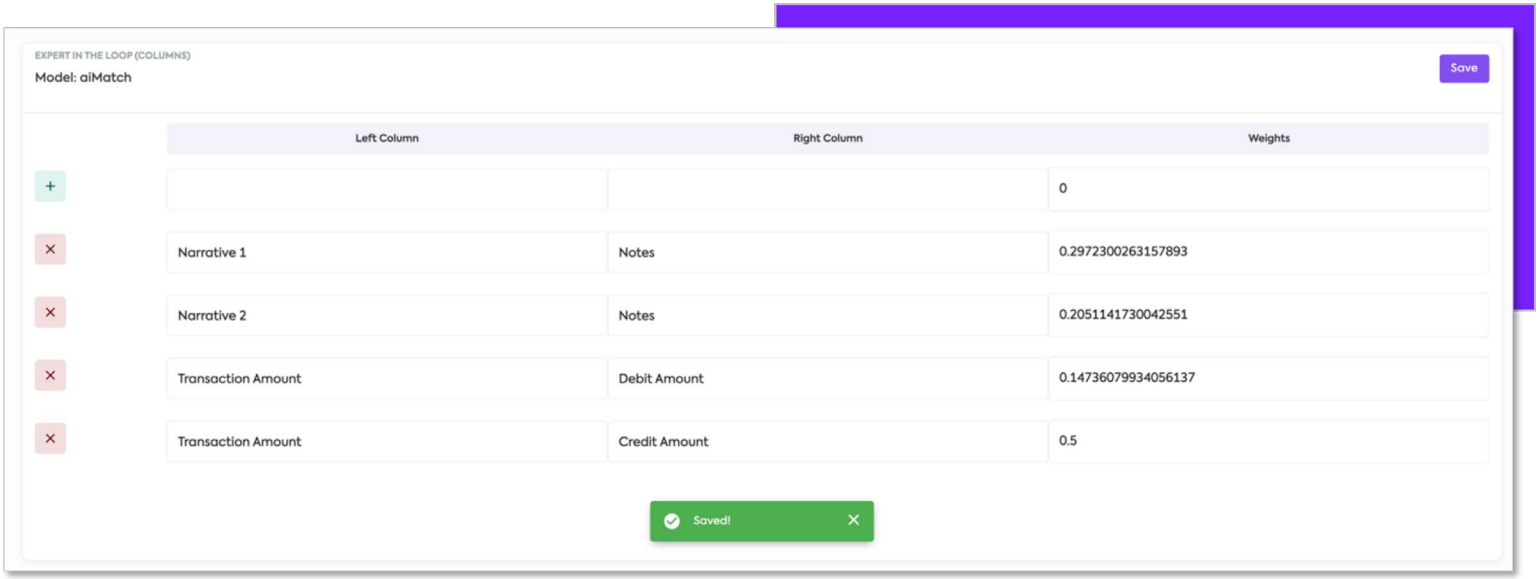


Image 4. Above we see the output of schema mapping upon few removals and an addition (Transaction Amount <> Credit Amount with 0.5 Weight).

STEP 3: ROW MATCHING

aiMatch uses the schema mapping to determine the matches of rows across Tables 1 and 2. This results in a certain number of rows being mapped while others remain unmatched. In the image below, we see that after the first round of matching, more than 10% of the rows remain unmatched across tables with the distribution of pair-wise row similarities depicted below.

