



# Data Search and Discovery in Energy and Oil & Gas

---

An Overview of AI Capabilities



A collaboration between Iron Mountain and Emerj



# ARTIFICIAL INTELLIGENCE AND THE ENERGY SECTOR

## An Introduction

The [increasing demand for energy, oil, and gas](#) puts pressures on businesses in these sectors to improve the efficiency in their production and business processes. One way oil and gas companies might be able to do this would be to organize their internal data in ways which make it easy for employees to search and discover patterns that could optimize their business processes.

The energy sector 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.”

Businesses face many challenges in terms of collecting and structuring their data, which is key to the successful imple-

mentation 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 oil and gas 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.

Recent advances in machine learning and NLP for data search applications have made it possible to extract structured information from free text, such as that found in well or reservoir reports and logs. However, the largest challenge involved in applying these techniques to unstructured data in the oil and gas sector seems to be the lack of access to labeled training sets that can help AI models understand industry specific jargon and terms.

Most data search and discovery applications for AI in the energy and oil and gas sectors would require businesses to ensure their data is prepared for input into a machine learning model.

Germán Sanchis-Trilles, who holds a PhD in Computer Science from the Polytechnic University of Valencia and co-founded Sciling, believes that consolidating all the data into a single format might prove challenging for most businesses:

“Depending on the size of the company the problems with integrating data into single format are a big challenge. Large scale organizations might have data of several types/formats different locations. This ends up forcing companies to create ‘data lakes’ where each AI software has its own dataset to work with.”

The number of total barrels produced or sold by a given plant in a year might be relatively simple to measure and track. Specific reports about the performance of a well or nuanced and intricate sensor data from heavy equipment might be more challenging to unify, particularly if there are different company norms for data collection and management at different plants or in different regions. A change in company culture can be harder than the technical hurdles of harmonizing data.

In the energy and oil and gas sectors there are also other challenges with data in terms of security and access, privacy, regulations and compliance, IP protection, and physical and digital barriers. Large oil and gas companies might need to think about how data is being collected in their various sites across the world, what regulations might affect access to data from a certain physical location, or what digital firewalls might constrain access to data.

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, a machine learning model might be trained on data from sensors in an oil and gas refinery to predict operational efficiency. If the model is trained on data sets that have incorrect data from a malfunctioning sensor, the output of that model will be altered and likely inaccurate. Insuring that only accurate data enters an AI system is critical.

**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. Using the same example as above, any missing sensor data in the datasets used to train the machine learning model would again result in deviations from the desired output of predicting operational efficiency.

**Is all of your historical data consistent?** Businesses need to ensure that their data is consistently collected in the same format. For example, If operators of different wells in an oil field filled out drilling reports in different formats, the data would need to be restructured to one uniform format before a machine learning model could be trained on it.

**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. For example, if an oil and gas company outfitted their refinery with new equipment that had additional sensors (such as vibration sensors) or that functioned differently than older equipment, the company might need to update the data accordingly for an AI system to make sense of the new working environments.

# 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.

Giacomo Domeniconi, a Post-Doctoral Researcher at IBM Watson TJ Research Center and Adjunct Professor for the course “High Performance Machine Learning” at NYU, lays out some of the possibilities of AI for data search and discovery:

“One of the bigger improvements I’ve seen in the last two or three years has been in utilizing unstructured data to gain insights. Unsupervised machine learning techniques might help businesses make sense of their unstructured data without the need for massive amounts of labeled datasets.

Image recognition might now move on to video where an algorithm might be taught to understand actions or movements being performed in videos.

NLP is also being applied for sentiment analysis of news or social media.”

Techniques such as semantic search can help identify and extract facts, attributes, concepts, and events from unstructured content for analysis. For example, an oil and gas company might be able to gain insights that can help reduce procurement costs by using NLP and machine learning for data discovery on historical supply chain documents, tenders, and bids from raw material suppliers.

It might be possible for oil and gas companies to use AI for applications such as patent analysis, customer sentiment analysis from call center notes, survey feedback, and online forums, or procurement analysis from supply chain contracts and other bid documents.

Oil and gas companies might need to track trends in government reports, geological studies, or social media in order to win against competitors. NLP can help automate and aggregate news and information.

Such AI applications come with technical challenges and data considerations as well. For instance, energy and oil and gas literature usually contain industry specific jargon, such as well depth or types of hydrocarbon. Subject matter experts in the field will likely need to work alongside data scientists to build a useful AI system.

Search and discovery AI might also prove highly valuable to oil and gas companies in ensuring that their legal and regulatory compliance processes are optimized to reduce costs and avoid penalties. Compliance teams on energy and oil and gas companies might need to comb through thousands of documents to keep track of existing and new regulations.

AI search and discovery operations can be used to democratize access to the legal and regulatory compliance information according to Richard Downe, who holds a PhD in Electrical and Computer Engineering from the University of Iowa, and is Director of Data Science at Loblaw Digital. He suggests that oil and gas companies can also use NLP and machine learning to generate queryable regulatory databases, making it easier for their employees to access the information. Downe states:

“Search engines like Google have shown that it is possible to create contextually aware search by studying the query that users typed in, the relationships between the words typed and the prior search history.

