expert.ai

Applying Natural Language Understanding Artificial Intelligence to **Complex Insurance Problems**

Applying Natural Language Understanding Artificial Intelligence to **Complex Insurance Problems**

Insurance companies consume massive amounts of data via their underwriting, contracts, and claims processes. While this poses an ongoing challenge for insurers, they have found both opportunity and value in the variety of artificial intelligence (AI) technologies available to them.

AI can include anything from image and voice recognition to robotics and predictive analytics, but it is natural language understanding (NLU) that has proven most applicable to the complex tasks associated with corporate and commercial insurance.

Natural Language Understanding

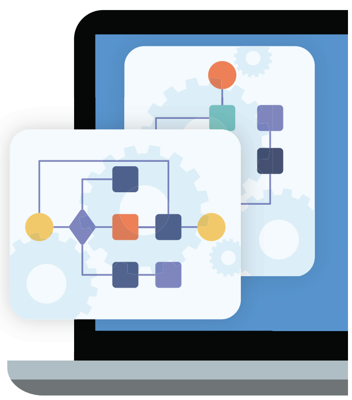
NLU is a subset of AI that refers to computing systems that execute processes requiring a human-like level of comprehension. This can include tasks like reading unstructured (e.g., articles, posts, emails) or semi-structured (e.g., invoices, order forms) documents as well as pattern recognition. For everyday situations related to insurance, such as evaluating risk, investigating claims, examining contracts or implementing corporate governance, NLU is the relevant AI discipline to consider. The closer the system can come to matching the decisions of a well-trained human, the better.

Machine Learning and Rule-Based Learning

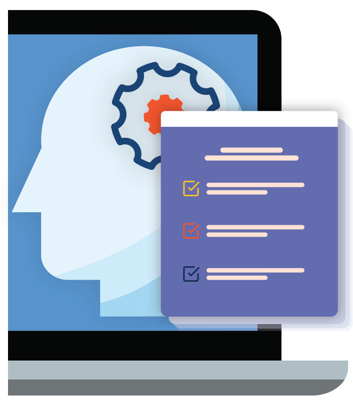
NLU approaches can be roughly classified into two main groups: symbolic "rule-based learning", and statistical "machine learning" (ML). Rule-based learning is often referred to as "linguistic" or "semantic" technology, because the system must understand language and extract meaning like a human. These technologies can support most use cases but require that rules and domain-specific knowledge of language be programmed to enable the system to properly understand content.

Statistics-based ML technologies are a more recent addition to AI, and they can only be applied if certain requirements are met. **Pure ML systems are not programmed with rules or logic on how to understand language.** ML samples input data in high volumes, and through statistical analysis, gradually discern which are the good outcomes. Learning is a long and computationally intensive process. Importantly, **ML systems are only as good as the breadth of learning data they consume.**

Recent advances from academic research, combined with the computational advances of cloud computing have resulted in a proliferation of machine learning systems and attention. However, setting the wrong expectation for ML not only undermines its potential, but risks slowing down investment in a technology that shows great promise.



Machine Learning



Rule-Based Learning

Complexity and Sample Data

NLU technologies for unstructured information management have proven valuable across many industries for applications such as content enrichment, process automation, and conversational AI.

For the insurance industry, NLU technology is being applied most notably to processing unstructured documents to extract, organize and highlight important data. Its ability to mimic human understanding when reading unstructured information makes it especially valuable for insurance applications.

To identify the best technology or methodology to use in a typical application of NLU technologies on unstructured information, use cases must be classified by two factors:

Complexity

When classifying NLU applications, complexity refers to the difficulty and ambiguity of a specific task, even if a human was to do it. This can be measured, for example, by the scope of questions a chatbot is required to answer, or how nested the knowledge to be extracted is and the level of reasoning need to apply to reach a conclusion from the analysis of documents in a claim automation project.