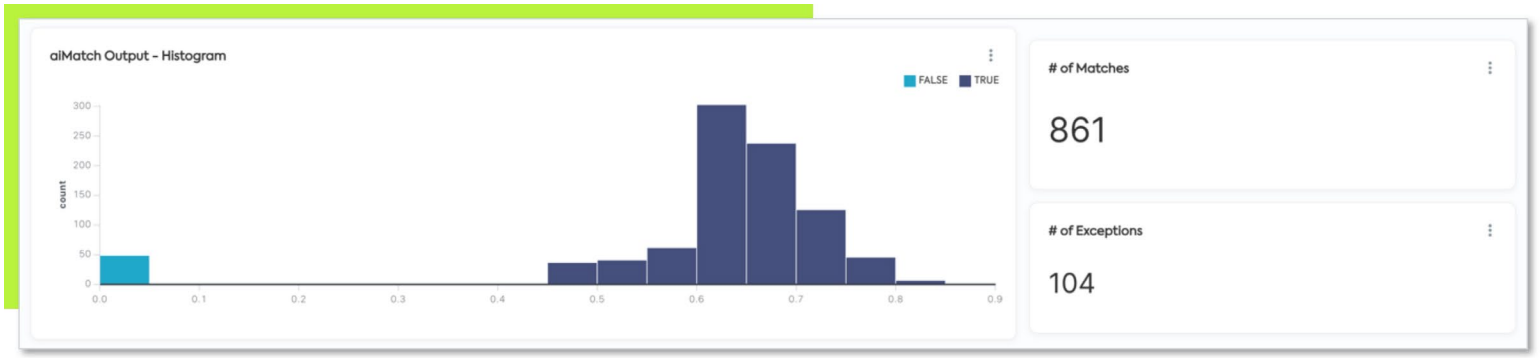
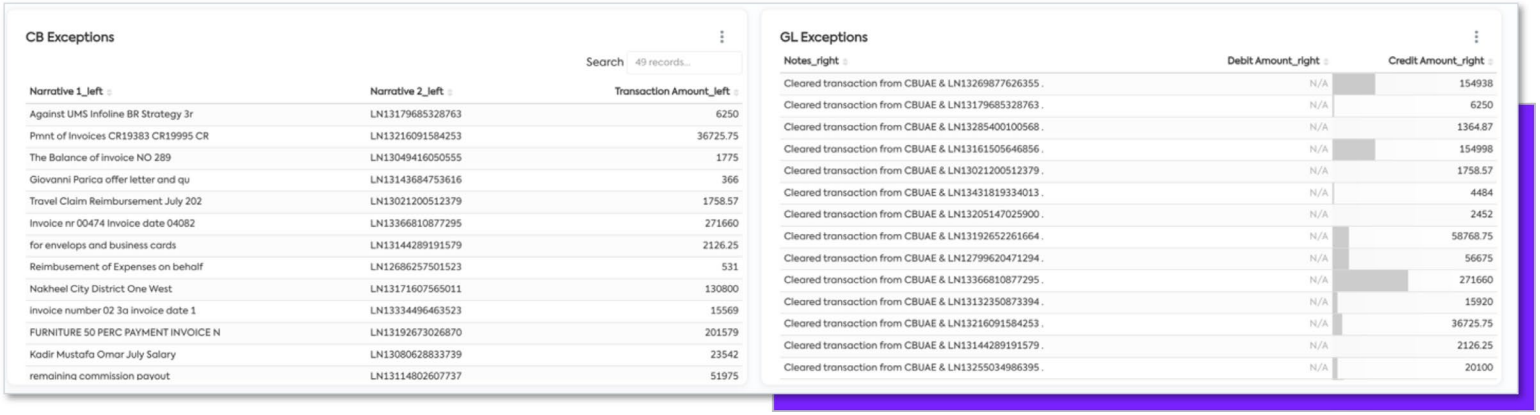


Image 5. By providing simple yes/no inputs as to the quality of the matches, it's easy to quickly reduce the number of exceptions to a manageable set for internal business users to review.

Once a user clicks into the data to see the examples of exceptions across Tables, it quickly becomes clear that there are some obvious examples in both tables that should be matched.



CB Exceptions

Search 49 records...

Narrative 1_left	Narrative 2_left	Transaction Amount_left
Against UMS Infoline BR Strategy 3r	LN13179685328763	6250
Pmnt of Invoices CR19383 CR19995 CR	LN13216091584253	36725.75
The Balance of invoice NO 289	LN13049416050555	1775
Giovanni Parica offer letter and qu	LN13143684753616	366
Travel Claim Reimbursement July 202	LN13021200512379	1758.57
Invoice nr 00474 Invoice date 04082	LN13366810877295	271660
for envelops and business cards	LN13144289191579	2126.25
Reimbusement of Expenses on behalf	LN12686257501523	531
Nakheel City District One West	LN13171607565011	130800
invoice number 02 3a invoice date 1	LN13334496463523	15569
FURNITURE 50 PERC PAYMENT INVOICE N	LN13192673026870	201579
Kadir Mustafa Omar July Salary	LN13080628833739	23542
remainina commission payout	LN13114802607737	51975

