

# ARTIFICIAL LAWYER

AI AND LEGAL AUTOMATION NEWS + VIEWS

ABOUT CONTACT HOME

NEWS TICKER > [ 16th August 2017 ] ILTACon17 Guest Blogs: Day Two – All

SEARCH ...

## Guest Post: Legal AI vs eDiscovery

🕒 9th May 2017 👤 artificiallawyer 📧 AI + eDiscovery 💬 0



This is a Guest Post by the well-known, UK-based, eDiscovery expert and consultant, **Jonathan Maas**, which looks into the increasingly integrated area where AI and legal discovery technologies meet and overlap.

**Artificial Lawyer** has to say that this strange parallel universe that legal data analysis has lived in, with one plane of existence called 'eDiscovery' and the other plane of existence called 'legal AI document review', has always seemed amiss.

### EMAIL ALERTS

SUBSCRIBE

### RECENT POSTS

484% Rise In Legal Tech Patents in 5 Years, Says Thomson Reuters

ILTACon17 Guest Blogs: Day Two – All Dressed Up for the AI Party

Global Legal Blockchain Consortium Forms, Integra Ledger A Key Member

### FOLLOW ME ON TWITTER



Though, as with all market segmentations there is a very good reason for it, which in part is an historical one, i.e. eDiscovery was a truly commercial enterprise long before 'Legal AI' hit the scene.

Jonathan has taken up the noble challenge of navigating between these two worlds and drawing up some insights from his voyage to share with us. Thank you, Jonathan.

### The World of eDiscovery

Every year sees a new buzzword relating to the use of technology within the legal profession. Traditionally that use has been related to the document-intensive process of disclosure (or discovery).

In Common Law jurisdictions, such as America and England & Wales, parties engaged in litigation are legally obliged to find and exchange all relevant documents prior to trial so that the outcome of the dispute is decided based on all the available facts. This process is known as discovery (America) or disclosure (England & Wales), and an increasingly painful set of sanctions/penalties are 'awarded' if a party fails to comply adequately with these obligations.

There's more to it than that, of course. Each party needs to find, preserve, collect, review for privilege and relevance and finally disclose those relevant, unprivileged documents, and then deal with the other party's incoming disclosure. Nowadays these documents are almost entirely in electronic format, but paper still rears its head.

On large matters this can easily be the most costly part of a dispute outside of the trial itself so technology has become the solution to the problem caused by the technology. The upside in this area for the lawyers is that they have usually been able to charge their clients for the use of this technology. Not so in other areas of the law.

Litigation lawyers have always earned an honest buck reviewing documents to ascertain whether they are relevant (penalties apply if your disclosure contains too many irrelevant documents) and to remove privileged documents (private communications between lawyer and client). In the paper world this meant putting a team of junior lawyers in

### Tweets by @ArtificialLawya



**Richard Tromans**  
@ArtificialLawya

Thanks very much @thomsonreuters - always great to have some solid data on legal tech. Much appreciated.

>>>

[artificiallawyer.com/2017/08/16/484](http://artificiallawyer.com/2017/08/16/484)

...

#### 484 % Rise In Legal Tech Pa...

There has been a 484% increase in the number of [artificiallawyer.com](http://artificiallawyer.com)

3h

[Embed](#)

[View on Twitter](#)

a room full of lever arch files of copy documents and letting them read and file documents according to various appropriate and pre-ordained criteria.

In today's world this means putting a team of junior lawyers in a room full of computer screens, using a document review system to tag documents to those same (but often more) criteria. We all know that the ease of creating and sharing documents nowadays means that there are considerably more documents in play than in the paper world. Lawyers therefore need to use technology to whittle that number down.

## **Lawyer Versus Technology**

This is where the lawyer-versus-technology tension kicks in: lawyers traditionally charge by the hour yet are being forced by their clients or, to a lesser extent, by prevailing business practices to introduce efficiencies. This is a tough call as this means fewer hours billed. For some litigators, this is still a step too far. However, a growing number now embrace the chance to use technological advances to allow them to work better for less cost and thus improve the whole litigation experience for their clients. Analytics is now the name of the game, and that's where the smart money is going.