Completeness of Sample Documents

This second factor is a proxy for how applicable pure machine learning would be to a specific scenario. An ML system must be trained, and its performance depends on two critical factors:

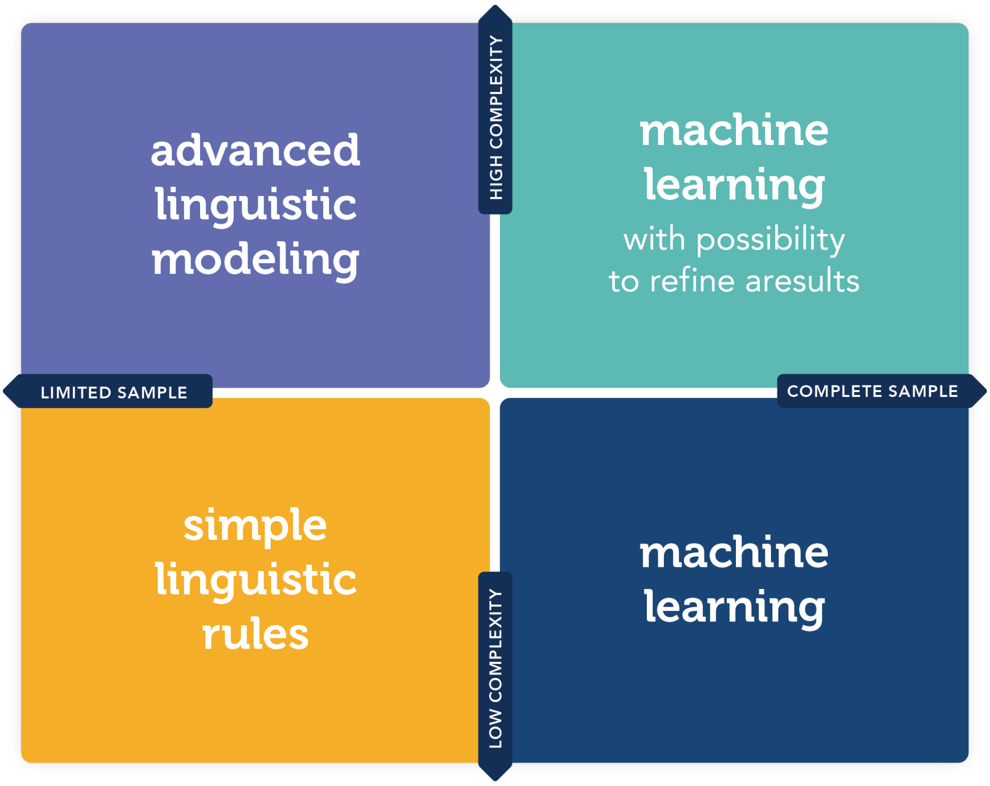
- The quantity of documents in the training set, and
- The representativeness of the training data and documents to the specific use case.

Completeness of the ML training set is a measure of these two factors. A project that is a true candidate for pure ML will have both a large training set and a balanced coverage distribution across all the possible outputs.

Typical Insurance Use Cases for Natural Language Understanding

As explained before, NLU technologies for automatic management of unstructured information can be roughly classified into two main groups: linguistic/semantic rule based, and statistics/machine learning based.

Linguistic rule-based technologies can support any of the different use cases independent of complexity and availability of relevant sets of sample documents. Statistics- and ML-based technologies are instead applicable only to use cases in which you meet the requirements of high quantity of documents and complete coverage of the problem space. Even so, for a given quantity and coverage of learning, the performance of ML systems tends to degrade as complexity is increased.



CAPABILITIES OF DIFFERENT NLU TECHNOLOGIES

Based on this balance between complexity and completeness, organizations should evaluate which NLU approach provides the highest possible business value.

