Legal Tech - Bringing Law into the "Twentieth" Century

by Tomasz Grzegory, Janos Puskas

1. Introduction

Legal professionals are renowned for their affection for the past. One could say that dogmatism is in the DNA of the legal profession. We still heavily rely on paper documents, attorney's letters delivered by registered mail, telephone calls and emails - all of which are 20th century inventions.

The global pandemic quickly levelled us up to video chats, online collaboration and cloud computing. Still, these are inventions from the past century, only, until now, largely disregarded by the legal industry.

However, the world outside is changing at an exponential rate. Innovation and disruption are the key leading drivers of human development. Technology is changing our everyday lives.

This applies to the legal profession as well. There are some exciting ideas to leverage technology and use it for the benefit of lawyers and their clients. Some of them are becoming working products, but innovations are yet to have a disruptive impact on everyday work.[1] Legal tech has a high potential when it comes to supporting law firms or legal departments in their processes and helping improve their effectiveness. However, the legal profession is still behind other businesses in the adoption of innovative technologies and the use of legal tech is not widespread in practice.[2] This is attributable only partially to the fact that statutory regulations do not keep up with the development of technology.

---

1 Helena Hallgarn, 'Legal Tech: Moving from ideas to execution' (2020) <https://www.legalbusinessworld.com/post/legal-tech-moving-from-ideas-to-execution> accessed 15 February 2021.
2 The Law Society of England and Wales, 'Lawtech Adoption Research Report' (2019) <https://www.lawsociety.org.uk/topics/research/lawtech-adoption-report> accessed 15 February 2021.

2. *"If Only I Had the Right Questions!" Where Are We and Where Do We Go in the Legal Tech?*

Lawyers face the challenges of transforming markets, data overload and ever-increasing complexity of information while meeting client's expectations and staying productive. In order to fulfil work objectives, it is already helpful to apply innovative changes and automate certain tasks. Introducing new technologies in daily practice becomes indispensable.

Most of the challenges stem from the ever-increasing and fast-changing regulatory environment. Over time, regulatory frameworks become more complex and detailed, with more and more new regulated areas. A perfect example of this are the recent (and future) regulations in the European Union affecting digital services, like GDPR and DSA. The expansion of consumer regulatory framework leads also to customer-facing companies receiving an increased number of complaints, which are more and more complex. According to Google's transparency report[3], Google has received a total of 1 011 696 'right to be forgotten' removal requests (concerning 3 971 081 URLs) by individuals since January 2015.

In addition to European Union regulations, national content monitoring laws like the German NetzDG generate further legal obligations on tech organizations. Digital service providers regularly receive requests to remove content from their products by courts and government agencies around the world. According to Facebook's transparency reports, Facebook has performed 10 content removals by the request of the Hungarian government[4] and 12 content removals by the request of the Polish government[5] since 2017.

It is important to point out that the regulatory burden has increased across industries in the past decades. If we consider the EU level legislation together with the national level legislation regarding a specific case, it may be necessary to review thousands of pages of legal material in order to answer a legal question.

The increasing number of regulations generates both legal and operational tasks. These tasks can be performed either by humans or by applying

---

3 According to data published on 25 February 2021 <https://transparencyreport.google.com/eu-privacy/overview?hl=en> accessed 26 February 2021.

4 According to data published on 25 February 2021 <https://transparency.facebook.com/content-restrictions/country/HU> accessed 26 February 2021.

5 According to data published on 25 February 2021 <https://transparency.facebook.com/content-restrictions/country/PL> accessed 26 February 2021.

legal tech solutions. As we explain below, deploying legal tech solutions will be more efficient and cost-effective in the long run.

Certain legal tasks or procedures (e.g., due diligence, compliance investigations) require lawyers to review large amounts of data. As companies accumulate more and more data and work with digital tools, the number of documents to be reviewed is also constantly increasing.

3. What's Hot in the Legal Tech?

The buzzwords of the early 21st century are among others: #bigdata, #blockchain, #machinelearning, #AI, #cybersecurity, #encryption, #collaboration, #cloudcomputing.

In our present review we group legal tech applications into two categories: the first contains generic software that is used by legal professionals to provide back-office support to teams of law firms' employees (such as automated billing software, timesheet software, project management tools, communication tools, online security tools), the second contains software applications that are designed and tailor made with solely the legal industry in mind (e.g. tools for legal research, e-discovery, contract management, predictive analytics).

Both groups however have one common denominator. They are using already existing ideas and solutions that were repurposed to serve legal functions.

