Data Accuracy & Human-in-the-loop

Data extraction

As a leading provider of contract data extraction services, Brightleaf understands the importance of machine learning models for improving accuracy, efficiency, and cost savings. However, we also know that having a human in the loop is critical for ensuring the reliability of the extracted data and the ethical implications of these models.

Machine learning algorithms are incredibly powerful and can make predictions that were once thought to be impossible. But, they are only as good as the data they are trained on. Without high-quality and unbiased data, these algorithms can produce inaccurate results, which can be costly and even harmful in sensitive sectors such as healthcare, finance, and legal.

A report from McKinsey & Company found that companies that used human-AI collaboration outperformed those that used either humans or AI alone. The report found that the combination of human expertise and AI capabilities led to better decision-making and improved business outcomes.

"Why can't we simply use better algorithms to gain higher accuracy?"

While using better algorithms can certainly improve the accuracy and confidence of machine learning models, relying solely on algorithmic improvements may not be sufficient. This is because algorithms alone cannot account for the complexities of human language, culture, and context. These are all factors that require human expertise, which is why having a human in the loop is so important.

Here are a few reasons why relying solely on algorithmic improvements may not be sufficient:

Quality of data: The accuracy and reliability of the output of a machine learning model depend heavily on the quality of the input data used to train the model. Poor-quality data can result in biased or inaccurate models, and even the best algorithms will be unable to overcome these limitations.

Complexity of the problem: Some problems are inherently complex, and even the most advanced algorithms may not be able to solve them with high accuracy. These problems may require the integration of multiple algorithms, domain-specific knowledge, and human expertise to achieve optimal results.

Interpretability: Some machine learning models are inherently difficult to interpret, and the factors that influence their decision-making process may not be clear or transparent. This can be particularly problematic in applications where ethical or legal considerations are at play.

Human feedback and oversight: Machine learning models operate in complex environments with ever-changing data and inputs. Incorporating human feedback and oversight can help to ensure that the model remains accurate and ethical in its decision-making.

> **Constant feedback can help machine learning algorithms learn and improve their results over time. However, the human-in-the-loop (HITL) is still essential in machine learning projects to ensure the algorithms' accuracy, reliability, and ethical use.**

Human-in-the-Loop (HITL) approaches are becoming increasingly important for machine learning (ML) applications because acquiring high-quality training data can be difficult and frequently requires human expertise. Human experts collaborate with machine learning algorithms to provide feedback, quality control, and domain expertise at various stages of the ML workflow in HITL approaches.

In contract data extraction, for example, HITL approaches in ML model selection and evaluation involve human experts who provide domain knowledge and assess the performance of the ML models.

HITL can assist in addressing the limitations of relying solely on machine learning algorithms, such as data quality, problem complexity, and interpretability. By providing human oversight and expertise, HITL can also help to improve the accuracy and reliability of machine learning models. Overall, HITL is becoming increasingly important for machine learning applications because it allows for the incorporation of human expertise and domain knowledge into the ML workflow, improving the accuracy and reliability of ML models.

About Brightleaf Solutions, Inc.

Brightleaf provides a technology-powered service to extract information using our own proprietary semantic intelligence/natural language processing technology, our own team of lawyers to check the output, and our own Six Sigma process to deliver end-to-end, highly accurate, extracted data.

Your legacy contractual documents along with this extracted meta-data can be migrated into a Contract Lifecycle Management (CLM) system for tracking and reporting. This greatly enhances the value of your investment in the system. The data can be virtually anything, and it is customized for each of your types of contracts. All meta-data, terms and conditions, legal provisions, and even all obligations (which are usually scattered throughout your contracts) are extracted by our software. Our team of lawyers checks each-and-every extracted data point against the original documents using a stringent Six Sigma quality process, which delivers highly accurate results **(download Strategy Brief)**.

This allows you to leverage the knowledge in your existing contracts, report on the extracted data, even recover hidden revenue (e.g. by policing penalty clauses in your supplier agreements) and comply with current and upcoming regulations.

Request Demo