




Data Search and Discovery in Insurance

An Overview of AI Capabilities



A collaboration between Iron Mountain and Emerj



IRON MOUNTAIN®



emerj

ARTIFICIAL INTELLIGENCE AND THE INSURANCE SECTOR

An Introduction

Historically, the insurance industry [has collected vast amounts of data](#) relevant to their customers, claims, and so on. This can be unstructured data in the form of PDFs, text documents, images, and videos, or structured data that has been organized for big data analytics.

As with other industries, the existence of such a trove of data in the insurance industry led many of the larger firms to adopt big data analytics and techniques to find patterns in the data that might reveal insights that drive business value.

Any such big data applications may require several steps of data management, including collection, cleansing, consolidation, and storage. Insurance firms that have worked with some form of big data analytics in the past might have access to structured data which can be ingested by AI algorithms with little additional effort on the part of data scientists.

The insurance industry might be ripe for AI applications due to the availability of vast amounts of historical data records and the existence of large global companies with the resources to implement complex AI projects. The data being collected by these companies comes from several channels and in different formats, and AI search and discovery projects in the space require several initial steps to organize and manage data.

Radim Rehurek, who earned his PhD in Computer Science from the Masaryk University Brno and founded RARE Technologies, points out:

“Theoretically a lot of applications are possible with AI in search and discovery, but in practice (in real world business situations) data is highly messy and businesses need to understand what might be possible, given their data constraints.”

A majority of the data that insurance firms collect is likely unstructured to some degree. This poses several challenges to insurance companies in terms of collecting and structuring data, which is key to the successful implementation of AI systems.

Giacomo Domeniconi, a post-doctoral researcher at IBM Watson TJ Research Center and Adjunct Professor for the course “High-Performance Machine Learning” at New York University, mentions structuring the data as the largest challenge for businesses:

“Businesses need to structure their information and create labeled datasets, which can be used to train the AI system. Yet creating this labeled dataset might be very challenging apply AI and in most cases would involve manually labeling a part of the data using the expertise of a specialist in the domain.”

Businesses face many challenges in terms of collecting and structuring their data, which is key to the successful implementation of AI systems. An AI application is only as good as the data it consumes.

Natural language processing (NLP) and machine learning models often need to be trained on large volumes of data. Data scientists tweak these models to improve their accuracy. This is a process that might last several months from start to finish, even in cases where the model is being taught relatively rudimentary tasks, such as identifying semantic trends in an insurance company's internal documentation.

Most AI systems necessarily require the data to be input into an AI system in a structured format. Businesses would need to collect, clean, and organize their data to meet these requirements. Although creating NLP and machine learning models to solve real-world business problems is by itself a challenging task, this process cannot be started without a plan for organizing and structuring enough data for these models to operate at reasonable accuracy levels.

Large insurance firms might need to think about how their data at different physical locations across the world might be affected by local data regulations or differences in data storage legacy systems at each location. Even with all the data being made accessible, businesses would find that data might still need to be scrubbed to remove any incorrect, incomplete, improperly formatted, duplicate, or outlying data.

Businesses would also find that in some cases regulations might mandate the signing of data sharing agreements between the involved parties or data might need to be moved to locations where it can be analyzed. Since the data is highly voluminous, moving the data accurately can prove to be a challenge by itself.

QUESTIONS TO ASK

to determine if your business's data is ready for use with AI

Is your data accurate? As mentioned at the start of this report, AI systems are largely dependent on the quality of the data on which they are trained. For example, if the addresses and names in a dataset for claims forms do not match, the output for any analysis on this data will be erroneous.

Do you have the right kind of data? The data must be relevant to the objective of the AI system, and the data necessary to achieve the objective must be present in the datasets that are fed to the AI model.

Is all of your historical data consistent? Businesses need to ensure that their data is consistently collected in the same format. Unstructured documents, such as handwritten inspection records, might be inconsistent in their formats and content. Insurance companies might need to ensure that the inspection records for a medical case lists the patient's name, address, and medical history details consistently before feeding the data into a machine learning model.

Does your data need to be updated? The data being fed into an AI system changes in most use cases, and the system may need to be updated. Businesses need to update their data regularly to ensure accurate insights are being derived from their AI system. A machine learning model trained on old data might provide less relevant or wholly inaccurate outputs than one updated frequently with new data. For example, a machine learning model trained on old data might not be able to detect newer methods of claims fraud.

AI FOR SEARCH AND DISCOVERY

There are several marquee commercial offerings for using NLP and machine learning for data search and discovery. One notable development here was [Google launching its Google Cloud Search](#) enterprise tool, which the company claims allows organizations to search and access their internal information more effectively.

For example, employees in an enterprise could type in questions such as, “Where are the docs shared by Mary Brown,” or, “Who is Bruce’s manager?” and Cloud Search could potentially return answer cards with the relevant information.