Within applied legal tech, some promising areas of development emerged:

- Text analysis: The automated process of sorting unstructured text data, making it easier to retrieve relevant information. A massive amount of paperwork can be generated when conducting various legally important processes (e.g., due diligence). Well-designed algorithms can analyze thousands of documents better and faster than the human workforce. Such tools search through an organization's documents to automatically pick out legal concepts - which helps reduce risk and make the due diligence process a whole lot faster.[6]

---

6 Alejandro Esteve de Miguel Anglada, 'AI and Machine Learning in Legal: Tools Every Lawyer Needs to Know (In a Nutshell)' (2020) <https://blog.biglelegal.com/en/ai-and-machine-learning-legal-tools-every-lawyer-needs-to-know> accessed 22 February 2021.

- Text analysis software is widely used in knowledge-driven organizations to discover new information or help answer specific research questions from large collections of documents.[7] First, text analysis gathers unstructured data from multiple text-based data sources. Then, it detects and removes anomalies by conducting pre-processing and cleansing operations and extracts all relevant information. After the extraction, the software converts the information into a structured form that can be further analyzed or presented directly.[8]

- Text analysis uses various methods to process texts, one of the most important is '*natural language processing*'.

- E.g., an advanced text analytics software[9] combines text analytics with machine learning technology and natural language processing in order to automatically identify and structure all possibly relevant information from all types of text data. In addition, the software allows its users to customize the system's capabilities and tailor the system to the user's own operational specifications and requirements.

- <u>Law firm practice and compliance management</u>: There are Task/Matter Management software, through which certain business tasks can be subdivided and distributed among different corporate departments/lawyers, this makes whole projects more transparent. The participants of the project can collaborate effectively and comment on the current work progress. This category also includes applications for document storage and automated billing. The most advanced software is capable of performing all the relevant tasks on their own, automating all business and legal processes.[10]

- Most of these software enables to track performance on ongoing legal matters, reporting and invoicing on client projects, remotely manage legal cases, automatically log working hours and manage all email and communication and other documents.

---

7 Kevin D. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (Cambridge University Press 2017) 88.
8 Linguamatics, 'What is Text Mining, Text Analytics and Natural Language Processing?' <https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing> accessed 22 February 2021.
9 An example is Kira Systems <https://kirasystems.com/> accessed 20 February 2021.
10 Salvatore Caserta and Mikael Madsen, 'The Legal Profession in the Era of Digital Capitalism: Disruption or New Dawn?' (2019) 149 iCourts Working Paper Series 10.

- E.g., an advanced practice management software[11] offers customized solutions not only for law firms, but also specifically for in-house legal teams and business teams. The contract management solution with a contracts database can support the alignment between legal and business units and significantly increase the transparency of contracting processes. Legal information processed by the software (on scale) may serve as a relevant data asset supporting business decisions.

- <u>Communication tools</u>: COVID-19 accelerated the use of online collaboration tools, cloud computing technologies, online video conferencing platforms and much more. The pandemic has forced an acceleration of the digital transformation in all industries. This created an interesting loopback - businesses migrated online and the legal industry had to follow.

4. *'All that Glitters Is Not Gold'*

Artificial intelligence (AI) has taken our world of expectations by storm. Nowadays, AI is often portrayed as a disruptive, almost magical technology that is going to entirely reform the legal profession as well.[12] It may even seem that AI could offer a solution to any problem concerning businesses or the legal profession. 'AI' is by far the most frequently used (and abused) buzzwords of the recent years. Even though it generates a constant buzz, the term 'artificial intelligence' can be a bit misleading when it comes to legal tech applications. In general, we can conclude that using the term 'machine intelligence' instead of 'artificial intelligence' is more accurate in the context of legal tech (please see detailed explanation of this distinction below, under the technology background).

Due to the current hype around AI, it may seem like every legal tech company is an AI solutions provider. The term '*AI washing*' refers to the marketing practice based on inaccurate labeling of certain products or services to imply that they deliver AI, when in fact they might be nothing more than a simple automation, basic statistics or a new marketing spin

---

11 An example is InvestCEE <https://investcee.hu/> accessed 22 February 2021.
12 Andrew Ng, 'What Artificial Intelligence Can and Can't Do Right Now' (2016) Harvard Business Review <https://hbr.org/2016/11/what-artificial-intelligence-can -and-cant-do-right-now> accessed 22 February 2021.

for an existing application.[13] Automation is the result of a machine being told by a human precisely what to do, and then successfully doing it. This is a simple execution of a command that cannot qualify as an AI, since it is not trained to improve by itself over time. For example, digital signatures and excel formulas are automations that cannot be considered as AI solutions.

