Imprima

AI that really works – For M&A, Real Estate and Legal Professionals

Occam's Razor vs. GPT-3

Pieter van der Made — Imprima

There is a clear need for AI

As a Due Diligence Technology provider, Imprima has worked with numerous law firms and in-house legal teams. They all share the same concern: repetitive manual work diminishes productivity and decreases job satisfaction.

So, how are law firms and legal teams tackling the challenge? AI, and machine learning, in particular, is a real focus because it's the obvious solution. It has the potential to automate repetitive manual processes and the great thing is, the more repetitive a process, the more effective ML can be. At the same time, humans experience this type of work as tedious and monotonous, which leads to reduced motivation and increased human error throughout the review process.

But is machine learning reliable?

So if we accept that AI and ML in due diligence have such potential, how do we apply it in an effective way? Firstly, its output needs to be accurate and reliable. If you cannot rely on the accuracy of AI in due diligence, you will still need to double-check its results manually, and no time is saved whatsoever. So, is it reliable?

Often none or very little information is provided by the suppliers of AI tools. Sometimes a "Recall" of 90% is claimed, which means that 10% of what you're looking for, is not picked up. For example, say there are 100 instances of a specific clause you're searching for, 90 would be picked up by AI, and 10 would not.

But is that enough? And with which "Precision" has that been achieved? Let's explain what is meant with Precision first: it is a measure of how many false positives are produced (by the AI algorithm) (as opposed to Recall, which is a measure of

the false negatives) [1]. In summary, think of it this way; 'Recall' is a measure of how many things were missed by AI and 'Precision' is a measure of how many things were incorrectly identified by AI.

Recall-precision tradeoff

Obviously, you want both Recall and Precision to be high: High Recall would mean that you miss very few relevant items (documents or clauses or paragraphs, or whatever you are looking for). At the same time, we don't want AI to include incorrect items that are not relevant for you (Precision).

However, in practice, with any AI, there is a tradeoff between Precision and Recall. If you really want to be sure you miss no or very few documents, you will have to investigate more documents. If you really can't afford to be looking at any non-relevant docs, you will likely end up missing more relevant docs.

In any case, statements about Recall are meaningless without specifying the Precision at which that Recall can be achieved. It is easy to achieve a high Recall if you allow the Precision to be very low. It is even easy to achieve 100% Recall, if you allow the Precision to be 0%: that would mean the algorithm just returns all items (docs or clauses etc.), and leaves it to the user to find and select the relevant items. Childish argument? Maybe, but it does make the point that specifying Recall without specifying Precision is meaningless.

Word2vec

Fun fact on Google's word2vec: it translates words into "vectors", mathematical representations of the semantic meaning of the words. As a result, you can add and subtract the words. Does that work? As it turns out: yes: classic example: "king" minus "man" plus "woman" becomes "queen".

Occam's razor

So, to optimize both Recall and Precision, surely the most sophisticated AI algorithms are required? As it turns out the answer is no. It's rare that you'd need to throw the 'deep-learning kitchen sink' at your problem. In many cases not actually.

Despite what many will tell you, it is not about the most sophisticated and nuanced algorithms. There are no silver bullets, no proprietary algorithms that make the difference. Moreover, there is an abundance of highly sophisticated ML algorithms out there in the public domain, open-source implementations even. For instance, Google has published their famous word2vec algorithm.

The point is to choose the right algorithm for the problem at hand. At Imprima we let ourselves be guided by "Occam's Razor" [2], which means, in essence: use the simplest solution that fits your problem, as that is, in all likelihood, your best solution.

Translating Occam's wisdom to modern-day ML technology: For some problems, you need a neural net. And sometimes

linear binary classifiers are the appropriate choice. For other classes of problems simple regular expression do the trick, despite some deeming that to be "outdated" technology. In a way GPT-3, though extremely promising and impressive [3], is in a way a great example of the applicability of Occam's razor. As quoted in a recent article in The Verge [4]: "... GPT-3 was only capable of automating trivial tasks that smaller, cheaper AI programs could do just as well …".

At Imprima, we indeed use a variety of techniques, Neural Nets, linear classifiers, regex, depending on the problem at hand. We are not ML purists. We try to solve our clients' problems in the best possible way. In a way that is reliable, and where the user is in control. **See also our white paper on XAI below**.

GPT-3

GPT-3 is in an enormous Neural Net, with as many as 175 billion (!) parameters, trained on a large portion of documents found on the internet. It predicts text most likely to follow input text (e.g. an answer to a question).

It seems to have great potential, but also has its flaws, making it unreliable. As quoted in MIT Technology Review: "GPT-3 often performs like a clever student who hasn't done their reading trying to [BS] their way through an exam. Some well-known facts, some half-truths, and some straight lies, strung together in what first looks like a smooth narrative." [5]

So how accurate is that?

At Imprima we strive for – and attain in our tests – 95% Recall with 10% "Effort" (Effort meaning the number of docs you need to inspect to hit that level of Recall, so an alternative to Precision for measuring false positives). We'll show you this in the graph below.

Below we show the result of an experiment we did, where we tried to find 100 Credit Agreements that included a Change of Control clause, in a data set of 3100 agreements (where the other 3000 agreements also included Credit Agreements without a Change of Control clause).

As you can see in the Recall-Effort graph below (an alternative way of looking at Recall-Precision trade-off), 95% Recall is achieved at about 5-6% Effort. As you can see from the graph, at c. 13% Effort you can even achieve almost 100% Recall, and 80% Recall can already be achieved at 3% effort.