Insurance companies can benefit from AI search and discovery in a similar manner by finding better ways to search for information in both structured and unstructured datasets. NLP and machine learning can be used to develop smart search tools that understand context and jargon specific to the insurance industry. Radim shares his experience working in AI projects with insurance companies and adds:

“A lot of documents in the insurance industry are unstructured in the form of handwritten notes or PDFs. The current demand from insurance companies is around using this data to replace legacy systems. Improving existing processes is the focus of these companies, as opposed to finding new ways of executing a certain process.”

A large amount of text-based data is collected by insurance firms in many different languages from an array of sources. This could include smartphone apps, insurance claims forms, emails to and from customers, social media data, insurance adjuster notes,

medical health records, police statements, and underwriter notes. This trove of data could likely be used to help insurance firms identify trends, spot new opportunities, and detect fraud.

For example, Insurance companies might have thousands of signed claims forms or contracts coming in every day. These documents need to be checked to see if there were any changes to the original. Traditionally, this task is done by humans, but machine learning could certainly handle this.

In certain applications, a component of human-aided training might be required, in which the NLP software “learns” by human review. Human operators might:

- Create or clean the input data
- Measure the accuracy of the output
- Create additional training or meta-data

In large-scale systems, these review processes can be repetitive, and using human review might not be feasible. In such cases, businesses can:

- Create an interface to speed up the review process
- Use a crowdsourcing solution like CrowdFlower for large-scale data tagging services
- Update the existing business process system to include the human review at a stage before the data is sent to the algorithm

We look at some of the use-cases where AI is being applied for data search and data discovery in the insurance sector below.

USE-CASES Applications of AI in insurance

Insurance industry processes seem to be emerging as one of the larger applications for artificial intelligence. Traditionally, the industry has been dominated by legacy systems and large brands whose processes and product offering haven't evolved much.

The insurance industry checks off several boxes that might indicate readiness with regards to AI adoption, such as a large trove of historical data and the existence of large businesses with the capital and resources to successfully implement AI.

AI could help insurance firms use the data collected by them from multiple channels, such as forms on their website, records of live chat with customers, emails, posts on social networks, documents and reports generated by insurance agents and customer care operatives.

All of this data is relevant to certain aspects of the insurance business, such as claims processing or customer service. Ensuring that this barrage of incoming data can be effectively utilized for AI ap-

plications requires the proper collection methods, storage, and management.

One of the key applications for NLP and machine learning in finance seems to be in ensuring the right kind of data is sent to the right department inside the firm. For example, an email from a customer citing a complaint needs to be sent to the complaints team for effective and fast response. Overlooking or missing such emails due to human error might eventually result in low customer satisfaction or even loss of revenue in some cases.

NLP could help insurance firms with the challenge of timely and accurate data access by automatically clustering and understanding the context of data from different channels. NLP and semantic search algorithms can ingest a lot of this insurance data, and over time accurately route the documents to the necessary departments in insurance firms.

According to Rehurek, there are two benefits to search and discovery applications in insurance for data management, as he explains with an example:

“Customer Support or complaint requests can come through many channels, such as emails, phones, the company’s website etc., and these need to be routed to the right departments. NLP and machine learning can help insurance firms classify these incoming requests by understanding their context.

At the same time, the customer service team can also gain useful information, such as... the top complaints from customers and maybe develop an FAQ for the most common ones.

Transforming legacy workflows like routing the complaints saves the company costs in terms of number of human hires. But Analytics and clustering of this data is what really drives revenue.”

For instance, one company already applying AI to identify business opportunities in auto insurance is Progressive Corporation. The company seems to be using machine learning to interpret automotive driver data to track market trends and identify business opportunities, such as allowing customers to personalize their insurance policies to their driving habits.

According to Pavan Divakarla, Data and Analytics Business Leader at Progressive, machine learning algorithms are starting to help the company better understand customer data in making predictions about what will happen in the insurance marketplace.

[Insurance Europe reports](#) fraudulent insurance claims accounts for 10% of all claims expenditures in Europe. The Association of British Insurers estimated that around \$2.48 billion worth of fraudulent claims goes undetected each year in the UK. The French Insurance Federation reported detecting around \$195 million worth of fraudulent claims in 2011.

AI software could help insurance companies mine data from insurance applications or claims forms. Insurance adjusters usually inspect property damage or personal injury claims to determine the payout to policyholders. It is common practice for these adjusters to keep notes on their inspections. These notes are often handwritten—in other words, unstructured. NLP could help insurers identify claims that might be fraudulent by searching for red flags within adjuster notes.

NLP could also really help insurers in identifying organized fraud. NLP-based software could identify similar phrases or sentence structures in the descriptions of incidents from several different claimants. These

kinds of patterns are difficult and highly time-consuming for human insurance officers to identify.

AI solutions designed to detect insurance fraud might help insurance companies upskill their human fraud investigators and augment their capabilities. NLP could help insurers identify more potential fraud cases than humans and, over time, improve the accuracy with which they detect fraud. An NLP solution would in turn allow human fraud investigators to focus their time on more serious and expensive cases of fraud.