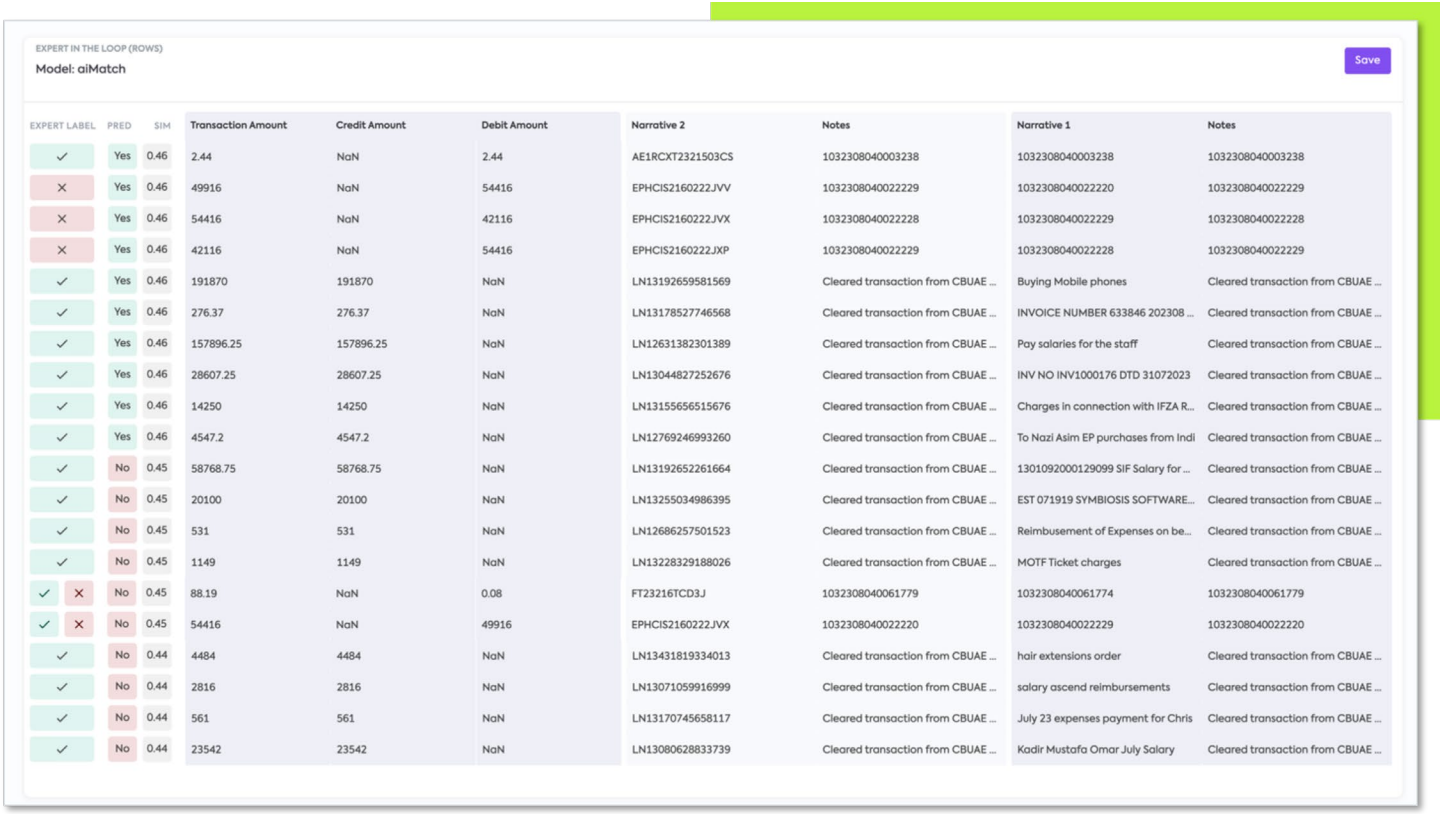
GL Exceptions

Notes_right	Debit Amount_right	Credit Amount_right
Cleared transaction from CBUAE & LN13269877626355 .	N/A	154938
Cleared transaction from CBUAE & LN13179685328763 .	N/A	6250
Cleared transaction from CBUAE & LN13285400100568 .	N/A	1364.87
Cleared transaction from CBUAE & LN13161505646856 .	N/A	154998
Cleared transaction from CBUAE & LN13021200512379 .	N/A	1758.57
Cleared transaction from CBUAE & LN13431819334013 .	N/A	4484
Cleared transaction from CBUAE & LN13205147025900 .	N/A	2452
Cleared transaction from CBUAE & LN13192652261664 .	N/A	58768.75
Cleared transaction from CBUAE & LN12799620471294 .	N/A	56675
Cleared transaction from CBUAE & LN13366810877295 .	N/A	271660
Cleared transaction from CBUAE & LN13132350873394 .	N/A	15920
Cleared transaction from CBUAE & LN13216091584253 .	N/A	36725.75
Cleared transaction from CBUAE & LN13144289191579 .	N/A	2126.25
Cleared transaction from CBUAE & LN13255034986395 .	N/A	20100

Image 6. Matching opportunities identified

STEP 4: IMPROVING ROW MATCHING

While aiMatch uses ML to identify potential matches between the fields across different systems, the real power of the tool lies in the way it solicits expert feedback. Rather than requiring users to go through an entire collection of matched/not-matched sets, aiMatch solicits input from reviewers in terms of thumb up or thumbs down on very few carefully chosen matches and non-matches.



EXPERT LABEL	PRED	SIM	Transaction Amount	Credit Amount	Debit Amount	Narrative 2	Notes	Narrative 1	Notes
✓	Yes	0.46	2.44	NaN	2.44	AE1RCXT2321503CS	1032308040003238	1032308040003238	1032308040003238
×	Yes	0.46	49916	NaN	54416	EPHCIS2160222JVV	1032308040022229	1032308040022220	1032308040022229
×	Yes	0.46	54416	NaN	42116	EPHCIS2160222JVX	1032308040022228	1032308040022229	1032308040022228
×	Yes	0.46	42116	NaN	54416	EPHCIS2160222JXP	1032308040022229	1032308040022228	1032308040022229
✓	Yes	0.46	191870	191870	NaN	LN13192659581569	Cleared transaction from CBUAE ...	Buying Mobile phones	Cleared transaction from CBUAE ...