What does this mean for you? 10% Effort equates to saving 90% of your time. And that is achieved by starting to review only a handful of documents, and telling the machine which docs you want to see more of. You do not have to train the Machine Learning algorithm in advance: the algorithm will just pick up your behaviour, and train itself. For the user it is really very simple. But it works!

In the next white paper, we will discuss how well it works in a variety of situations, in any language, for any kind of problem (not just pretrained clause searches), and that it can also deal with the issue of "Transfer Learning", i.e. the issue that a model trained on one data set is not applicable for another, slightly different data set.

95% recall or 90% recall; what's the big deal!?

Note that 95% Recall may look similar to 90% Recall at first sight. In fact, it is not. 90% Recall means you miss 10% of the items that you are looking for (docs, clauses, whatever it is that you are looking for). 95% Recall means you miss only 5% of the items that you are looking for. So, in fact, 95% Recall is twice as accurate as 90% Recall.
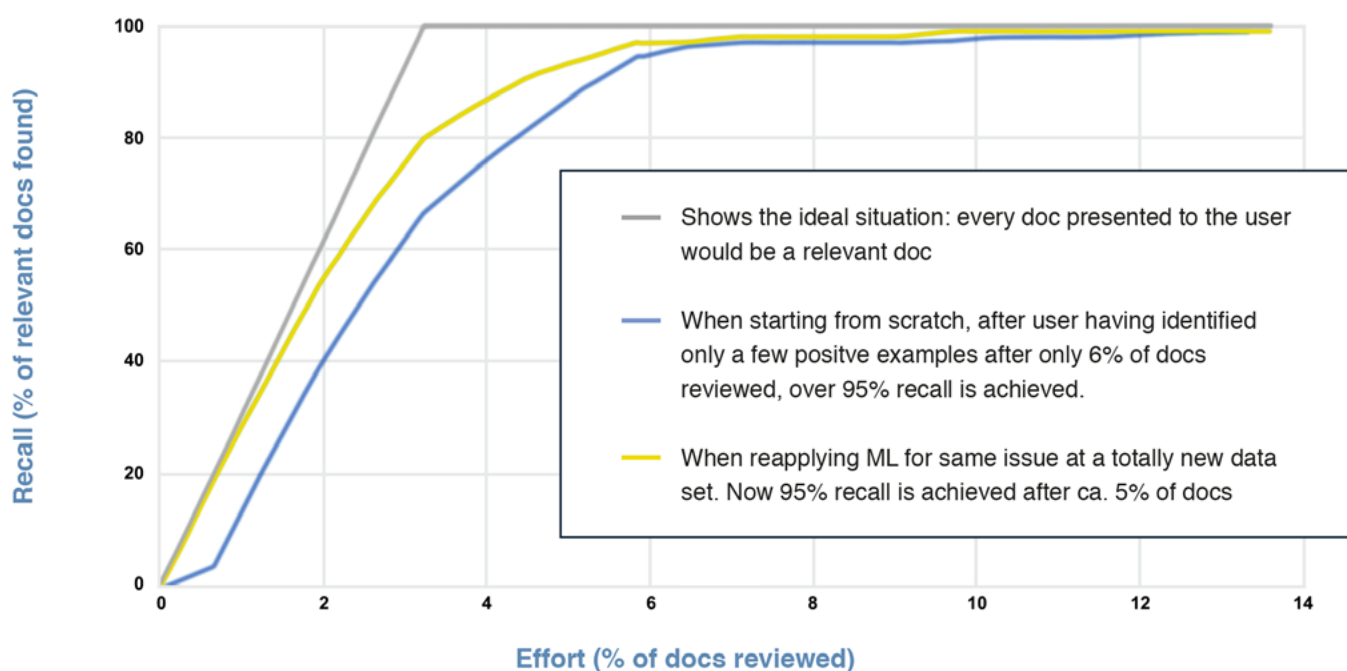
Credit agreements with change of control (image source: proprietary research conducted by Imprima)

Citations

1   https://en.wikipedia.org/wiki/Precision_and_recall
2   https://en.wikipedia.org/wiki/Occam%27s_razor
3   https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/
4   https://www.theverge.com/21346343/gpt-3-explainer-openai-examples-errors-agi-potential
5   https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/

Disclaimer

About Imprima

Imprima is a leading Virtual Data Room provider, handling over $1 trillion-worth of transactions in over 160 countries.

Our clients include high-profile corporations, financial institutions, and advisors.

Whether it's a complex M&A transaction, managing portfolios of assets or completing a time-sensitive restructure, our innovative products enable secure and fast deal execution.

- **Imprima Virtual Data Room (VDR)** is a highly secure, fast and intuitive virtual data room that preserves the confidentiality of mission-critical documents and communications during any transaction.

- **Imprima Asset Lifecycle Management (ALM)** empowers you to organise, manage and track your assets in an efficient and cost-effective way, ensuring they are sale-ready at any given time.

- **Imprima AI** uses the power of Artificial Intelligence to automate Legal, M&A and Real Estate due diligence. It delivers 90% faster and more accurate contract review and virtual data room structuring.

For more information, contact:

Lidia Nakonechnaya
Marketing Manager
Imprima
T +44 7734 962 655 | E Lidia.Nakonechnaya@imprima.com