LOWER RIGHT QUADRANT:

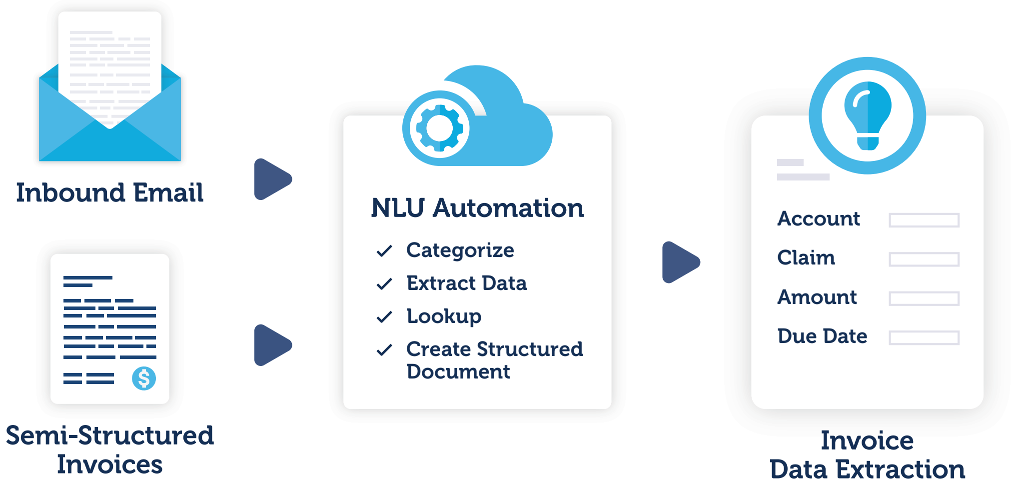
Best Fit for Machine Learning

The combination of low complexity and a highly complete sample marks the business cases that are the best fit for pure machine learning techniques.

For insurance, an example for this scenario is the management of supplier invoices by the purchasing department. Usually, the process consists of extracting data (name, dates, quantities, payment terms) from documents that, even if different from one supplier to the other, are semi-structured, meaning they contain the same data and the difference in format is minimal.

In this case, we have a significant number of sample documents. In addition, there is a relatively small number of data elements to be extracted and the differences in the format are limited. These factors — large number of sample documents and low complexity — make it the ideal scenario for pure ML.

In this scenario we can select a big enough sample to perfectly represent each type of invoice. In addition, the abundance of samples ensures that if the results of a single iteration are unsatisfactory, the sample size is large enough to retrain the system until you reach the point of best possible results.



INVOICE MANAGEMENT EXAMPLE

UPPER RIGHT QUADRANT:

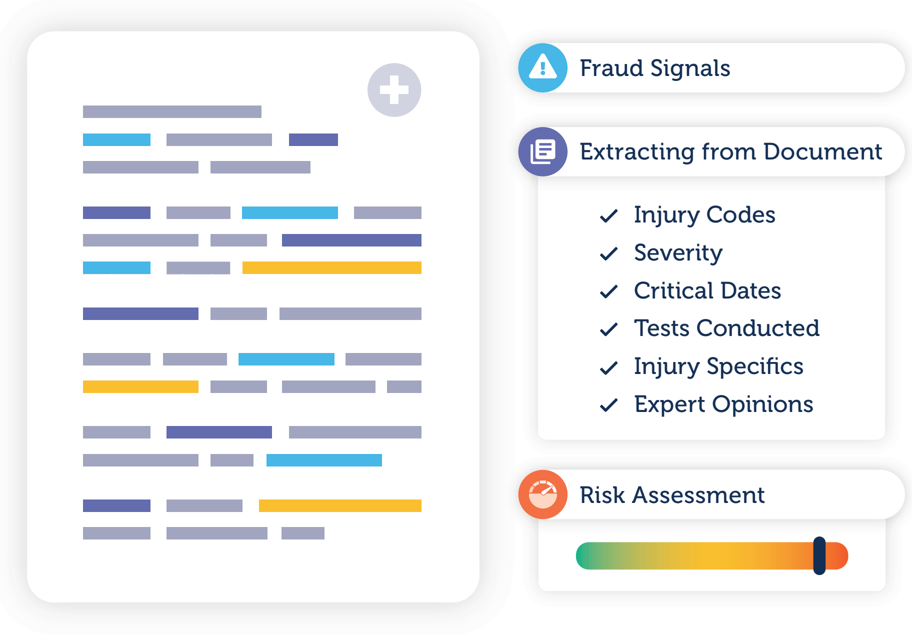
Machine Learning with the Possibility to Refine Results

Moving to the top right of the diagram, we still have many well-distributed sample documents, but the subject matter is much more complex. This use case is for projects that deal, for example, with a rich and variable set of questions for a question answering system or a rich and deep taxonomy, or system used to classify documents, for an automatic categorization task.