By way of background, there are a number of common processes applied to the electronic documents to reduce their volume and therefore the cost of review and disclosure:-

1. Remove all computer-generated content ("de-NIST") by applying a filter of known computer programme file types using a public list compiled and shared by America's National Institute of Standards and Technology.
2. Remove all exact duplicates according to their MD5# or SHA-1 value (alpha-numeric fingerprints unique to every document).
3. Optionally, also remove near-duplicates based solely on the text within the documents. The level of similarity can be set by the lawyers on that matter but, as for all disclosure decisions made, this must be defensible in court if required. An example is weeding out an email that has exactly the same text as another, but bears a different time-stamp due to different system times on the sending and receiving servers.
4. Email threading is then applied. This is the removal of all individual emails that make up the chain of correspondence so that one "master" is retained that accurately reflects the entire

discussion, plus any attachments. Hijacked chains that occur when someone uses an earlier email to begin a different discussion are treated separately.

5. Usually the lawyers will have developed a set of “keywords”, perhaps agreed with the opposing lawyers, that when used in a search would quickly isolate potentially relevant documents for legal review. In a product dispute, for instance, these keywords would include variations of the product’s name, the names of project team, and so on.
6. An extension to this process would be to use keywords to identify potentially privileged documents (such as known lawyers’ names, names of law firms, legal phrases, etc.)
7. A further extension could be searching for keywords designed to exclude obviously irrelevant material (such as known senders of spam email).
8. The final common stage is restricting the document population based either on the dates within which the events that led to the dispute are agreed to have taken place and/or the key players involved in those events.

This is when it starts getting interesting: our new Wild West is the world of developing methods of increasing technical complexity to reduce the volume or to focus the legal review team on the most potentially relevant documents first: the use of Technology (or Computer) Assisted Review. This is what US Magistrate Judge Facciola declared in a judgment in 2008 as somewhere “angels fear to tread”<sup>[1]</sup>.

Predictive coding is one type of TAR: a lawyer with a deep understanding of the issues in dispute reviews a “seed set” of documents and hands them, with their decisions as to relevance, back to the system which then tries to emulate that reasoning on a different batch of documents. The lawyer then marks those decisions and returns the corrected result back to the system for it to have another go. This continues until the overturn rate by the lawyer of the system’s decisions reaches an acceptable rate.

As ever, the lawyer must be able to defend this process in court if required. The end result is that the system stands a good chance of directing the legal review team to the most relevant documents first. It does not conduct the review itself (a common misconception within the profession). This has been the subject of a number of recent US, English, Irish and Australian judgments where judges have shown approval for the use of predictive coding in appropriate cases.

This is a one-off exercise in that the system applies what it has been taught. If that which is relevant changes, or additional documents from a different part of the client's business are added, then the system needs to be re-taught what is relevant by a human. The technology has moved on from then and we are now watching the rise of the algorithm through machine learning. This is where the system, once taught, can continue the learning process on its own.

In 2016 **Forbes** magazine described machine learning thus:

*'Very basically, a machine learning algorithm is given a 'teaching set' of data, then asked to use that data to answer a question. For example, you might provide a computer a teaching set of photographs, some of which say, 'this is a cat' and some of which say, 'this is not a cat'. Then you could show the computer a series of new photos and it would begin to identify which photos were of cats.'*

*'Machine learning then continues to add to its teaching set. Every photo that it identifies – correctly or incorrectly – gets added to the teaching set, and the program effectively gets 'smarter' and better at completing its task over time.'* [2]

As the judge said, we are where angels fear to tread: this process is currently regarded as the most reliable way of a party discharging its legal obligations cost-effectively when faced with today's often eye-watering volumes of documents. Many simply accept that this is currently the best cost-effective way of legitimately coping with the volumes of data involved in disputes nowadays. I'm sure even better approaches will evolve as the technology itself evolves and humans accept that computers can be trained to make good decisions. Lawyers of a scientific bent, especially in the US, are picking over the mathematics and science behind the technology to be assured that the technical processes are legally acceptable.

## **AI and Litigation**

Within litigation artificial intelligence is finding a role, too. RAVN[3] recently launched their ACE powered robot for LPP (legal professional privilege). This seeks to automate the review for privileged material, removing another income stream from lawyers (but not entirely, as the computer's decisions rightly need to be quality controlled by humans. For the time being). I believe most lawyers would like to move away from the world of repetitive drudgery inherent in large-scale document exercises and actually use their training and experience where it really matters.