This can be applied on a much smaller scale to domain specific legal context to create a searchable database that augments the capabilities of human compliance officers. For instance, if a user typed in “obviousness,” the AI search might emphasize results that are relevant to the meaning of that word in a specific sub-domain of law, such as patent law.”

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 energy and oil and gas sectors below.

## USE-CASES Applications of AI in the energy and oil and gas sectors

The oil and gas industry is usually divided into three major operational sectors: upstream, midstream, and downstream. Upstream involves the exploration and production of oil and natural gas. Midstream usually refers to transportation and storage stages. Downstream encompasses the various processes involved in refining and selling oil.

In the upstream, there may be opportunities for applying AI to optimize the operating conditions of rigs dynamically, leading to reduced production costs and minimizing the downtime of machinery. Alejandro Betancourt, who holds a PhD in Interactive and Cognitive Environments from the Eindhoven University of Technology and leads the analytics team at Columbian oil and gas company Ecopetrol, suggests that oil and gas companies can potentially improve their enhanced oil recovery (EOR) rates by using NLP, computer vision, and machine learning for drilling and exploration data.

According to Betancourt, the oil and gas industry has collected data over the years that might have been curated and analyzed by human experts, making it ripe for feeding into AI systems. This data might include exploration production and reservoir data logs, such as seismic surveys, well logs, conventional and special core analyses, fluid analyses, static and flowing pressure measurements,

pressure-transient tests, periodic well production tests, records of the monthly produced volumes of fluids (oil, gas, and water), and records of the monthly injected volumes of EOR fluids (water, gas, CO<sub>2</sub>, steam, chemicals).

Oil and gas companies could potentially gain insights from the above mentioned data sources, such as the ideal operating conditions for each well under different working conditions. Over a period of time, AI software might ingest the curated data records and condense information from data sources, such as those mentioned above, which might be in the form of structured documents, PDFs, hand written notes, audio or video files. This type of software might let oil and gas companies search and discover information from all their data irrespective of channel or type to gain the most accurate search results. Alejandro adds:

“Due to the amount and complexity of the data, the petrotechnical databases (such as well logs) are difficult to navigate and search. I see a future with smart petrotechnical databases which can help geologists easily find data for a particular well/reservoir by using a couple of intuitive commands. Additionally these smart databases might also be very capable in analysis of this data.”

[There is some evidence for natural language processing and machine learning being applied to advanced data search applications.](#) Extracting structured information from unstructured free text documents, such as handwritten maintenance notes or drilling reports, might be possible with NLP.

For example, in [a report from four researchers](#), including PhDs from Stanford University and Texas A&M University, we found evidence of NLP being used to extract information from an oil and gas firm's drilling reports. The researchers claim to have tested their methodology on 9670 drilling reports from 303 wells in an oil field.

The report offers a methodology for using NLP to automatically classify sentences in old drilling reports from oil and gas companies and identify patterns in the oil and gas companies' operational behavior. Oil and gas drilling reports are usually unstructured documents that might contain information about oil well status, the kind of drilling that is being done, or oil and natural gas production figures. Seeing this information displayed over time might help companies determine clear patterns that seem to correlate with efficiency or inefficiency.

The report claims the tool developed by the researchers can be used offline by an energy company that might want to gain insights from historical drilling

reports for identifying positive operation patterns. Additionally, they claim this use case might evolve into a real-time decision support system, where oil and gas companies might be able to reduce drilling costs associated with malfunctions or system downtimes.

The researchers used daily reports with information about equipment inspections, which were created by well operators. Their algorithm could reportedly extract information from wells that were currently operating and wells that had equipment breakdowns.

The NLP model could classify sentences that are present in drilling reports as EVENTS, SYMPTOMS, or ACTIONS. Oil and gas companies could use this data to identify the best practices to reduce downtimes and consequently improve production efficiencies.

The core business value that this use case might drive for oil and gas companies lies in cost reduction and mitigating accidents by identifying patterns in drilling reports for the most successful operational practices. This would then allow companies to avoid less optimal practices.

This use case might also be used by oil and gas companies to find patterns between text extracted from drilling reports and other well data, such as rock samples or well logs, to create comparable technical reports for each of their wells.

Quick access to records and information might be vital for large oil and gas companies that operate out of multiple locations across several countries. In many cases these might include older physical records to be managed along with the digital records. This makes information access a challenging problem for larger firms due to issues such as missing or incomplete records, records in several different formats, or unindexed records.

AI search and discovery tools can help oil and gas companies digitize records and segment the information based on geographical location, country, or project. Visual representations of such data might potentially lead to identification of issues, such as pipeline erosion or increased equipment usage.

