

# AI BASED MATCHING FOR LEGAL ENTITY MASTERING

## EXECUTIVE SUMMARY

Machine Learning (ML) and artificial intelligence (AI) are revolutionary tools that are dominating the application space and are appearing in many aspects of our lives. Every financial institution stores multiple security master files from a multitude of vendors. Ensuring data quality and making decisions based upon best-in-class data sets, verifying the accuracy, and having the ability to impose overrides for proprietary company data sets is one of the most challenging problems firms face today. This white paper describes how we are using ML and AI to improve quality in legal entity (LE) matching for the investment banking industry.

## INTRODUCTION

West Highland's (WHSS) Legal Entity Management (LEM) service via our Intelligent Data Platform (IDP), combines multiple data vendors into a single, aggregated feed, providing our clients with REST API access to multiple data points relating to pre-matched public and private entities from around the world.

THESE DATA POINTS INCLUDE, BUT ARE NOT LIMITED, TO:

<p>COMPANY PROFILE</p> 	<p>COMPANY HIERARCHY</p> 	<p>AFFILIATES</p> 	<p>FUNDAMENTALS</p> 
<p>PEOPLE</p> 	<p>DEALS</p> 	<p>NEWS</p> 	

**84%**  
of businesses think AI will help them to gain or maintain a competitive advantage.

These fields are defined according to client requirements and are typically ingested into their applications or customer relationship management (CRM) systems. IDP provides an intuitive administrative GUI front end for direct access to the data, enabling firms to manually override or adjust as the need arises.

The matching is performed using our sophisticated machine learning algorithm, allowing us to accurately match and dynamically link fields from traditional or alternate data sources. Each legal entity is assigned a unique WHSS ID.

**80%**  
of business and technology leaders feel AI improves productivity and creates jobs.

## MACHINE LEARNING

Public and private LEs from different vendors are represented in different ways, from the name of the LE to the entity type, country of domicile and other attributes. This data can be prone to errors with differences in name, company legal structure, and other fields, with some vendors even providing additional comments as part of the name field. So, a fixed rules-based / fuzzy-logic approach does not always work reliably as all variations in attributes cannot be anticipated. An AI solution learns to identify matched entities in a more general way and is better at handling unforeseen or unexpected situations.

We utilize AI to identify possible errors that we analyze and feed back to individual vendors for correction, ensuring a higher quality of reference data to our customers.

## TECHNOLOGY

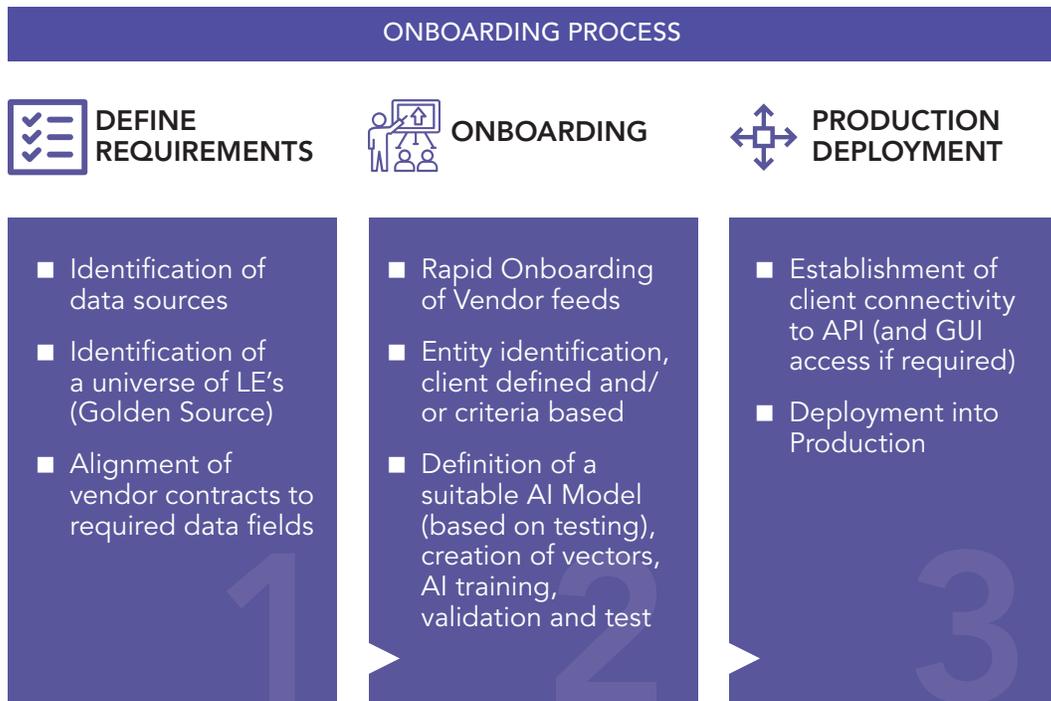
The AI is developed using the industry leading, and most popular framework – TensorFlow. It is Open Source, developed by Google and has become a favorite AI tool of some of the top global companies. IDP is cloud based and therefore highly scalable.

## METHODOLOGY

We cleanse, normalize, and then convert all relevant attributes, that allow for a unique identification of a LE, into vectors, using the latest ML techniques and our proprietary algos. We also generate pseudo random data to create large training data sets. These data sets not only ensure a generalized solution but also ensure we have sufficient examples of all required attribute combinations.

The vectors are then presented to a deep learning AI neural network for training and verification. Once a robust and highly accurate solution is reached, the AI is further tested against never-before-seen data, and ultimately deployed for production use.

The neural network is monitored for accuracy, and will be re-trained as needed, e.g., as datasets are added, or significantly change.



**79%**  
of the global  
business leaders  
say AI would  
make their  
jobs easier and  
efficient.

## MAINTENANCE

IDP is designed from the ground up for scalability and ease of expansion, and so new data fields and vendor sources can be quickly on-boarded at low cost, depending on complexity. Additional API calls can also be rapidly added as needed.

WHSS subscribes to daily vendor feed updates (delta files). These are processed and entity changes are re-matched and available to the production feed before start of day.

There are a number of administrative functions available to manage IDP, which WHSS can perform on behalf of the client. These include:

- LEs can be manually linked or unlinked.
- AI matching thresholds can be defined and changed.
- AI linked LEs can be overridden.
- Possible AI identified vendor errors can be highlighted for analysis.
- Changes to client LEs can be electronically communicated back to vendors on behalf of the client.

## CONCLUSION

West Highland's ML based matching solution represents the next generation of LE mastering capability, providing the benefits of higher accuracy to the client as well as more seamless integration of alternate and less well-structured data sources.

## ABOUT WEST HIGHLAND SUPPORT SERVICES

Bringing over 20 years of market data optimization, operational efficiency, and expense reduction to financial firms through innovative tools and fit-for-purpose solutions.

## AUTHOR

**Steve Massel** is Development Manager at West Highland Support Services. Steve and his team are focused on applying new technologies to the WHSS product range, as well as building new and advanced solutions utilizing the benefits available with Machine Learning and Artificial Intelligence.

## WEST HIGHLAND SUPPORT SERVICES

Wall Street Plaza  
88 Pine Street, Suite 720  
New York, NY 10005  
212.867.1848  
[www.westhighland.net](http://www.westhighland.net)