Artificial Intelligence in Due Diligence

How to make it work in **real life**

imprima | White Paper

Published June 2019

Working in due diligence

Junior lawyers, analysts and associates working at financial or legal advisory firms in M&A look forward to rewarding careers. Whilst the rewards are high for those who make it, this line of work can be extremely demanding. Working against the clock in high-stakes environments, these individuals work long hours and sacrifice weekends; with work/life balance taking such a hit, it is important to identify the culprit – the lengthy, laborious and sometimes inefficient due diligence (DD) process. Juniors are starved of both their time and exposure to more in-depth analysis thanks to the lengthy review of agreements, financial documents and contracts. Not only can the workload be suffocating, its mundane, monotonous nature is a catalyst for human error. So not only are individuals suffering, quality of work is too. That is not good news, either for the development of the professionals doing the work or the client experience.

There is a great deal of talk about technology's role in addressing this issue, specifically Artificial Intelligence (AI) and Machine Learning (ML). The latter is the subset of AI that is often touted as the answer to due diligence challenges – and that is because it turns that same repetitiveness of the work into an advantage: it is fuelled by what it learns from mundane, repetitive tasks.

What's preventing AI from helping?

Many see great potential for AI and ML to play an important role in DD. But there is a major hurdle preventing the adoption of AI in this space: it is not clear how to apply the technology in live DD situations effectively. We will discuss that issue in this paper, sharing with you how we have applied ML and what the results have been.

It is important to recognise that some see AI as a threat to their business model, as it may reduce billable hours or jobs, as manual work is automated. We do not believe that ML has to pose this threat and will touch upon this topic later in this paper and also, address it in depth in a separate paper to follow.

However, before we even begin to discuss those issues, let's see how AI and ML can be effectively deployed and applied within DD in the first place.

Breaking down the challenge

At Imprima we have been researching this topic for almost three years, are engaged in ongoing collaborations with two universities and have conducted our own research. Our traditional line of business (Virtual Data Rooms) gave us a great opportunity to meet with legal experts, M&A advisors and RE advisors: our clients.

These conversations, combined with our research, provided a powerful feedback loop that led to the innovative development of an AI toolkit for DD.

A huge challenge during DD is the time and effort it takes to identify documents that contain a specific characteristic, a certain issue, or 'red flag'. These characteristics can help shape the value of the deal (positively or negatively) but the nature of the work involved in unearthing them can lead to human error (see our Intro section). We wanted to explore how to address this challenge using AI and ML.

The question became, **"how can we use AI to help find documents of a certain type, with certain phrases, provisions or issues (red flags) that can impact the value of a deal, more quickly and accurately than is currently achieved?"**

Additional factors we considered when designing a solution:

1. The tool cannot demand a fundamentally new way of working – it has to fit into current processes and add value from the start.

The high-stakes nature of the task at hand means we can't ask people to change the way they work. They wanted a tool that would fit into their processes and begin adding value immediately.

2. To further complicate this, clients told us that they were not interested and had no time for training algorithms before using them. Traditional ML demands this.

This means our tool should learn how to operate based on typical DD work/tasks/ processes carried out by the people involved. The repetitive nature of these tasks gives the ML algorithm the opportunity to "learn" how to perform and to eventually start suggesting items/documents for the user to review.

3. Importing and exporting data between a VDR - the key repository of data used in DD - and AI tools is not a good option, also not when via an API.

The movement of data between applications is time consuming and cumbersome. Often, for those on the buy-side of a deal, it is not allowed because of data security concerns. Moreover, it will be impossible to address the two considerations above if our tool was not integrated into the VDR, as we will discuss below. That also holds when the data is exchanged between applications via an API, as that would limit the data being exchanged in terms of granularity as well as interactivity.

4. The AI results should be at least as accurate as manual review by lawyers.

This one's quite important…if the tool didn't improve on the current process in terms of speed, accuracy and effectiveness, what would be the point?

Early Findings

We discovered that we could build something that does not require complex ML techniques. In fact, the simplest solution that explains the data gives the best result. A more complex solution would only lead to a phenomenon called "over fitting": The risk would be that the ML tool might be trained to solve the problem for one data set, but not for others. In addition, it would lead to an unrealistic semblance of accuracy. This is discussed in more detail in our white paper of July of last year. <https://www.imprima.com/white-paper/white-paper-ai-streamline-legal-due-diligence/>

We realised the key to a successful ML tool is not in its sophistication, but in its access to data. This data is needed as 'fuel' to train the ML algorithm automatically – something that is critical to the success of the tool, based on what we learnt from clients (see the previous section!).