The higher complexity means that it is more difficult to identify samples that uniquely represent each node of the taxonomy or desired outcome of the automation task. **The training data sample may have the quantity, but the nature of the task makes complete coverage impractical.** This makes the pure ML model much more difficult, or practically impossible, to train to the level of performance required.

Insurance examples would include unstructured medical reports for accident claims. While there are standard terms and patterns that can be specifically identified, documents from outside sources are unstructured and content varies widely. Achieving the required level of accuracy here requires additional rule-based learning beyond the range of pure ML training on the input sample.

The practical implication of this is a requirement to essentially give the ML process some knowledge that it cannot attain purely by analyzing the samples. In the example, injury codes, critical dates, tests and results are detected by ML analysis, even from an unstructured input document with an unknown format. More subtle signals such as fraud indications and expert opinions depend on rule-based logic to augment the ML data sample.



MEDICAL RISK ASSESSMENT

Machine Learning is a Black Box

This case highlights an important aspect of pure ML systems that is less advertised or understood: ML systems are black boxes. This means that you cannot apply your "human" understanding of the phenomena that you are trying to automate to improve the results. For machine learning systems, there is no tool with which to refine the algorithm.

With pure ML, the only option you have is to feed more training examples to the system and depend on the algorithm to adjust results statistically over time. Unfortunately, this doesn't always guarantee that you will improve your results or reach the level of accuracy required.

Ideally the use cases in this quadrant could still leverage the advantages of ML. If only you could adjust the model...

Maintenance

To provide a complete picture of the different AI techniques and approaches, it is important to emphasize that these systems cannot be configured and then just left without any further attention. Once the level of accuracy and performance required is reached and the solution is implemented, different factors can contribute to the deterioration of its performance. Thus, it is important to regularly monitor the results and to plan for maintenance activities.

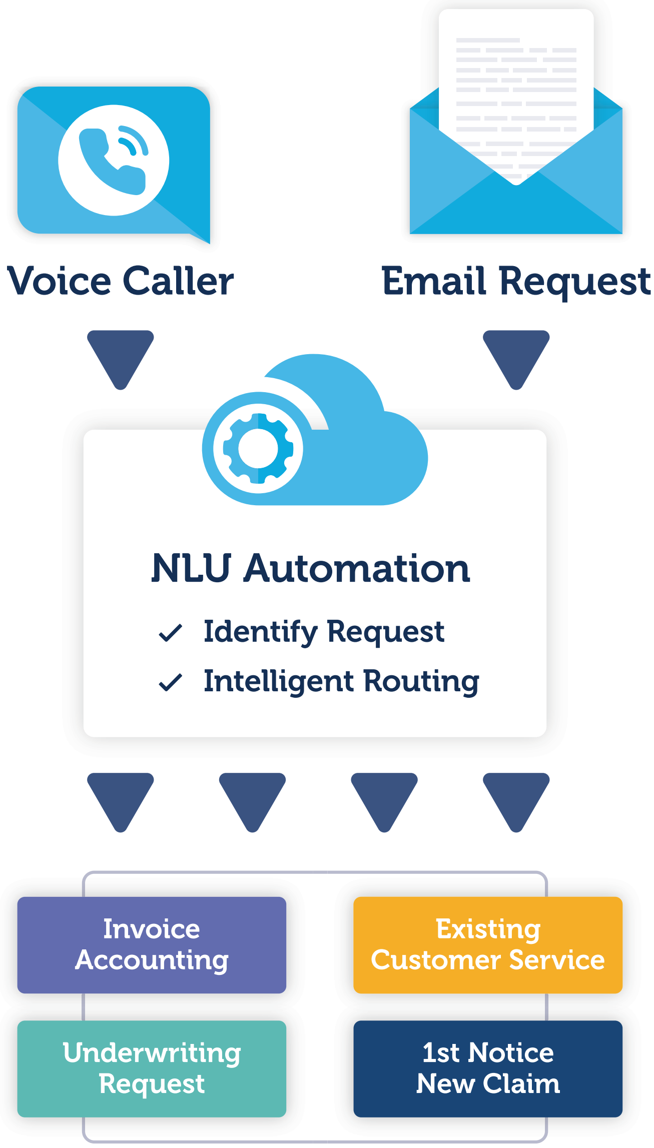
Different approaches must be taken when maintaining a semantic/linguistic technique versus a pure ML technique. The semantic/linguistic approach generally allows for incremental changes to the rule set that, in most situations, are sufficient in replicating the original performances. Instead, the ML approach usually calls for a retraining based on a training set that is representative of the changed scenario.