The phenomenon of "*washing*" isn't new – we had experienced it with '*green washing*' and '*cloud washing*'. Moreover, the tech industry has always been infatuated with buzzwords: before AI, '*big data*', '*cloud computing*', '*web 2.0*' were trending.

The reality is that some "*AI companies*" don't employ AI at all and only using the word to get funding or to generate attention and sales, leaving their customers dissatisfied by failing to live up to their own promises. AI washing risks turning artificial intelligence into nothing more than a broadly ignored marketing term by overinflating expectations[14] and creating widespread misconceptions related to AI, including its powers and what it can and cannot do.[15] At the end of the day, this commercial opportunism may lead people to lose trust in AI's functionality and struggle to accept and adopt AI into their lives.

5. So, What Does 'Artificial Intelligence' Really Mean in the Legal Tech?

In addition to the AI-washing phenomenon, AI's unclear definition[16] certainly aids in the confusion and deception regarding its applications. AI is a broad, '*umbrella*' term covering plenty of different technology solutions (from the state-of-the-art deep learning models to human coded rules), thus, it can be complicated to determine which tool should count as AI and which shouldn't.

---

13 Stephanie Overby, 'AI vs. automation: 6 ways to spot fake AI' (2016) The Enterprisers Project <https://enterprisersproject.com/article/2020/3/ai-vs-automation-6-ways-spot-fake-ai> accessed 23 February 2021.
14 Kaveh Waddell, 'The Dangers of "AI washing"' (2019) <https://www.axios.com/ai-washing-hidden-people-00ab65c0-ea2a-4034-bd82-b4747567cba7.html> accessed 23 February 2021.
15 Dr Bimal Roy Bhanu, 'AI-washing: is it Machine Learning … or Worse?' (2019) <https://financialit.net/blog/artifical-intelligence/ai-washing-it-machine-learning-or-worse> accessed 23 February 2021.
16 Michael Legg, Felicity Bell, 'Artificial Intelligence and the Legal Profession: Becoming the AI-Enhanced Lawyer' (2019) 38(2) University of Tasmania Law Review 38.

Researchers in the field of artificial intelligence largely agree that AI means computer science methods that allow machines to complete tasks normally viewed as requiring human intellect, in other words, the ability of computers to exhibit human-like cognitive abilities. So, AI mimics certain operations of the human mind. Also, there is a difference between Strong AI (or AGI) and Weak AI (or Narrow AI): Weak AI focuses on performing a specific task[17], while Strong AI can perform a variety of functions, and over time, it might (or might not) develop a human-like consciousness instead of simulating it, like Weak AI does.[18] AI currently used in legal technology have defined functions, so they are considered as Weak AI.

The terms "*machine learning*", "*machine intelligence*" and "*artificial intelligence*" are often used interchangeably.[19] Instead of these three terms being identical, machine learning and machine intelligence are subsets of AI.

In particular, machine intelligence (as a form of Weak AI) means that the machine is programmed with some (but not all) aspects of human intelligence, including problem solving, prioritization and learning. It enables a technology (a machine, an algorithm or a device) to interact with its environment intelligently, meaning that it can take actions to maximize its chance of successfully accomplishing its objectives.[20] It is the intersection of artificial intelligence and machine learning.

Machine intelligence is a more exact term for legal tech applications in general because it covers the full spectrum of functions (learning, problem solving, prioritization) that legal tech applications may offer at the moment, contrarily to 'AI' which is a broader and less accurate term. The functions of machine intelligence are still very far from what a Strong AI (i.e., a hypothetical machine that exhibits behavior at least as skillful and flexible as a human's) could perform.

And machine learning (ML) is one specific element of machine intelligence that centers around data, in which the computer uses algorithms

---

17 The Law Society of England and Wales, 'AI: Artificial intelligence and the Legal Profession' (2018) <https://www.lawsociety.org.uk/topics/research/ai-artificial-intelligence-and-the-legal-profession> accessed 24 February 2021.

18 Marcelo Corrales, Mark Fenwick, Nikolaus Forgó, *Robotics, AI and the Future of Law.* (Springer 2018) 59.

19 Kate Prengel, 'AI washing: Why Does it Matter?' (2020) <https://pivotalpredictions.com/ai-washing-why-does-it-matter/> accessed 24 February 2021.

