Knowledge Management and AI: a Match Made in Heaven

Published on 2023-06-26 byJack Shepherd

UNHELPFUL

"Give me an example of an insolvency event of default clause" → → "Here you go – I don't care if you are acting for lender or borrower"
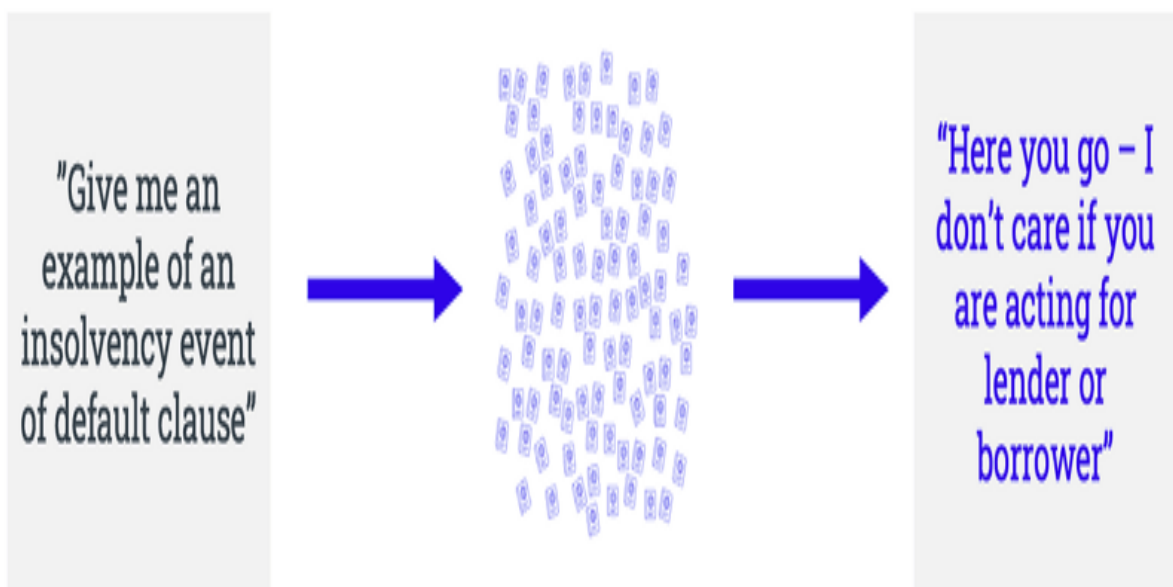
The ability to automatically classify content and then find and share just-in-time knowledge through AI is a dream many knowledge managers have held close for years. With the evolution of large language models, some have posited that AI is now advanced enough to be able to take over core knowledge processes without the need for human intervention or content tagging. In fact, though this advanced AI stands to vastly improve the efficacy of knowledge work, it is those firms that have focused on knowledge management and put effort into building structured databases of content that stand to gain the most from the technology.

Spotlight on Knowledge Management

During the pandemic, knowledge sharing became more difficult for law firms and legal teams. Traditional ways of sharing knowledge, such as listening in on meetings and calls and in-person

training events, no longer worked. As a result, in the past few years, there has been an increased focus on knowledge management strategies in the legal industry.

Such strategies can be far-reaching, such as how to define working practices, cultural improvement, and harmonizing business data sources.

This article focuses on another strand: how a firm harnesses the experiences of individuals to build collective knowledge. Often, this might take the form of collecting (or "curating") useful work product from lawyers that might be repurposed in the future. It might also involve specialist knowledge teams producing content that incorporates the competitive intelligence of their organization.

Knowledge powered by AI

As part of this discussion, most teams have also had one eye (or, in some cases, both) on AI. For many years now, it has been the vision of many teams that the capture and sharing of knowledge management is supported (or driven) by AI.

Indeed, many teams that do not have a culture of knowledge sharing (e.g. "our lawyers will never click a button to share information") or lack the resources to successfully implement a knowledge management system (e.g. "we do not have the time to tag thousands of documents) have put all knowledge management activities on pause until AI can do all of this for them. The emergence of large language models (LLMs) means that many of these teams may not have to wait for much longer.

Metadata

There is a common thread that holds all types of knowledge management activities together. That thread is the quality and level of categorisation of the content, often represented by the level of metadata and context surrounding knowledge content.

Whether you are looking to make the knowledge content you already have more findable, or to develop and drive your knowledge strategy, the basic principles that (1) your own content matters, and (2) cleaner data creates better results, remain as true as ever in the context of advanced LLMs.

Metadata in the real world

Law firms and legal teams are facing a data explosion. For organizations that have deployed purpose-built document management systems (DMS), much of it will be held in the same place. Many believe that there are huge insights to be had from that data.