It is not the objective of this paper to provide an analysis of the advantages or disadvantages of one technique versus the other, but it is important to note and understand that these maintenance activities should be considered when looking at the total cost of ownership of these systems.

LOWER LEFT QUADRANT:
Simple Linguistic Rules

This scenario does not require a very sophisticated linguistic engine. Shallow linguistic systems or even simple 'bag of words' classification tools are sufficient enough to reach a reasonable quality. The relative simplicity of the taxonomies or extraction objectives allows you to configure the rules, usually with good results.

Robotic process automation is the most common application of this scenario. These scenarios are less interesting for business and commercial insurance today though because, due to their simplicity, the automation value has already been achieved.



ROBOTIC PROCESS AUTOMATION
ROUTING TO CORRECT DEPARTMENT

UPPER LEFT QUADRANT:
Advanced Linguistic Modeling

There are several complex scenarios that involve small sample sets distributed in a non-uniform fashion. For these use cases, you need both a linguistic engine sophisticated enough to ensure deep understanding of content and a set of tools powerful enough to ensure the development and effective application of advanced linguistic rules.

Because of the complexity here (i.e., homogeneity between documents to be classified or data to be extracted), **simple keywords and Boolean operators are ineffective in distinguishing the differences between similar concepts.**

Risk analysis reports are a good example of this highly complex, low-volume scenario. Engineering analysis of risk factors in an industrial or large public venue, or the comparison of policies and contracts to a standard are common situations for commercial and corporate insurance underwriting. In most cases, a human underwriter must spend many hours, or even days, reading, comparing and evaluating these complex documents.

A rule-based system that embeds specific industry and company knowledge can automate this analysis for you. However, risks such as sensitivity to language differences across documents, and the sparse coverage of possible outcomes in each sample make this a poor candidate for ML.

EXTRACTING & RATING RISK FACTORS

Another example of a high-complexity, low-volume scenario is one related to policy review. Insurance carriers are often asked to review and analyze nonstandard policies — especially at the time of negotiation or renewal — to evaluate unexpected exposure. An NLU-based system enables insurers to streamline this manual, error-prone process by using predefined checklists to identify implicit and explicit coverages and exclusions. The complexity of these documents and the limited number of examples make it a good case for the top left quadrant of our diagram.

POLICY REVIEW
RISK IDENTIFICATION

Insurance Recommendations for Natural Language Understanding AI

There is not a single AI technique that can ensure high performance for every insurance situation. With that said, commercial and business insurance generally lie in the high complexity range given the high value and risk associated with each policy and the unique conditions to consider.

For complex insurance cases, machine learning offers great efficiencies for business cases that fall under the scenarios on the right side of the diagram (large training set), provided there is a way to steer the ML training appropriately. **A solid semantic NLU technology integrated with machine learning capabilities is the ideal way to address insurance needs and achieve desired results from NLU computing Investments.**

Advanced semantic understanding systems built specifically for the insurance industry already have the native flexibility to address all the diverse use cases in each area of the matrix. So, why not apply ML functionalities selectively where it will add the most value? The desired combination is a technology stack that provides a combination of rule based NLU technology capabilities and ML based algorithms to address the most common use cases for unstructured information.