We soon realised being a VDR provider is a huge advantage. Why? Because all relevant data for DD and user information is stored and processed within the VDR. We then came to the realisation that this ML tool had to be built within the VDR if we were to satisfy all design considerations.

Building an ML Tool for DD

We wanted to ensure that all information the ML tool needs to be successful is given to it during a normal document review process by a DD professional – again, a key design consideration. From a sell-side perspective, that process might be document review to help select documents to include in the VDR, or whilst conducting vendor DD. From a buy-side perspective, it could be document review as part of the formal DD process.

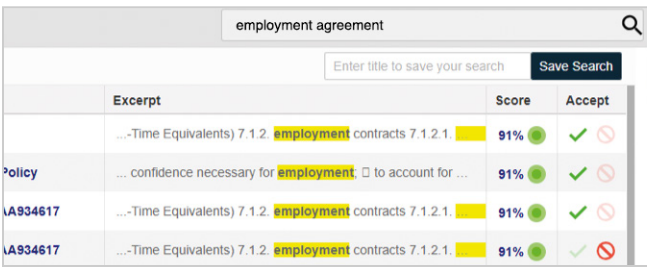
We decided to design the ML tool such that it is trained on the basis of information that is generated during these processes. Therefore, it will cost people no extra effort, time or money; no separate process needs to be set up to 'train' a tool or an algorithm.

To help fit into current process, we had to get to work with flexibility front of mind. Users of the tool simply enter a search term and start reviewing the documents presented (as they would anyway). And then, as the users review the document, they mark it as 'relevant' or 'not relevant' depending on the issue that is under review. All of this (document review and marking of documents as relevant or not) sits within a simple, intuitive user interface, like that of popular websites, to ensure it feels familiar.

Applying the Tool in a Real-Life Situation

Let's document how this looks in practice: A lawyer or analyst needs to review employment agreements, specifically those that do not contain a non-compete clause. This may happen because a new standard employment agreement has been introduced at the target company, but not all the old agreements have been revised.

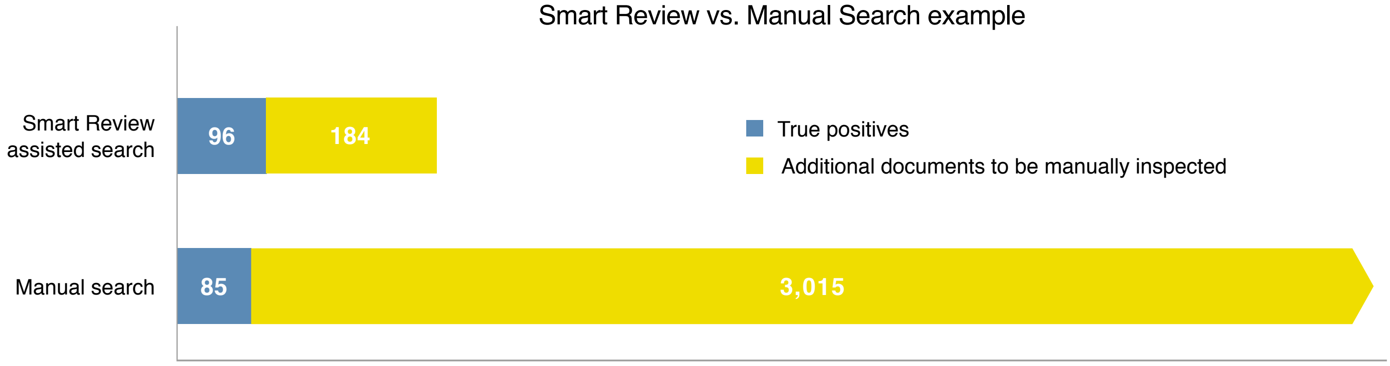
The ML-driven review process could begin, for instance, with the user typing, 'employment agreement' into the search bar. There's no need to put together complex "regular expressions" (such as: "employment" AND "agreement" AND NOT "non-compete"), as the Machine Learning will not rely on that. Additionally, such complex regular expressions would not result in accurate results anyway; the risk of missing documents (in other words, low "recall") would be very high.



The tool then observes what the user does with the results and learns from it. As the user marks documents as 'relevant' or 'not relevant' the ML tool develops its understanding of the user's needs. The tool then applies its enhanced understanding of the request and searches again, this time presenting more and more relevant results to the user.

The results…

Smart Review vs. Manual Search example



We tested the above on a set of approximately 3,000 of all types of legal documents. These documents included, amongst others, employment agreements, of which a small portion were employment agreements without a non-compete clause (note that the other legal documents also included employment agreements with a non-compete clause).