20 Stephanie Enders, Cathy King, Spencer Murray, Anna Koop, 'Machine Intelligence, Artificial Intelligence & Machine Learning' (2016) <https://www.amii.ca/latest-from-amii/machine-intelligence-artificial-intelligence-machine-learning/> accessed 24 February 2021.

planted in the software to learn from data and improve automatically through experience. The aim of ML algorithms is to build a model based on sample data ("*training data*")[21] that can make decisions or predictions about new, previously unseen sample data without being explicitly programmed to do so. ML is used in a wide variety of applications, e.g., such as e-mail filtering, image or speech recognition, automatic language translation, recommender systems and autonomous vehicles.

There is a difference also between 'supervised', 'unsupervised' and 'reinforcement' learning. Supervised learning means that the algorithm is trained with labeled data, i.e., the algorithm is using some data that is already tagged with the correct answer (output variables are given besides input variables).[22] Unsupervised learning means that no labels are given to the learning algorithm, the algorithm needs to find a structure in its input on its own. Reinforcement learning means that the algorithm interacts with a dynamic environment in which it must perform a certain goal (like driving a vehicle) while constantly receiving feedback.[23]

Machine learning is based around '*pattern recognition*'. This means that the algorithms can learn to recognize certain patterns based on a set of previously observed data. Once the algorithm is trained to identify a pattern, it uses that as a basis of comparison for all new data.[24] Pattern recognition can also detect anomalies, for example, it can point out to lawyers the areas that tend to be overlooked in legal documents, as well as highlighting missing pages, incorrect or strange words and extra clauses - among other things - so lawyers know exactly where to look while double-checking contracts. Face recognition and visual search are among the top uses for pattern recognition, where the algorithm describes pictures so they can become searchable.[25] The ultimate aim of pattern recognition technology is to simplify and connect analytics' or search results.

---

21 Lauri Donahue, 'A Primer on Using Artificial Intelligence in the Legal Profession' (2018) Jolt Digest <https://jolt.law.harvard.edu/digest/a-primer-on-using-artificial-intelligence-in-the-legal-profession> accessed 24 February 2021.

22 Harry Surden, 'Machine Learning and Law' (2014) 89(1) Washington Law Review 103.

23 Exigent-Group, 'Why Machine Learning is Key to Legal Transformation' (2020) <https://www.exigent-group.com/blog/why-machine-learning-is-key-to-legal-transformation/> accessed 24 February 2021.

24 Zee Gimon, 'What Is Pattern Recognition in Machine Learning' (2019) <https://huspi.com/blog-open/pattern-recognition-in-machine-learning> accessed 25 February 2021.

25 Yulia Gavrilova, 'Pattern Recognition and Machine Learning' (2020) <https://serokell.io/blog/pattern-recognition> accessed 25 February 2021.

Since ML algorithms are based on the training data, machine learning becomes more effective as the size of the training dataset grows. *'Big data'* is a term that refers to the large volume of data, and it is also a collective name for methods of processing data that is so large, fast or complex that it would be difficult or impossible to process using traditional methods.[26] The ML process can be improved by feeding big data to the algorithm, because the more training data is shared with the algorithm, the more accurate the results will be.[27] If big data is involved, pattern recognition may identify hidden patterns and analytics even solely based on indirect data that we would never be able to find using a traditional rule-based search.

6. Do Machines Understand Us?

We've mentioned natural language processing (NLP) as the main method applied by text analysis that helps in structuring text-based data. NLP is a subfield of machine intelligence and linguistics concerned with the interaction between algorithms and human language. NLP helps the algorithm read and derive meaning from text (or another input such as speech – like Siri, Alexa and Google's voice search) by simulating the human ability to understand a certain language, such as English.[28] It works by learning human language, using context and prior queries and results to predict what attorneys may need in their searches. For example, NLP is applied by search engine autofill, spell checkers, speech recognition with voice assistants and chatbots. There are several areas of legal activity where NLP might play a significant role[29]:

- Legal research: NLP can translate plain English search terms into legal search. When searching case-law, the algorithm continues to learn ba-

---

26 Patanjali Kashyap, *Machine Learning for Decision Makers* (Apress 2017) 12.

27 Hans Weber, *Big Data and Artificial Intelligence: Complete Guide to Data Science, AI, Big Data and Machine Learning* (2020) 100.

28 Benjamin Alarie, Anthony Niblett, Albert Yoon, 'How Artificial Intelligence Will Affect the Practice of Law' (2018) 68 (supp.1) University of Toronto Law Journal 106.