For instance, an oil and gas company can digitize all their seismic records and other exploration and production (E&P) data to create a searchable database. AI software might help identify patterns in the data that indicate if an oil well asset might pose economic risks to the company.

In the oil and gas industry, data such as seismic or subsurface charts are usually in the form of image files. Apart from taking geologists a long time to study, these files might be stored in different locations on-premises, cloud storage, or even microfiche. AI search and discovery tools can help find and access information from these image files in a single

place, allowing oil and gas companies to engage in analytics projects.

[Total Oil recently announced an agreement with Google Cloud to jointly develop an AI system to analyze subsurface data aimed at improving their E&P processes.](#) According to the announcement, the partnership might lead to the development of AI systems that can interpret subsurface images from seismic studies using computer vision and automate the analysis of technical documents using NLP. The business value driving this application would be in augmenting the capabilities of the company's geologists and reservoir engineers enabling them to work more efficiently.

The project will reportedly bring together geologists from Total and machine learning experts from Google Cloud at Google's Advanced Solutions Lab in California. The project's long term goal seems to be developing an AI assistant for geoscientists at Total that can aid with information search or potentially even recommend suggestions to improve operational efficiency in various oil and gas processes.

Total claims to have had a history of applying machine learning to exploration and production activities, including predictive maintenance for turbines, pumps, and compressors at its industrial facilities production profile forecasting, automated analysis of satellite images, and analysis of rock sample images.

[According to Accenture](#), categorization of the costs of different goods and raw materials are not usually organized well in oil and gas companies. Traditionally, companies use human employees to classify the company's total spend manually or with a rules-based software.

Both these methods are labor intensive. On average, a typical set of annual spending data from an oil and gas company might contain a total of 800,000 to 1.2 million line items with 200,000 to 400,000 unclassified line items. It is common in the oil and gas industry to collect data on spending only every two to four years due to the labor intensive nature of the task, which means that companies have less data.

Many global energy and oil and gas companies might find that this problem is compounded by the fact that their business units (including oil wells and corporate offices) in different locations might input spending data, such as vendor names or item descriptions, in different formats.

NLP and machine learning might enable oil and gas companies to easily search and track spending data from unstructured reports. Algorithms like the one from SAS might be capable of automatically categorizing the unclassified line items in annual spending reports.

The core business value for this application seems to lie in freeing up valuable human time and resources from a labor intensive classification task.

In downstream operations for the oil and gas industry, corrosion by crude oil is a huge risk for equipment failures. Depending on the chemical composition of the crude or the environment in which it is stored, veteran corrosion engineers can devise methods to avoid equipment downtimes.

Digitizing this knowledge and delivering maintenance insights to new engineers might now be possible with AI. For detecting oil corrosion risks, companies can use NLP and machine learning develop a searchable database of maintenance information from data such as refinery incident reports and physical properties different types of crude. This data can be structured or, in most cases, unstructured in the form of word documents or PDFs.

An NLP system could parse through maintenance data to aggregate knowledge from veteran corrosion engineers,

for example. Engineers could use an NLP-based search system to find information about how to avoid corrosion for different types of crude by interfacing with a dashboard.

Inspection repair and maintenance are areas where oil and gas firms collect vast amounts of data, a majority of which is usually unstructured or even hand-written. NLP and text mining can be used to extract, classify, and correlate the vast amount of knowledge gained by engineers over the years.

This is especially important in the oil and gas industry due to the Great Crew Change, a phenomenon that refers to the large age gap in the oil & gas workforce, where most engineers and geoscientists are either over 55 or under 35. This means that the experience of the veteran engineers is not always fully transferred to the next generation of engineers, and AI data search and discovery might be a strong fit for resolving this issue.

## USE-CASES Applications of AI in the energy and oil and gas sectors

## AI-DRIVEN BUSINESS VALUE

---

AI-enhanced data search and discovery might soon be applied in the oil and gas industry to applications across many business functions, such as actual production and maintenance, front-end sales, and customer services. As Alejandro put it:

“Production optimization, downtime minimizing, reservoir understanding and modeling, and finally commercial support by combining internal and external data sources will be the areas where AI will drive maximum business value for oil and gas companies.

There are plenty of measurable business results in successfully applying data analysis in Oil & Gas particularly in reducing equipment downtime, deferred production, and improving the overall efficiency of systems. There are also two crucial KPIs that companies can track namely the processing time and analysis time, several of the AI applications will target these two areas.”

Experts we spoke to seem to caution business executives that despite the enormous opportunities for AI in the oil and gas industry, it is a challenging area for many companies in the space due to the technical complexity of the analysis and the computational resources required.

Improving production efficiencies or gaining operational insights from drilling data might well be the “low-hanging fruit” applications for AI in the energy and oil and gas industry. Applications of AI that will truly drive business value from the perspective of business leaders in the oil and gas or energy industries would be those that help in improving sales or closing better deals with existing customers. Corporate departments in the oil and gas industry might end up being the largest end users for AI simply because of the direct value being driven from improved sales.

## 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](mailto: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.