The test results showed that after reviewing a small portion of the documents, virtually all employment agreements without a non-compete clause had been found. Bear in mind, this was done using an algorithm that had not been trained in advance by anyone. Still it saved more than 90% of the user's time when compared to the traditional review process.

Additionally, the AI-powered search can be saved for future use, to find similar issues in similar agreements, which will lead to even greater time savings and accuracy.

We have found that many issues of a similar nature can be handled with this approach, such as finding credit agreements with a change-of-control clause, license agreements without a proper limitation of liability clause, etc. In general, documents of the same type that contain an issue, or a red flag.

Assessing the results

We found that with the tool we can retrieve relevant documents for complex queries with an accuracy ("recall") of 95% or higher. That is much more accurate than manual review: scientific studies (Grossmann and Cormack, 2011) show that the average for human review during DD is 85% or less.

It is also important to note, the accuracy achieved with ML is consistent: the tool doesn't succumb to fatigue or the monotonous nature of the task at hand – quite the opposite, it thrives on it.

Not only does it help the junior lawyers or analysts with achieving higher and more consistent accuracy, senior staff who need to review work of the people working for them are also helped. By checking (and correcting) the tagging of the documents as relevant or not, the ML algorithm is further trained to comply with those improved tags. It's interesting to consider that, when ML is employed, this senior-level review may be needed much less frequently, as the ML will "remember" what was corrected.

The impact on business models, efficiency gains

We should also attempt to address the sensitive topic as to whether ML could be a threat to legal or other advisory firms, or to their business model, or to jobs. It is a topic of much debate and concern lately. We believe however that, from a technical perspective, it does not have to be a threat, as ML may be used to increase value, rather than reduce cost.

For instance, a use case could be to apply the ML approach to increase the reliability of identifying high-risk queries by combining AI-based queries with exhaustive manual search. The lawyer or analyst would still manually, and exhaustively, review all agreements, but is assisted by ML to boost his or her accuracy of finding issues (“recall”), increasing confidence that all issues are found. In other words, it would lead to increased accuracy at the same cost.

Here’s another example where value could be added by using ML: In many cases, only the most important agreements need detailed review; for example, the agreements in place with a company’s top 20 to 50 customers. A client would then perhaps assume that the remaining customer agreements are similar and don’t contain additional risks. That could be a dangerous assumption, and one that does not need to be made if ML is used.

With the ML method we have developed, such a review could be enhanced by executing the following 2-step approach:

Step 1:

The top 20 – 50 agreements are reviewed manually with the support of our ML tool. As a result;

- Find information within documents such as those that present potential risk.
- The ML algorithm has been trained to perform a similar review in similar agreements

Step 2:

The remainder of the agreements can then be automatically reviewed, adding value to the law firm’s or advisor’s service at no additional cost.

On top of that, of course any automation of tasks through the application of ML should lead to Lawyers and Advisors being able to spend more time on in-depth and strategic analysis of the documentation and the target company itself.

Conclusion

1. ML leads to substantial time savings:

- ✔ **90% or more time saved versus traditional DD, or, in other words, professionals spend only 1/10th of the time they currently do, on review.**

2. ML significantly improves review accuracy:

- ✔ **Typically a recall of 95% or higher is achieved versus 50% - 85% in traditional DD.**

3. When implemented as part of the VDR workflow and integrated within the VDR, you do not need to spend time training algorithms in advance of their use:

- ✔ **Additionally, the tool provides full flexibility in how users can search for issues.**

4. Complex issues can be found with simple queries, ML does the hard part.

5. Data remains in a secure environment – the VDR:

- ✔ **No import or export is necessary.**

In a future whitepaper, we will be showing how AI can be used to automatically categorise and index documents uploaded to the VDR. This automates the laborious and lengthy VDR prep phase. Additionally, we will discuss our technology for automatically generating summaries of agreements.

www.imprima.com

Contact us today to find out more:

London
Tel: +44 20 7965 4700
E: londonsales@imprima.com

Paris
Tel: +33 1 58 36 06 60
E: parissales@imprima.com

New York
Tel: +1 646 844 2993
E: newyork@imprima.com

Frankfurt
Tel: +49 69 915 0980
E: frankfurt@imprima.com

Amsterdam
Tel: +31 207 155 600
E: amsterdam@imprima.com

Sydney
Tel: +61 283 110 266
E: sydney.sales@imprima.com

Official London Stock Exchange, Marketplace Provider:

London
Stock Exchange

Imprima is proud to have been selected by the London Stock Exchange to provide due diligence services to corporations listed on the Exchange. This reflects the quality of our technology, services and customer support.