

## SUCCESS STORY



# SIMPLIFY COMPLEX NUMBER CRUNCHING WITH LLM AND TEX.AI: ACHIEVE 16X TIME SAVINGS AND 95% ACCURACY

## PROJECT OVERVIEW

The client wanted to extract, classify, and summarize value interest fields from more than 50 PDF (legal, finance, and due diligence) documents. The client used AI in combination with other plug-and-play text extraction solutions from Amazon and Azure, which did not meet their requirements.

## SOLUTION DELIVERED

Text analytics leveraging teX.ai and LLM

## CLIENT DOMAIN

Fintech

## KEY HIGHLIGHTS

- The basic entity extraction achieved 95% accuracy.
- Extracted medium and complicated files with 90–95% and 80–90% accuracy, respectively.
- Using Indium's solution for 60% of document extraction saved an incredible 16x time.
- Multi-lingual model enabled clients efficiently handle global customers.

## ABOUT CLIENT

The client is a leading international debt recovery organization, that collaborates with 117 law firms across 100+ countries. Their customer-centric approach prioritizes out-of-court solutions for debt recovery, fostering strong business relations. Backed by skilled in-house lawyers, they've served 3,000+ European clients, managed 10,000+ positions, and 73 million debts in 2 years.

## BUSINESS CHALLENGES

- The client encountered difficulties extracting critical information from a variety of unstructured legal, financial, operational, due diligence, and performance-related documents.
- These documents are collected from diverse locations, and they frequently come in different languages, formats, structures, and currencies.
- The extraction process required adapting off-the-shelf tools such as Textract, Cognizer, and Document AI to meet their requirements.
- However, because these text extraction systems are limited to single-page data extraction, they experienced difficulties when dealing with legal, financial, or due diligence documents.
- Furthermore, managing the large number of data fields inside their documents adds to the complexity.

Indium augmented the client's existing AI platform to perform API-based extraction leveraging a Python OCR pipeline combined with an NLP pipeline. Indium created a pipeline and solution that seamlessly integrate with the client's current documentation platform.

### Document classification

Assigned documents to one of a predefined set of categories, such as legal, finance, operational, and due diligence. Document classification helped organize and identify documents that are similar to each other, and segregate irrelevant ones.

### Page classification

Assigned a page to one of a predefined set of categories, such as rent rolls, asset and liability ledgers, balance sheets, and market research reports.

### LLM-based text extraction for documents

Extracted text from documents using a large language model (LLM) trained on a massive dataset of legal and financial documents. LLM extracts text by identifying patterns and generating the text that matches those patterns.

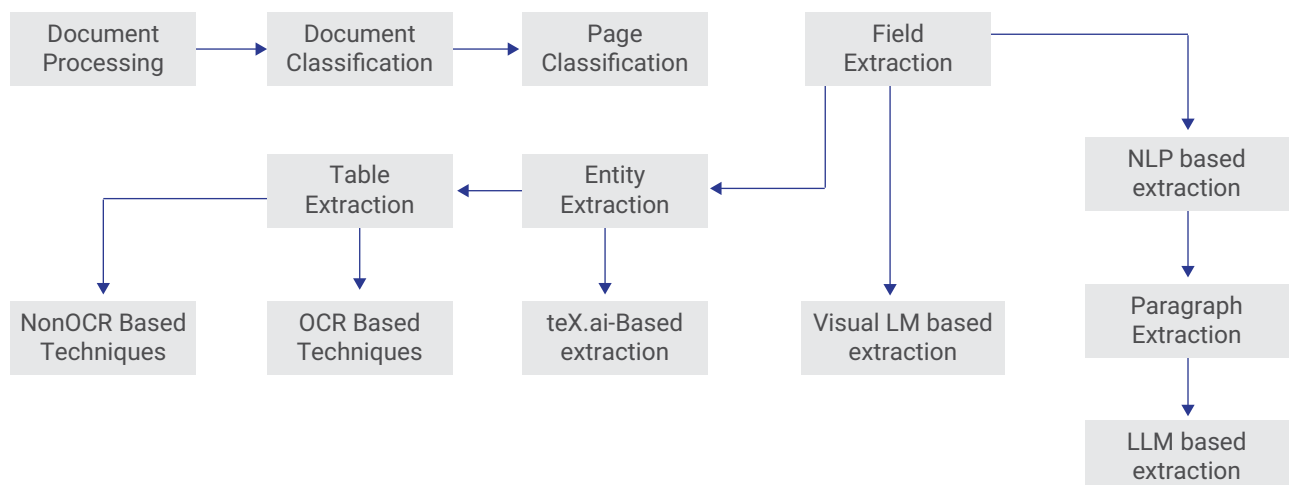
### teX.ai implementation

Leveraged teX.ai—a powerful, customizable preprocessor engine to extract, summarize, and classify the required documents.

### TROCR/Detectron2 table extraction

Leveraged TROCR and Detectron2 (open-source ML frameworks) to automate table extraction from documents. They identify the table cells and then extract the text from the cells.

## SOLUTION HIGHLIGHTS



## BUSINESS IMPACT

- The adoption of the Indium solution has resulted in an impressive 16-fold reduction in the time required for document processing.
- With 60% of document extractions being automatically handled by Indium's solution, tasks that previously required significant human intervention can now be completed in an hour and with reduced effort.
- Automation capabilities have led to a substantial decrease in human error during the document extraction process. This improvement has resulted in higher data accuracy, reduced rework, and minimized compliance risks associated with incorrect data handling.
- The capacity to extract data from 345 fields across 15 distinct document types enables greater analysis, insights, and decision-making, ultimately contributing to more informed business strategies.
- Indium solution's ability to support multiple languages has enabled the client to seamlessly process documents in different languages.
- Business teams now have access to accurate, relevant, and up-to-date information that supports their decision-making processes, fostering a data-driven culture within the organization.
- The ability to quickly process documents and access accurate information enables the company to respond quickly to market changes, customer demands, and emerging opportunities.
- Efficiency gains, reduced errors, and access to comprehensive data have collectively contributed to an increase in the company's agility.

## TECH STACK





# ABOUT INDIUM

Indium Software is a fast-growing Digital Engineering company, focused on building modern solutions across Applications, Data, and Gaming for its clients. With deep expertise in next-gen offerings combining data and applications, Indium offers a wide range of services including Product Engineering, Low-Code development, Data Engineering, Ai/ML, Digital Assurance, and end-to-end Gaming services.



## USA

Cupertino | Princeton  
Toll-free: +1-888-207-5969

## INDIA

Chennai | Bengaluru | Mumbai | Hyderabad  
Toll-free: 1800-123-1191

## UK

London  
Ph: +44 1420 300014

## SINGAPORE

Singapore  
Ph: +65 6812 7888

[www.indiumsoftware.com](http://www.indiumsoftware.com)



For Sales Inquiries  
[sales@indiumsoftware.com](mailto:sales@indiumsoftware.com)



For General Inquiries  
[info@indiumsoftware.com](mailto:info@indiumsoftware.com)