29 Robert Dale, 'Law and Word Order: NLP in Legal Tech' (2018) <https://towardsdatascience.com/law-and-word-order-nlp-in-legal-tech-bd14257ebd06> accessed 24 February 2021.

sed on what cases are clicked on and viewed. Moreover, it can reveal where specific phrases appear in a lengthy document.[30]
- Case predictions: NLP can provide predictive models to help better understand how a given judge or court may rule. This can help attorneys better tailor their arguments to support or combat the prediction. Based on the predictions, lawyers could structure their argument around what the judge will find most persuasive.[31]
- Electronic discovery: Determining the relevance of documents to an information request.
- Contract review: Checking that a contract is complete and avoids risk.[32]
- Document automation: Creating routine legal documents.
- Legal writing: Drafting legal advice (e.g., using question-and-answer dialogs to provide tailored advice), e-mails, court notes and case notes.

7. *E-discovery - Where Work Became Technology Assisted*

Organizations usually face the discovery process in complex business disputes, litigations or regulatory investigations. Discovery is an initial process where relevant information and records, along with all other evidence related to a case, are sought.[33]

With the development of computer technologies, most information that is produced, distributed, and stored by businesses now exists in electronic form. E-discovery (the short form of the term "electronic discovery") encompasses what most often is referred to as electronically stored information, or ESI. The purpose of e-discovery is to locate, secure and preserve ESI

---

30 Alison Wilkinson, 'How Natural Language Processing Can Improve Legal Search Results' (2020) <https://kirasystems.com/learn/how-natural-language-processing-improving-can-improve-legal-search-results/> accessed 24 February 2021.

31 Daniel Martin Katz, 'Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry' (2013) 62 Emory Law Journal 936.

32 Mark Sears, 'AI Challenges and Why Legal Is A Great Place to Kick-Start Great NLP' (2019) <https://www.forbes.com/sites/marksears1/2019/05/14/ai-challenges-and-why-legal-is-a-great-place-to-kick-start-great-nlp/?sh=79ec487f4408> accessed 24 February 2021.

33 Burke Ward, Janice Sipior, Jamie Hopkins, Carolyn Purwin, Linda Volonino, 'Electronic Discovery: Rules for a Digital Age' (2012) 18(150) Boston University Journal of Science and Technology Law.

to be used as evidence as part of a legal case.[34] Data of all types can serve as evidence, e.g., database files, docs, e-mails, videos, images, instant messaging chats, voicemails, calendar files, spreadsheets, web sites, computer programs, animations, GPS and location data, IoT devices, mobile data, social media posts, audio files and even malware such as viruses, trojans and spyware. ESI is layered and often contains metadata such as timestamps, author and recipient information, and file properties. Metadata can be essential[35] when deciding on legal matters since it may contain the date and time a document was written as well.

Besides litigations and regulatory investigations, e-discovery software can be deployed also for early case assessment, internal investigations, and public records requests.[36]

E-discovery can be very complex and multi-layered depending on the current organization and the legal case i.e., how data is stored and how complex the data is.[37] E-discovery is more than one single task, it is rather a series of diligent steps and actions that together, as a consistent process, help build solid evidence for a legal case.[38]

E-discovery usually runs from the time a lawsuit is foreseeable to the time the digital evidence is presented in the litigation. The Electronic Discovery Reference Model (EDRM)[39] is a ubiquitous and widely referenced standard that represents a conceptual view of the stages involved in the e-discovery process[40]:

1. *Information governance*: It serves as a foundation for the e-discovery process and refers to the set of policies implemented to manage a company's data.
2. *Identification:* It is a phase when potentially responsive documents are identified. There are various methods to identify sources of potentially

---

34 Kimberly Williams, John M. Facciola, Peter McCann, Vincent M. Catanzaro, *The Legal Technology Guidebook* (Springer 2017) 44.

35 Philip J. Favro, 'A New Frontier in Electronic Discovery: Preserving and Obtaining Metadata' (2007) 13 (1) B.U.J. Sci. & Tech.

36 Logikcull, 'The Ultimate Guide to eDiscovery' <https://www.logikcull.com/guide> accessed 25 February 2021.

37 'The Basics of e-discovery' <https://www.exterro.com/basics-of-e-discovery> accessed 25 February 2021.

38 'What is e-discovery?' <https://www.barracuda.com/glossary/e-discovery> accessed 25 February 2021.

39 The EDRM model is available at: <https://edrm.net/resources/frameworks-and-sta ndards/edrm-model/> accessed 25 February 2021.

40 'The Basics of e-discovery. Chapter 1: Process' <https://www.exterro.com/basics-of -e-discovery/e-discovery-process> accessed 25 February 2021.

relevant ESI, e.g., interviewing key participants and reviewing case facts.

