

# CASE STUDY: HERETIK

## Automating the contract process with document annotation

### OVERVIEW

Heretik, a legal machine learning company, turned to CloudFactory to accelerate training of its contract review application. Our team successfully labeled thousands of legal documents for various elements, including contract structure, key entities, and clauses.

#### Services Used

- + Natural Language Processing
- Managed Workforce



#### Web

heretik.com



#### Industry

Insurance



#### Headquarters

Illinois, USA



#### Company Size

11–50

### MEET OUR CLIENT

When Andy Abbott was selling the first company he co-founded, he discovered how cumbersome the contract review process was. Could there be a more effective way to wade through the unstructured data encapsulated in contracts, leases, employment agreements, and other legal documents?

That was the impetus for Heretik, a lightweight yet powerful contract review application that pairs machine learning technology with workflow capabilities and integrates with existing contract management tools to reduce the time it takes to extract key information from a contract by two-thirds.

To train the AI models Heretik developed, Abbott turned to CloudFactory.

### THEIR CHALLENGE

“We are dealing with very complex corporate agreements, and there’s some structure to them,” Abbott explains. “But there’s a lot of nuances and differences between each agreement. We realized that we needed to identify certain elements in order to train our models effectively. We also realized that we wanted to bring a very general solution to the market, which required us to have training data across many areas of law.”

With hundreds of thousands of documents to label in order to train the model, Heretik looked for data labeling vendors. Abbott nixed the crowdsourcing option because legal documents require a certain level of domain knowledge and reliability to tag correctly.

He wasn’t necessarily looking for people who already had that knowledge. Instead, he wanted a group that could be trained and would stick with the work. “As opposed to a crowdsourcing solution where people roll on and off pretty frequently,” Abbott says.

### OUR SOLUTION

Heretik chose CloudFactory for its managed solution approach. Working with its Kenya-managed workforce, CloudFactory selects employees that are best suited to this type of work and provides customized training

## STATS

# 50%

REDUCTION IN DATA  
PREPARATION TIME

# 1,500

LEGAL DOCUMENTS  
ANNOTATED PER MONTH

### OUR SOLUTION CONTINUED...

(with client input) that helps them reach a basic level of domain knowledge. The team works with agile methods including daily sprints and feedback loops.

**“We wanted to work with a group to build domain knowledge so they could become more effective as time went on. And that’s why CloudFactory’s managed workforce was the best approach.”**

In the last 20 months, the CloudFactory team has labeled more than 30,000 documents involving five use cases including contract structure (cover pages, recitals, table of contents, definitions, signature pages and areas, exhibits, etc.), key entities and clauses, and other components of legal documents.

“I think a lot of people in startups love to tackle everything themselves. They think it is frowned upon to receive outside development help or data annotation assistance. But before you decline to seek outside help ask this question, **‘Can outside assistance help me get that product to market faster?’** I have always leveraged external teams, especially in areas that I don’t have the needed skill set. I don’t know how to effectively run a large data and labeling team. If I’m attempting to do that, there’s going to be a lot of issues that I’m going to come across. Whereas if I just outsource that and look to experts in the field, we don’t have to worry about those hiccups as I would have otherwise.”

“We’ve been able to significantly accelerate our data science research. That has sped up product development, especially early on where we cut out half the time it took to do some initial training of the data.”

– Andy Abbott  
Heretik Co-founder and CEO

## THE RESULTS

Abbott says that data scientists tend to spend upwards of 80% of their time on a typical data project performing tedious tasks related to training the data such as annotation and organization. He estimates that **CloudFactory’s help has cut data preparation time by more than half.** “We’ve been able to significantly accelerate our data science research. That’s sped up product development, especially early on where we cut in half the time it took to do some initial training of the data,” Abbott says.

Abbott has also been pleased with the CloudFactory team. “They’ve been on top of things and very responsive when we needed them.”

“CloudFactory’s help has been impactful in getting our product to market. It’s helped us get customers and revenue sooner than we probably would have without their help.”