Most document management systems will have automatically apply a certain amount of metadata to content, e.g. "filename", "date modified", "checked out by", "client ID" etc. This kind of metadata is essential for the purpose of document management.

File metadata

Filename	Lemon - Underwriting Agreement.docx
Date modified	2022-07-16T19:20:30.45+01:00
Author	RCROSS
File type	.DOCX

Metadata that might exist on a file

DMS metadata

File type	
Checked out by	MVANGERWEN
Document class	CONTRACT | Contract
Client ID	994387
Comments	[No comments]

Metadata that might exist in a DMS

I have seen first-hand, and heard of many more, organizations that have taken a knowledge management system and plugged it into their document management system without doing anything to the underlying content. This kind of deployment works well if users have a high degree of capability when it comes to using complex search syntax. It might also work well for more retrieval-based searches, where people know in advance what keywords to type into a search bar, or if they know the specific content they are looking for.

However, when it comes to true knowledge management—where users need to be more exploratory in their research—this approach can only deliver so much. The main reason for that tends not to be the knowledge management tool of choice, but rather the quality of the data being fed into it.
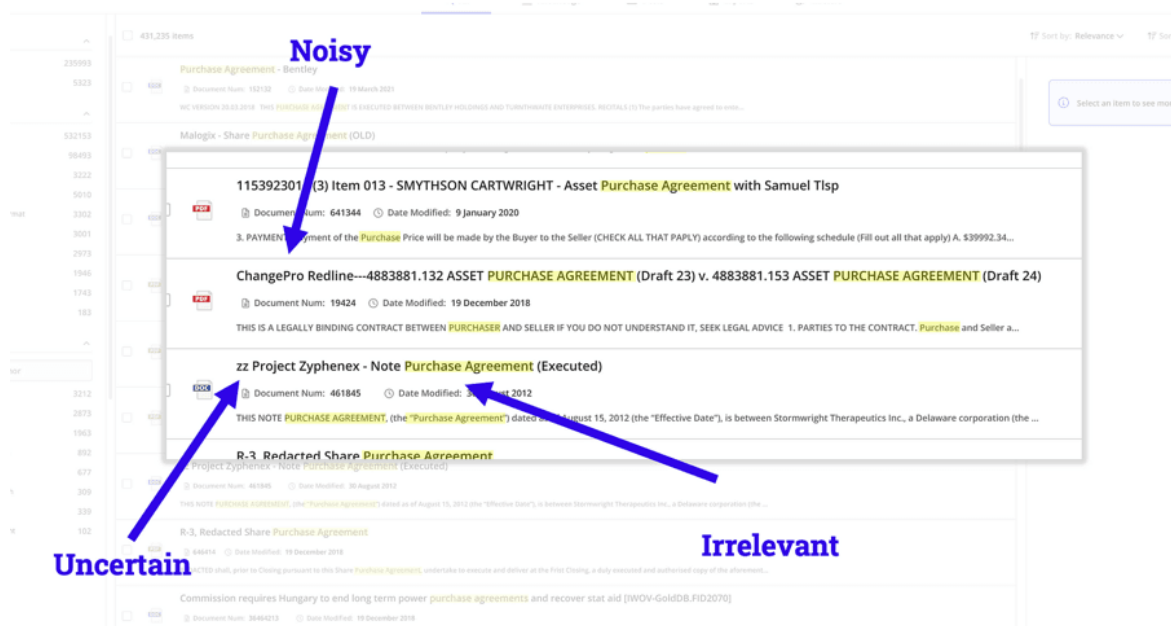
KM systems with poor content tend to lead to poor results