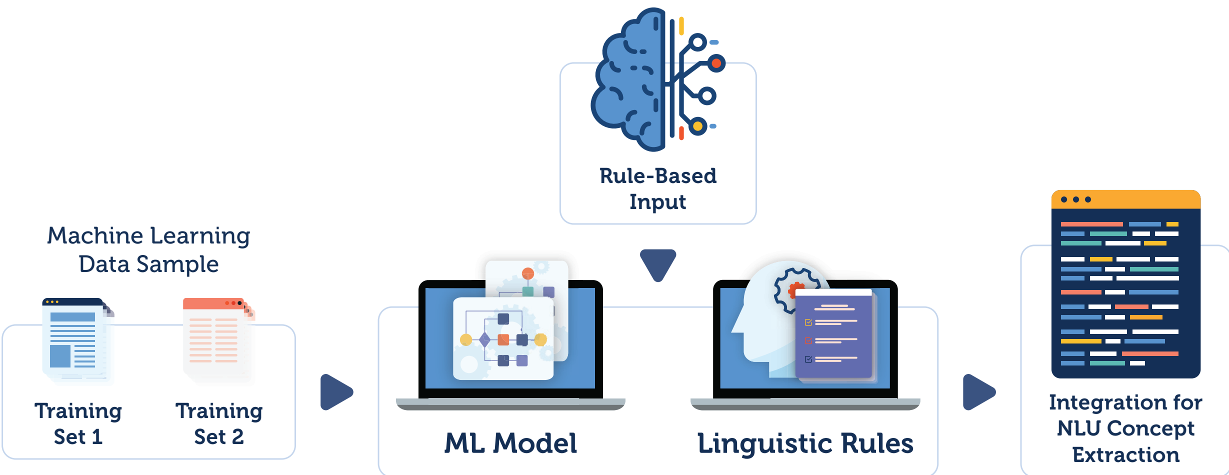
Essential Elements to Combine Include:

- Statistics-based machine learning environment as a starting point
- Deep, human-like understanding of written text
- Insurance-specific knowledge graph based on industry experience and use cases
- Visibility of the rules created through ML algorithms to make results explainable, eliminating the ML black box
- Ability to use rule-based algorithms to add precision to the learning algorithms directly, rather than only by adjusting sample data

Moving Forward with Hybrid NL

The ability for NLU to mimic human understanding has the potential to significantly improve decisions in insurance underwriting, contracts and claims. While pure ML systems provide value for consumer-facing and simple automation efforts, rules-based systems with deep "linguistic" understanding of unstructured information are necessary to address the risk and complexity of corporate and commercial insurance.

A better alternative is to deploy a hybrid system that can adapt to the variations of coverage and complexity present across a corporate and commercial insurance environment. Rather than employ a separate technology for each application, build all applications on a single system that combines rule-based and ML properties when needed.



HYBRID SYSTEM
ML MODIFIED BY RULE-BASED INPUT

The ideal NLU solution is one that combines the automated learning and breadth of an ML system with the human precision and industry knowledge of a rule-based one. This hybrid system requires that you can explicitly change rules based on organizational knowledge, rather than depend on the system to learn solely from examples.

Get
Started

Interested in learning how NLU AI will transform your insurance company? Get started here.

See what expert.ai can do for you!

expert.ai

About us
Expert.ai is the premier artificial intelligence platform for language understanding that augments business operations, scales data science capabilities, simplifies AI adoption and provides the insight required to improve decision making throughout organizations. The expert.ai brand is owned by Expert System (EXSY:MIL), that has cemented itself at the forefront of AI-based natural language solutions across Insurance, Banking, Publishing, Defence & Intelligence, Life Science & Pharma, Oil Gas & Energy, and more.

www.expert.ai **info@expert.ai**