Thus, artificial intelligence is gaining footholds in other legal areas such as contract management, BREXIT and GDPR compliance. Other companies in play include (but are not limited to) **Kira** [4], **Neota Logic** [5] and **Brainspace** [6]. Increasingly natural language algorithms are being used in tandem with the other methods mentioned here to allow lawyers to get quickly to the motherload of relevant documents, and most document review systems are building this level of analytics into their software. A good example of the integration of traditional and modern methods in a document review tool is Servient [7].

By way of example, I was recently engaged on a matter that saw us collect some 50 million documents. We reduced that to 2.9 million documents using the common processes I mentioned earlier (de-duplication, keywords, date ranges, etc.). Under the agreed charging mechanism the cost to review those 2.9 million documents would have been around £4.5 million. Although we had already reduced the document set by some 95% we realised that the cost of review was prohibitive (even though the claim and counterclaim combined amounted to nearly £100 million). We therefore suggested it would be worthwhile to throw some analytics at the problem to see how much further we could defensibly reduce the amount of documents for review.

We used a combination of in-house scripts and Brainspace, and added some natural language algorithms (for instance looking for documents that contained the agreed keywords but only in a negative context). We managed to reduce the 2.9 million to 700,000 documents, which went through to review at a cost of £1.1 million. We also analysed a statistical sample of that result, which showed the error margin was within the acceptable threshold agreed with the opposing lawyers. Of those documents, we disclosed about 300,000. We did not use predictive coding as we were confident that the documents did not lend themselves to that approach. Analytics saved the client £3.4 million of review costs.

Many firms are now embracing technology as a differentiator, either by using it to add value to or reduce the cost of their services or by sponsoring start-ups in some way to become early adopters of new technology. Stand-out examples include **McCann Fitzgerald's Data Investigations Group** [8], **Simmons & Simmons' eDiscovery Solutions team** [9] and **Allen & Overy's Fuse** [10]. This has been an extraordinary turn-around of a traditionally slow-moving and change-averse profession. However, this is also polarising the profession. I was at a

conference at the Law Society recently addressing “The Business of Change”.

**Isabel Parker, Freshfields’ Director of Legal Services Innovation<sup>[11]</sup>**, gave a jaw-dropping presentation of their impressive use of technology across their practice, highlighting the benefits to their “customers” of embracing new ways of doing things and yet one of the questions put to Isabel was “Should we be investing in technology during these straitened times?”

Artificial intelligence is the perfect tool to allow lawyers, and other providers of professional services to fill the gaps between the boring parts we are rightly developing computers to do for us so we can focus on the places where our intelligence can genuinely make a difference. These are opportunities, not threats. Too many times lawyers, frightened by the pace of change around them, reach out for technology, any technology, and pointlessly install it thinking they are still in the game. The point is to identify the problem and then find the right technology to solve it. Artificial intelligence is the perfect solution waiting for the right problem.

---

*References used in this article:*

**[1]** <http://www.insidecounsel.com/2008/09/05/where-angels-fear-to-tread>

**[2]** <https://www.forbes.com/sites/bernardmarr/2016/09/30/what-are-the-top-10-use-cases-for-machine-learning-and-ai/#196d53ca94c9>

**[3]** <https://www.ravn.co.uk/>

**[4]** <https://kirasystems.com/>

**[5]** <https://www.neotalogic.com/>

**[6]** <https://www.brainspace.com/>

**[7]** <http://servient.com/>

**[8]** <http://www.mccannfitzgerald.com/expertise/dispute-resolution—litigation/data-investigations-group.aspx>

[9] <http://www.simmons-simmons.com/en/online-services/our-ediscovery-solution>

[10]

<http://www.allenoverly.com/advanceddelivery/Fuse/Pages/default.aspx>

[11] <http://www.freshfields.com/en-gb/about-us/innovation/>

Share this:

Tweet
Share 4
G+
submit

Share
167
Email
Print

Related

Legal AI Enters Audio-Visual Realm With Veritone  
 20th April 2017  
 In "Audio Visual AI"

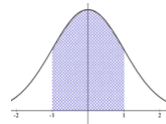
Neota Nabs High Profile COO; Wins Global Law Firm Freshfields as Client  
 30th January 2017  
 In "News"

The Artificial Lawyer Year in Review: The New Era of Legal AI Begins  
 19th December 2016  
 In "News"



« **PREVIOUS**  
 Guest Post: The 'Contract Stack' Revolution Begins

**NEXT** »  
 Welcome to Legal AI's 'New Normal'




---

**ARTIFICIAL LAWYER**

---