This should not be a surprise. If you are feeding a set of documents with metadata optimized for document management rather than knowledge management, users will not be able to refine searches or browse to content. For example, you cannot refine content according to which practice area it belongs to, which legal topic it deals with, which specific types of contracts are required or the jurisdiction(s) applicable to the content.
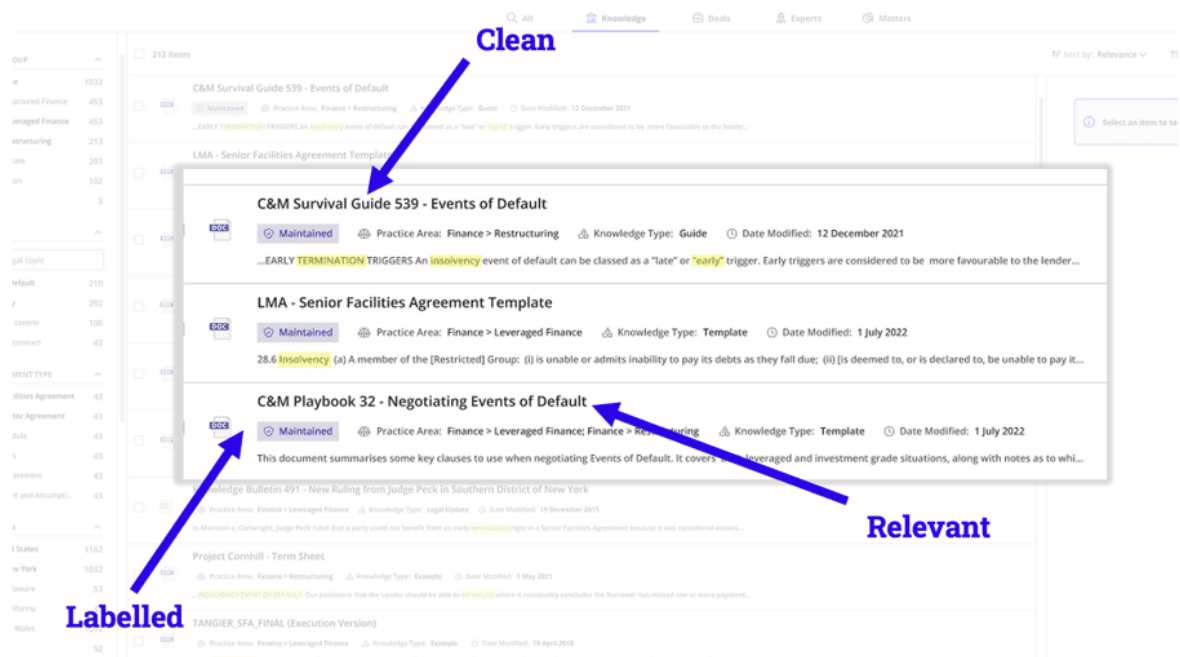

Contextual metadata can improve content quality for KM purposes

By putting in place robust processes for knowledge to be collected, shared and labelled, organizations can apply contextual metadata to content, making it suitable for knowledge management purposes.

Results tend to improve with contextual metadata

Using AI to classify content

Many would agree with me at this point, but note that it should not be the role of a human to label or spot knowledge content. This is a job optimized for AI.

The critical questions then become, "what can AI tell about a document", and "what can only a human tell about a document"? For example, AI might be able to tell you with a good level of accuracy that a document is an Underwriting Agreement. It can probably tell you what the applicable jurisdiction and governing law is. All of these things are hugely helpful for knowledge management purposes, because they help people refine their search accordingly.

The difficulty is that most DMSs are "noisy". They are full of half-baked drafts, aborted research memos and documents drafted in a rush. While AI can tell you basic facts about a document, as things stand right now, it cannot tell you whether a given example of a contract is reliable to use or not. Ending up with 1000s of potentially terrible drafts of Underwriting Agreements is less helpful for knowledge management than you might think.

Furthermore, when it comes to "example" documents, we too often think of the example itself being the extent knowledge, when the real value lies in the war stories, negotiations and tactics that led to its completion. Unless Alexa really is listening in while you work, it will be difficult for AI to be able to pick up this kind of context.

While AI can be undoubtedly be used to improve the cleanliness of content, we may be some way away from being able to exclude humans from the process. At the very least, as things stand now, we need a human to instigate or trigger the knowledge sharing process.
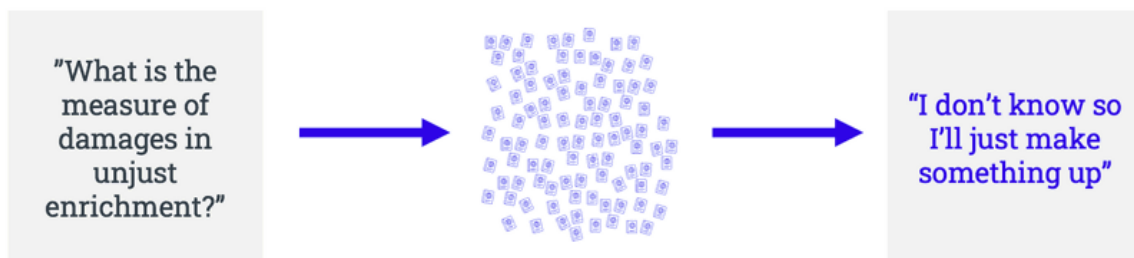
Using AI to produce knowledge

I have seen many bold claims when it comes to the capabilities of LLMs, for example "GPT can produce a first draft of a contract", "GPT can write a research memo for you". Often, I am left wondering a little whether people are really thinking about what happens when this output ends up in the hands of lawyers.