✓	Yes	0.46	276.37	276.37	NaN	LN13178527746568	Cleared transaction from CBUAE ...	INVOICE NUMBER 633846 202308 ...	Cleared transaction from CBUAE ...
✓	Yes	0.46	157896.25	157896.25	NaN	LN12631382301389	Cleared transaction from CBUAE ...	Pay salaries for the staff	Cleared transaction from CBUAE ...
✓	Yes	0.46	28607.25	28607.25	NaN	LN13044827252676	Cleared transaction from CBUAE ...	INV NO INV1000176 DTD 31072023	Cleared transaction from CBUAE ...
✓	Yes	0.46	14250	14250	NaN	LN13155656515676	Cleared transaction from CBUAE ...	Charges in connection with IFZA R...	Cleared transaction from CBUAE ...
✓	Yes	0.46	4547.2	4547.2	NaN	LN12769246993260	Cleared transaction from CBUAE ...	To Nazi Asim EP purchases from Indi	Cleared transaction from CBUAE ...
✓	No	0.45	58768.75	58768.75	NaN	LN13192652261664	Cleared transaction from CBUAE ...	1301092000129099 SIF Salary for ...	Cleared transaction from CBUAE ...
✓	No	0.45	20100	20100	NaN	LN13255034986395	Cleared transaction from CBUAE ...	EST 071919 SYMBIOSIS SOFTWARE...	Cleared transaction from CBUAE ...
✓	No	0.45	531	531	NaN	LN12686257501523	Cleared transaction from CBUAE ...	Reimbusement of Expenses on be...	Cleared transaction from CBUAE ...
✓	No	0.45	1149	1149	NaN	LN13228329188026	Cleared transaction from CBUAE ...	MOTF Ticket charges	Cleared transaction from CBUAE ...
✓ ×	No	0.45	88.19	NaN	0.08	FT23216TCD3J	1032308040061779	1032308040061774	1032308040061779
✓ ×	No	0.45	54416	NaN	49916	EPHCIS2160222JVX	1032308040022220	1032308040022229	1032308040022220
✓	No	0.44	4484	4484	NaN	LN13431819334013	Cleared transaction from CBUAE ...	hair extensions order	Cleared transaction from CBUAE ...
✓	No	0.44	2816	2816	NaN	LN13071059916999	Cleared transaction from CBUAE ...	salary ascend reimbursements	Cleared transaction from CBUAE ...
✓	No	0.44	561	561	NaN	LN13170745658117	Cleared transaction from CBUAE ...	July 23 expenses payment for Chris	Cleared transaction from CBUAE ...
✓	No	0.44	23542	23542	NaN	LN13080628833739	Cleared transaction from CBUAE ...	Kadir Mustafa Omar July Salary	Cleared transaction from CBUAE ...