3. *Preservation:* After relevant ESI is identified, it needs to be protected from spoliation (any destruction or alteration of evidence). The most common way to preserve ESI is through a legal hold process. A duty to preserve information begins upon reasonable anticipation of litigation. According to *Zubulake v. UBS Warburg*[41] failure to issue a written legal hold notice whenever litigation is reasonably anticipated will be deemed grossly negligent.

4. *Collection*: It means the transfer of the relevant ESI from a company to their legal counsel. Some companies that deal with frequent litigation have software in place to quickly place legal holds when an event (such as legal notice) is triggered and begin the collection process immediately. The collection method must ensure that the contents and metadata are not altered as a result of the process.

5. *Processing*: It means the reduction the volume of ESI and converting it, if necessary, to forms more suitable for attorney review.

6. *Review*: It means the evaluation of ESI for relevance and attorney-client privilege (which is exempt from e-discovery). Considering exponentially growing data volumes, manual review became impractical. There are artificial intelligence (AI) tools that can distinguish between relevant, non-relevant and privileged documents, thus, making the review process more viable and efficient.

7. *Analysis*: Evaluating ESI for content and context, including key patterns, topics, people and discussion.

8. *Production*: ESI determined to be relevant must be produced for use of potential evidence in appropriate forms and using appropriate delivery mechanisms. ESI can be produced either as native files or in a petrified format (such as PDF).

9. *Presentation*: Displaying ESI before audiences (trials, hearings, etc.).

As mentioned above, legal teams can apply AI tools in order to save time and money on the review process and expedite e-discovery by prioritizing documents and reducing the number of documents required for manual review. These tools typically use predictive coding which leverages machine learning to surface likely relevant documents based on prior review decisions. Predictive coding goes by various names, including "*technology assisted review*" (TAR) and "*computer aided review*" (CAR). In order for

---

41 Zubulake v. UBS Warburg LLC - 220 F.R.D. 212 (S.D.N.Y. 2003)

the predictive algorithm to work properly, it needs to be 'trained' first. This means that reviewers pull a representative cross-section of documents ('seed set')[42] from the pool of ESI that need to be reviewed. Reviewers code each document in the seed set as relevant or not relevant and input those results into the predictive coding software. By these inputs, the software generates an internal algorithm for predicting the responsiveness of future ESI.[43] As the review goes on, machine learning enables the software to continually learn from reviewer decisions and make more accurate results in significantly less time than human reviewers.[44] Of course, if the seed set is coded incorrectly then the algorithm will multiply those errors across the entire document set, thus, it will lead to a 'garbage in, garbage out' situation.

US and UK courts have already adopted decisions on predictive coding. *Monique Da Silva Moore, et al. v. Publicis Groupe & MSL Group*[45], a decision from the United States District Court for the Southern District of New York, is the first decision that explicitly recognized the use of predictive coding technology as an appropriate method to satisfy a producing party's review obligation. The court held that the use of predictive coding techno-logy "*is an acceptable way to search for relevant [electronically stored information] in appropriate cases*" and also that "*it is not a case of machine replacing humans: it is the process used and the interaction of man and machine that the courts need to examine*". The court highlighted that the use of predictive coding is acceptable where it is applied in a manner that produces reliable and proportional results. In this case, the judge based his ultimate approval of its use on five factors: "*(1) the parties' agreement, (2) the vast amount of ESI to be reviewed (over three million documents), (3) the superiority of computer-assisted review to the available alternatives (i.e., linear manual review or keyword searches), (4) the need for cost effectiveness and proportionality […], and (5) the transparent process proposed by [the Defendants].*"

Regarding proportionality, the judge stated that:

"*In order to determine proportionality, it is necessary to have more informati-on than the parties (or the Court) now have, including how many relevant docu-ments will be produced and at what cost […] In the final sample of documents*

---

42 Gideon Christian, 'Predictive Coding: Adopting and Adapting Artificial Intelli-gence (AI) In Civil Litigation' (2019) 97 The Canadian Bar Review 490.

43 Dana A. Remus, 'The Uncertain Promise of Predictive Coding' (2014) 99 Iowa Law Review 1072.

44 Aaron Goodman, 'Predictive Coding: A Better Way to Deal with Electronically Stored Information' (2016) 43(1) Litigation 23.

45 Da Silva Moore v. Publicis Groupe - 287 F.R.D. 182 (S.D.N.Y. 2012)

deemed irrelevant, are any relevant documents found that are "hot," "smoking gun" documents (i.e., highly relevant)? Or are the only relevant documents more of the same thing? One hot document may require the software to be re-trained (or some other search method employed), while several documents that really do not add anything to the case might not matter. These types of questions are better decided "down the road," when real information is available to the parties and the Court."