The advanced crop of LLMs have not been around for a long time, but are already infamous for making up erroneous laws, missing or omitting context and giving output that lacks any indication of supporting materials or source. It is ironic that these values—namely, accuracy of legal advice, tailored to a client, supported by authority—are values that most lawyers hold dear.
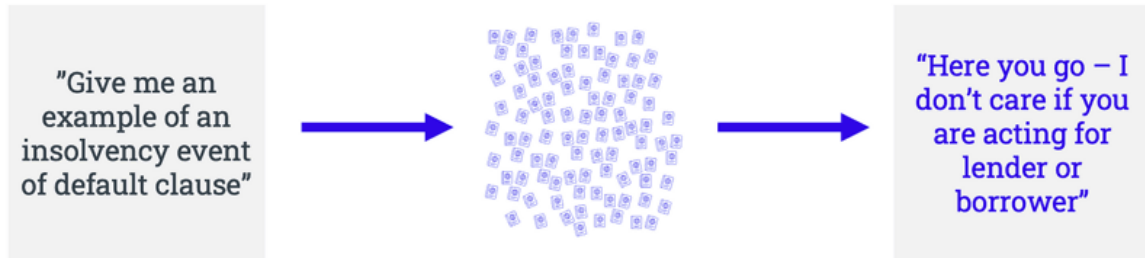

MISTAKEN

"What is the measure of damages in unjust enrichment?" → "I don't know so I'll just make something up"

The output of LLMs can sometimes be...wrong

UNHELPFUL

"Give me an example of an insolvency event of default clause" → "Here you go – I don't care if you are acting for lender or borrower"

The training data of LLMs does not always get you to the right place

UNPROVENANCED

"Show me some favourable terms I can negotiate for an investment side letter" → "Sure – I'll leave it to you to explain these when the client asks about them"
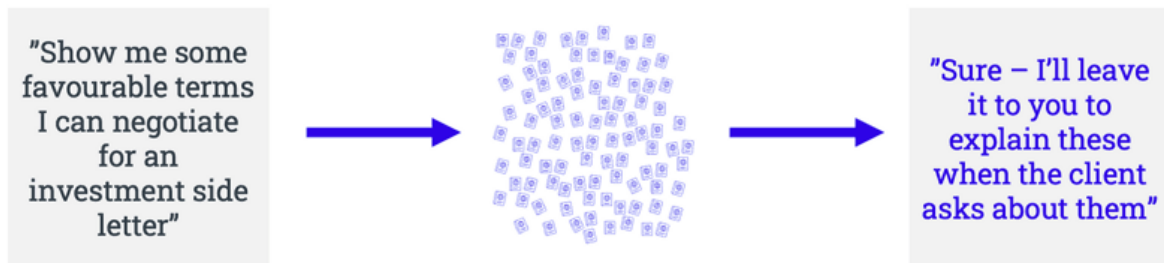
Follow-up questions or citations are challenging with today's foundational LLMs

While I do not agree that this can be fixed by simply improving the "prompt" you give an LLM, I still think LLMs have an incredible promise when it comes to knowledge in law. At a basic level, they can be used to help people find key knowledge content, e.g. by drawing up a list of keywords somebody might search for if they are looking for a given document. This, in combination with AI classification models, will undoubtedly be extremely powerful—and in these use cases, it does not matter hugely whether the output is completely correct.

LLM training data	General docs	Contextual docs
Almost incomprehensible scale, but contents are unknown	Large dataset, but lacks context and of varying quality	Smaller dataset, but has much more context and higher quality

Combining LLM training data with contextual data might be the best way forward

But the potential of LLMs does not stop there. One exciting area of research is what happens when you supplement the training data of an LLM such as GPT-4 with legal specific information. Casetext's CoCounsel was one of the first to do this with its structured database of legal content. A number of law firms are actively working on building data models that supplement foundational LLMs with the legal content from their own systems. It seems a reliable bet that firms with well-structured knowledge bases (rather than noisy, unclassified DMSs) will be at a huge advantage here, because they have handpicked and contextualized the exact content they want to use to drive the output.

Bringing the two strands together

If you are waiting for AI to sort all of your knowledge management problems out for you, it might be time to rethink things. Yes, you might get some limited use cases from "out of the box" LLMs that can produce half-baked memos—but you will still be left with lingering doubts around accuracy and provenance.

It will be interesting to watch as existing leaders in knowledge research platforms - such as Thomson Reuters, which has just acquired Casetext, LexisNexis, and Bloomberg - use their high levels of curated content to optimize the results of generative AI models. Firms that wish to use their own insights to gain a competitive advantage might be able to further optimise the results with their own "crown jewels".

To do this, they need to start building high quality, well-labelled and contextual repositories of knowledge. Those that embark on this journey will get a whole host of benefits on the way— namely a very accessible knowledge base system that can be used for research. This combination of "traditional" and "future" knowledge management strategies is surely a smart move for any organization thinking about a way to manage knowledge that is "AI-ready".