Another challenge that many companies face lies in the fact that social media data available to them is in the form of ambiguous conversational language. As a result, any insights that might be generated from reading thousands of these posts would be hard to identify. Improvements in NLP-based text mining software with insurance applications could allow insurance companies to gain useful insights from unstructured social media data in the next two to five years, including an understanding of customer sentiment surrounding their brand.

Ensuring that each claim has complete, accurate, and up-to-date information is a highly repetitive task for human investigators at large insurance firms. These firms often receive thousands of claims forms every day, and they need a way to process them faster. Claims processing is one of the most common applications for which large insurance companies are applying AI.

For instance, healthcare insurance claims largely include text-based information in the form of descriptions entered by call center operators and notes from adjusters about specific cases.

These health insurance claims contain text relevant to standard treatment procedures and diagnostic terminology, including medical abbreviations and acronyms. NLP can be used to rapidly mine information from these claims, identify missing information, and classify and route the claims to the relevant department.

There already exist B2B AI vendors that offer NLP software they claim can transcribe a customer's speech during a phone call with an insurance agent and automatically fill out that customer's claim form. This software could also automatically contact customers through email for information that is incomplete or filled out incorrectly on the claim form. This might enable insurance firms to help customers file claims faster and more accurately. That said, most of these AI solutions currently require a human manager to review the decisions they make.

Large insurance firms which deal with thousands of claims forms everyday might find AI-enhanced automation of claims processing helpful for reducing costs and improving accuracy because the scale of doing so manually might make it difficult for human analysts alone.

Another challenge for insurance firms looking to automate their claims processing is that the historical data needed to train the AI might still be in the form of paper documentation. Machine vision solutions and optical character recognition technologies can help insurance firms digitize these paper documents automatically and structure them in a machine-readable format.

A select few AI vendors today also offer a more end-to-end solution for insurance firms that allows for document digitization and uses NLP to find and retrieve information from documents in a contextual manner.

Belgian insurance firm Ageas [recently partnered with an AI vendor](#) in a pilot project aimed at automating the visual appraisal for motor claims. Ageas reportedly used a machine vision software to automatically assess vehicle damage from images by using historical claims payout data and the images attached to those claims that showed vehicle damage. Ageas's press release states the company might see both cost and time benefits using the AI solution it implemented.

AI-DRIVEN BUSINESS VALUE

The insurance industry might be slightly unique due to the fact that most customers only gauge the performance of an insurance firm when they make a claim and need to be paid and not when they invest in a policy. This makes customer service critical for insurance firms.

Many AI applications for improving search and discovery in insurance, such as processing claims faster or automating the routing of incoming data, might be challenging for large corporations. As Germán Sanchis-Trilles, who holds a PhD in Computer Science from the Polytechnic University of Valencia and co-founded Sciling, put it:

"In large corporate organizations, data is usually spread across several departments in different formats and types. Consolidating all of that data to create labelled training feedstock for AI algorithms might prove challenging. At large scales, this might be hard for companies to do and hence we see the existence of data lakes—where each software is trained on a compartmentalized dataset."

Experts we spoke to seem to agree that we might start to see insurance companies lead industry on how they can

leverage and structure their data. Insurance companies already collect large amounts of data. AI could help with digitizing and indexing all of this data for better search and for finding patterns that might not be easily visible to human analysts.

Although NLP and computer vision can be used to extract information from unstructured data, insurance companies might want to upgrade their databases and aggregation systems with an eye on what might be possible with structured data in the future.

The several steps involved in data management including collection, digitization, segregation and storage might be challenging for insurance businesses to automate. With several vendors offering AI services for each of these steps individually, it might make the integration process lengthier in addition to throwing up compatibility issues.

Some vendors offer complete solutions for data search and discovery including document digitization, data management, storage and analysis. Insurance firms can leverage AI across their business processes using historical data, even if the data is on paper, and gain insights to help reduce cost and time, improve accessibility and security for these processes.

About Emerj

Emerj is where business leaders turn to understand how AI is impacting their business and industry and what to do about it. We're the industry source for authoritative market research and competitive intelligence for the business applications of artificial intelligence.

Our objective, jargon-free research and industry overviews are designed to give executives and decision-makers exactly what they need for competitive insight, informed AI technology procurement, and strategic planning around AI.

AI disruption will open up new vistas for opportunity, creating new business paradigms and making others extinct. Emerj is the best place for leaders to get the facts and strategic insight they need to stay on the winning end of AI's inevitable disruptions.

Emerj offers custom research and thought leadership for brands that are innovating in artificial intelligence in various sectors and industries. If you're interested in credible, objective thought leadership in artificial intelligence, email us at services@emerj.com.

About Iron Mountain

Iron Mountain Incorporated (NYSE: IRM) is the global leader for storage and information management services. Trusted by more than 225,000 organizations around the world, the company helps customers lower cost and risk, comply with regulations, recover from disaster, and enable a more digital way of working.

Iron Mountain InSight is a content services platform that provides actionable business insights and predictive analytics through Machine Learning (ML)-based classification of a company's physical and digital information, which adds structure, context, and meta-data to information to make it more usable. The resulting enriched content can then enable enhanced automated governance and workflow throughout an oil and gas organization.