Da Silva Moore also set out guidelines for the appropriate use of predictive coding. First, parties intending to use this technology must choose a reliable e-discovery vendor and software, as well as design an "*appropriate process*" that includes "*appropriate quality control testing*". Furthermore, in accordance with the decisions, the parties should consider the following steps[46]:

- Allow the requesting party to view ESI that were used to train and additional ESI that were used to stabilize the predictive analytics system (marked both as responsive and non-responsive).
- Do not limit the number of iterative reviews used to 'train' the system up front, but rather assess whether the system has stabilized before stopping the iterative reviews.
- Do not adhere to an arbitrary number of ESI that will be produced without reference to the statistical results.
- Bring the vendor experts to the court hearing to respond to the judge's questions.

The first English decision on the use of predictive coding software was made in the *Pyrrho Investments Ltd & ors v MWB Property Ltd & ors case*[47], where the High Court of Justice held that predictive coding is permitted and could amount to a reasonable search. In this case the judge stated that what fundamentally matters in the disclosure process is the scope and quality of the search, rather than the listing and production for inspection of the relevant documents, which were found during the process. The judge listed ten factors that favored the use of predictive coding in the case[48], including, for instance, the fact that experience with predictive

---

46 Special Counsel, 'What Predictive Coding Court Rulings Can Teach Us', blog.specialcounsel.com (3 November 2016); <https://blog.specialcounsel.com/ediscovery/what-predictive-coding-court-rulings-can-teach-us/> accessed 26 February 2021.

47 Pyrrho Investments Limited v MWB Property Limited [2016] EWHC 256 (Ch).

48 Oliver Browne and Hayley Pizzey 'Pyrrho Investments Ltd v MWB Property Ltd: A Landmark Decision on Predictive Coding in e-Discovery' (Latham.London, 15

coding in other jurisdictions has shown that the software can be useful in appropriate cases; the number of electronic documents in the case (amounting to over 3 million); the circumstance that the parties agreed on the use of the software and how to use it; the use of a computer to apply the approach of a senior lawyer towards the initial sample; and the conclusion that there is no evidence that predictive coding leads to less accurate disclosure compared to manual review and there is some evidence that it is more accurate than manual review.

Later, in *Triumph Controls UK Ltd & anr v Primus International Holding Co & ors*[49], the High Court of Justice held that a party should not act unilaterally to use CAR in the disclosure process as there is a risk that the "*unilateral decision will be carefully scrutinized by the court at a later date, and a different course may be ordered.*" Also, when a party has acted unilaterally, that party should provide the other party with details about how the algorithm was set up and how it was operated. That had not happened in this case, which the judge found to be "*unsatisfactory*". The court referred to *Pyrrho Investments Ltd.* where the court stated that the "*best practice would be for a single, senior lawyer who has mastered the issues in the case to consider the whole [teaching] sample*". With respect to that, *Triumph Controls UK Ltd* held that CAR technology must be 'taught' properly, and the best practice involves a senior lawyer oversight. The court also ordered a manual review of 25 % (equal to 55 000 disputed documents) considering this a proportionate amount under the circumstances.

Based on the relevant case law, there is no doubt that the use of predictive coding will increase in the future as a global trend. However, there are several barriers that have kept predictive coding from being widely adopted in the legal industry.

Firstly, because predictive coding relies on highly complex algorithms based on advance data science and statistical sampling, it is often seen as a 'black box'. Although humans train the algorithm, the internal processes are mostly opaque, causing the machine decision-making to be non-transparent.[50] The technology is understood by only a limited number of experts – and few of them are lawyers.

July 2016) <https://www.latham.london/2016/07/pyrrho-investments-ltd-v-mwb-p roperty-ltd-a-landmark-decision-on-predictive-coding-in-e-discovery/> accessed 26 February 2021.

49 Triumph Controls UK Ltd & anr v Primus International Holding Co & ors [2019] EWHC 565 (TCC).

50 Michael Weather, 'Predictive coding: the current landscape' disputeresolution blog.practicallaw.com/ (Thomson Reuters 21 July 2016) <http://disputeresolut

Secondly, predictive coding is yet slow and expensive to implement. It may be beneficial as a long-term investment; but it may also be difficult to finance the technology. It takes significant manual-review time to train machine learning how to categorize ESI effectively.[51] As mentioned earlier, if the setup is faulty, the algorithm will produce poor results. Therefore, the setup needs senior attorneys with plenty of experience who are knowledgeable about the facts of the case; often, a more senior attorney whose time is even more expensive.