Image 7: Incorporating eXpert-in-the-loop for human-centric decision-making.

STEP 4: IMPROVING ROW MATCHING

As feedback is provided, aiMatch instantly incorporates and learns from the feedback, improving its ability to find additional matches.

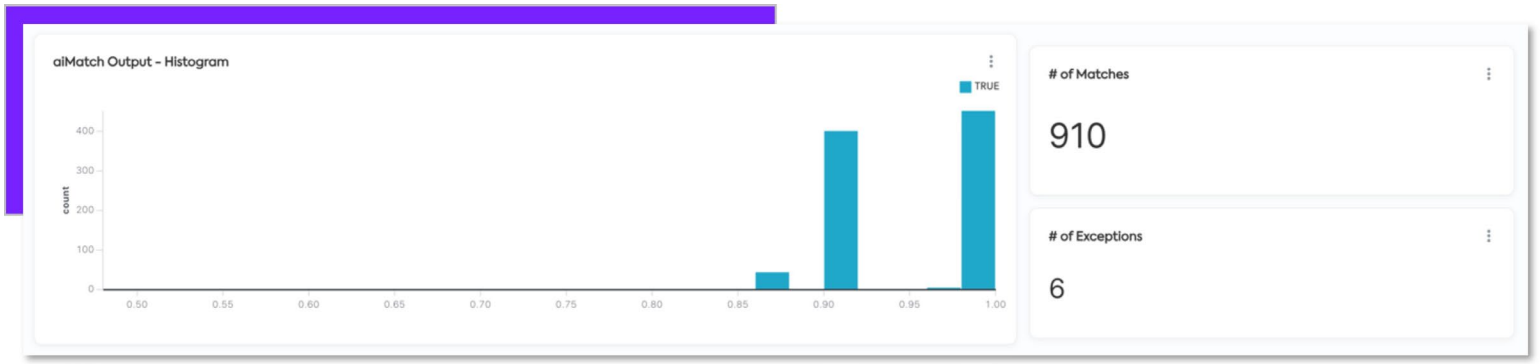


Image 8. Results of eXpert-in-the-Loop intervention

Using the new information provided from the expert review, the system will now go back to find all possible matches. In the image below, we can see that all data has been matched outside of 6 exceptions, which were only left unmatched because there was nothing left to match to in Table 1.

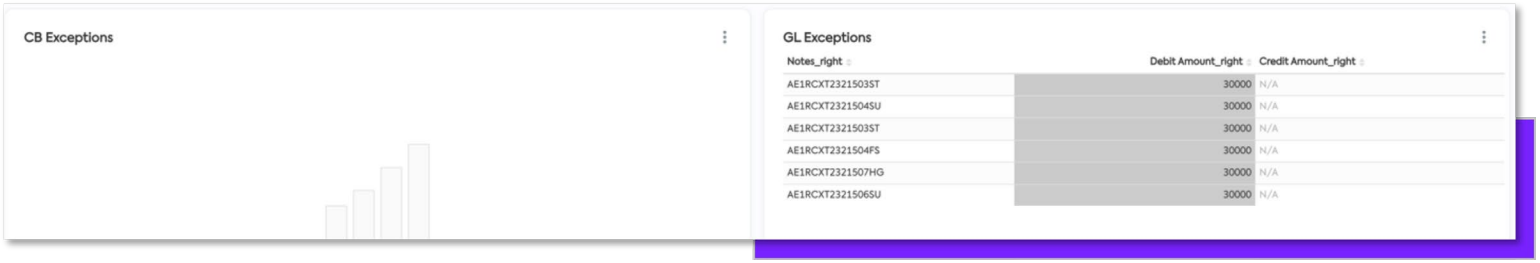


Image 9. Above you can see a small set of exceptions from the General Ledger exist; all other data has been matched.