Thirdly, predictive coding only understands logic and cannot make nuanced content distinctions, for instance, whether a document contains privileged information. Moreover, lengthy documents often include a multitude of topics, which may adversely affect their classification by the predictive algorithm. If a long document contains one short (although relevant) reference, the algorithm may leave this document out of consideration due to the relative noise of the other content in the source.[52]

Lastly, although most courts (at least in the US and in the UK) accept a predictive coding workflow, there is still some legal risk depending on the procedures employed and the results obtained. Especially if there is no agreement between the parties regarding the details of how predictive coding will be used in the case. It is sure that not all legal cases are appropriate for predictive coding.

8. *OK Google, Negotiate! When Do Lawyers Become Obsolete?*

Recent technology trends have led some lawyers to raise that their jobs may be under threat by legal tech innovations.[53] But can we really expect that lawyers will eventually be replaced by technology? We believe that technology complements the work of lawyers, rather than replacing it, thus, we see that legal tech offers more benefits than threats to legal

---

ionblog.practicallaw.com/predictive-coding-the-current-landscape/> accessed 26 February 2021.

51 Thomson Reuters Legal, 'How to make the e-discovery process more efficient with predictive coding?' <https://legal.thomsonreuters.com/en/insights/articles/how-predictive-coding-makes-e-discovery-more-efficient> accessed 26 February 2021.

52 PRISM Litigation Technology, 'Predictive Coding: The Good, the Bad, and the Ugly' (Prism Blog, 7 August 2019) <https://prismlit.com/predictive-coding/> accessed 26 February 2021;

53 Richard Susskind, Daniel Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts* (Oxford University Press 2015) 71.

jobs. It is very likely that legal tech tools will drastically change the way lawyers provide services for their clients. Technology can free lawyers from administrative and routine tasks, which will result in a decrease in their overall workload and more time for higher-value work. So, using technology will be a competitive advantage by increasing productivity and reducing human work hours. Therefore, even if technology won't replace lawyers, lawyers who use technology could replace those lawyers who don't.[54]

At the same time, human lawyers will always be needed for more critical, complex tasks with higher-level cognitive demands, like advising clients, formulating arguments, negotiating and representing in court. It could be argued that the consultative nature of legal work where human to human relationships matter, makes it immune to full automation.

For that reason, technology is not yet an imminent threat to legal jobs. Right now, legal tech tools could only decrease the work hours spent on more routine-oriented tasks, like contract review or document management; allowing lawyers to focus on tasks that truly add value.

However, most legal tech innovations are still in their early days, so we should not underestimate the power of technology. The legal profession is sure to change a lot in the near future.

Every business will be disrupted someday, and digital literacy will be a necessity for any of them to survive. We live in rapidly changing times, and now, we are also facing a period of fundamental and irreversible transformation in the world of law and technology.[55] In order to successfully cope with new challenges in the legal industry, lawyers and legal teams need to be open-minded to innovations, dismiss outdated practices and quickly adapt themselves to the fast-changing business environment by exploiting new tools and products. This means, in particular, that lawyers need to develop new competencies such as software skills, as well as understanding other disciplines of business including technology, analytics and data science. Lawyers will have to cooperate more and more with professionals from different fields to solve legal problems.

The need for disruption may not only stem from the changes inside the legal services industry, but also directly from clients' digital expectations. This includes the most trivial requests such as using videoconference applications or cloud collaboration tools as well as providing complex docu-

---

54 Omni Legal, 'Artificial Intelligence Won't Replace Lawyers—It Will Free Them' (Law Technology Today 27 February 2018).

55 Richard Susskind, Tomorrow's Lawyers. An Introduction to Your Future (Oxford University Press 2nd edn, 2017).

ment automation solutions. For law firms to meet these expectations, they need to be prepared to digitize their services by creating and implementing digital transformation strategies.

Digital transformation is a thing of the present and will be a thing of the future. Undoubtedly, legal professionals must follow suit to stay afloat. Advising without understanding clients' business is impossible. Moreover, one of the law industry's success factors called *knowledge asymmetry*, is fading away. Mostly thanks to the Internet proliferation and democratization of knowledge that it brings. Content digitization, search engines, remote learning make access to knowledge easier than ever before.

The change of the law firm leaders will be a key to their success in the near future. The legal industry must open both to innovation AND disruption. Technology is changing business across the board and the pandemic is fueling that process. In 2021 we can expect even more online chatting, greater collaboration in the cloud, more emails and more data to process in general. Legal innovators out there should focus on developing more advanced NLP solutions, privacy preserving technologies for data security and improving remote work modes.