USE CASE: RETAIL PROMOTION EFFECTIVENESS

A retailer is looking to better understand how different promotions have driven demand across a wide set of products sold across online and in-person stores, as well as if there are trends across geographies and consumer categories. To accomplish this, the retailer will need to look across a wide variety of data sets including:

- Promotion details such as start/end dates, discount amounts, and promotion types
- Promotion targeting information such as customer segments and loyalty tiers, geographic regions, or channels
- Promotion performance metrics such as sales revenue generated, units sold, redemption rates, and profit margins

Bringing the data together – and ensuring it's fit for analysis – is no easy task, as much of the data lives in different systems, and is governed by different schemas and naming conventions.

Basic promotion details (name, dates, and eligible products) are stored in the ERP, while a recently implemented CRM, with its own schema for storing promotional data, houses targeted customer segments and creative assets.

To further complicate matters, the retailer has recently acquired a smaller company that uses a legacy POS system. This system has a different schema for product information, which differs from the retailer's custom PIM system. Integrating the acquired company's data into the existing infrastructure adds another layer of complexity to the already diverse data landscape.

Across all these systems, the data schemas lack uniformity, and the degree of data completeness varies depending on the type of information each system stores. Despite these challenges, harmonizing these disparate data sources is crucial for gaining a comprehensive view of promotions and making data driven decisions.

This is the perfect application of aiMatch. Rather than spending valuable resources to manually reconcile data across sources, the retailer can use aiMatch to largely automate the

task. With the review of an expert in the loop, the retailer will benefit from automation as well as accuracy, making quick work of identifying all possible matches, and preparing the data for further analysis.

CONCLUSION

Connecting and reconciling disparate data sources is a common problem for all businesses across all industries, often accounting for more than 80% of a data analyst's time. Ikigai automates this process by leveraging its patented Large Language Model, aiMatch, to harmonize data across tables for greater efficiency and accuracy. aiMatch integrates human intuition and expertise with its eXpert-in-the-loop feature to quickly address anomalies and exceptions, continuously improving model confidence for increased data quality and improved decision-making.

Additional resources

- aiMatch data reconciliation
- How to transform your business outcomes with generative AI

GLOSSARY

aiCast	aiCast is a forecasting AI model based on patented Large Graphical Models (LGM). It is designed to predict future trends and outcomes based on both historical tabular and time series data and real-time data. aiCast generates 20% more accurate forecasts than traditional models and methods, even with sparse data.
aiMatch	aiMatch is a data reconciliation AI model based on patented Large Graphical Models (LGM). It automates the process of connecting and harmonizing disparate datasets, ensuring consistency and accuracy across multiple sources. By utilizing advanced pattern recognition and probabilistic techniques, aiMatch enables identification and resolution of inconsistent data and can synthesize new data to address missing or incorrect data.
aiPlan	aiPlan is a scenario planning AI model based on Large Graphical Models (LGM) which can generate and evaluate up to 10^{19} scenarios based on complex datasets. By simulating various potential outcomes and their likelihoods, aiPlan enhances scenario planning by providing insights into risks, opportunities, and strategic responses for organizations to navigate uncertainties.
eXpert-in-the-Loop	"eXpert-in-the-loop" (XitI) refers to a hybrid approach in artificial intelligence where human expertise is integrated into the machine learning process. This methodology involves combining the capabilities of machine learning algorithms with human domain knowledge or judgment to improve the accuracy, efficiency, and interpretability of AI systems.
Large Graphical Model (LGM)	A Large Graphical Model is a generative AI model that produces a graph to represent the conditional dependencies between a set of random variables. It is designed to work with enterprise-specific or proprietary data sources, such as tabular and time series data used in data reconciliation, forecasting, and scenario planning.

ikigai

LEARN MORE

www.ikigailabs.io