Canon Discovery Services

ARTIFICIAL INTELLIGENCE IN eDISCOVERY:
MARKET FORCES, FUTURE USES, IMPLEMENTATION AND TRUST

AI has steadily been bleeding into areas of legal practices, such as research, document preparation, contract analysis, data analytics, and eDiscovery. Legal innovators view this technology as a requirement for staying competitive and winning work. They realize that the use of AI-based applications and services will generate insights into the relationships of data, process management, and the very nature of legal work that will not be available to those who do not adopt them. These insights will, in turn, drive efficiencies in the preparation and management of the litigation process, delivery of legal work product, and consistent improvements in the quality and reliability of legal work. These efficiencies will provide litigation professionals better overall control and outcomes in cost containment, legal project management, and improved service to clients.

"Artificial Intelligence: The theory and development of computer systems able to perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."*

Legal professionals have been coming to unsteady grips with this new reality for a while. That unsteady grip does not apply to other, retail-focused, AI-based tools. The most skeptical attorney may still follow purchase suggestions from Amazon, use voice-based assistants such as Cortana, Google, or Siri to provide answers or initiate applications, or use autocorrect and grammar recommendations while creating content. These "simple" tools that contain elements of AI, have blended into the background of acceptance and paved the way for other, more advanced, AI-based tools to enter practice. While AI in the practice is no longer in its infancy, it has yet to fade into this background acceptance, as its application to legal and discovery processes has yet to be smoothly and seamlessly integrated. AI is not a pervasive component of infrastructure, but is localized by application or service, each with a unique use case and behavior. Until an AI framework

Canon

is implemented to coordinate and manage information at an infrastructural level, the implementation of discrete AI-based tools and services will require specialized skill sets and expertise to realize and exploit its growing capabilities in litigation.

AI GROWTH AND ACCEPTANCE THROUGH SEEPAGE

To understand how AI can grow in a practice, it is essential to remember the growth of computing power over just the past 20 years, and how it started slowly and accelerated, with the acceleration having never stopped. Computers were hobbyist passions, then word processors, then processing machines, then smartphones.

The growth of AI is following a similar curve, but at a much faster rate due to the amount of research and financing being poured into its development, and the abundance of data that can be applied to teach these learning engines. A simple illustration of this is the explosive growth in the legal tech startup market. In 2016, there was $224 million invested, and at the close of 2018, there was $1.6 billion of investment capital put forward.[1] This investment will push itself deeply into both the operational and retail markets for legal services. Combine this investment with the explosive amount of data generated by the profession and the courts to serve as training material for AI researchers and developers, and eDiscovery support experts can begin to understand the waves of change AI represents for the practicing litigation professional.

A Thomson Reuters survey of corporate counsel resulted in an assessment that AI tools will be in routine use within ten years.[2] A separate, but complementary analysis by Deloitte predicts that within that same ten-year period, 100,000 legal sector jobs will be automated, with AI-based tools and services providing the foundation for the changes in the market.[3] Three brief examples to illustrate areas of practice that would have been considered to be out of bounds for automation a few short years ago:

- **In a 2016 TED talk Andrew Arruda,** the CEO of Ross Intelligence, provided an overview of AI's reach into legal research.[4] The Ross research AI responds to natural language questions on case law to deliver detailed research information. It can also provide a detailed citation analysis, complete with negative treatments and recommendations of other case law from a submitted brief, or generate memorandums on legal issues to further an attorney's understanding of an issue.

- **Ravel Law (now branded as Context)** created an AI-based judicial and expert witness analytics tool capable of providing detailed information on citation patterns, common language use in judicial decisions, decisions by the attorney, and expert witness utilization assessments and challenges.

- **Scissero** is a document analysis and contract drafting tool designed to accelerate the creation of complex contracts, provide due diligence assessments, create risk assessment summaries, and provide support for large-scale repapering projects.
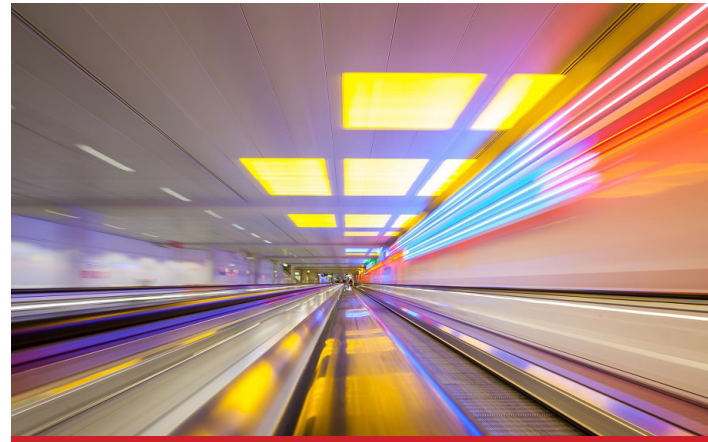
AI IN eDISCOVERY AND THE POTENTIALS FOR USE IN COMPLEX ESI ASSESSMENTS

Like other areas of legal practice, AI tools started with a small footprint in eDiscovery as a "simple" search tool used to sift megabytes of structured data to locate specific terms or keywords in a given data set, and identify potential duplicates or relationships. It evolved into the review phase with technology-assisted review (TAR) and predictive coding. With TAR 1.0, AI tools used human guidance and feedback during a defined training stage, and based on static data models that prepared it to support an upcoming review. In TAR 2.0, continuous active learning was unleashed to continually monitor a review from the outset, and automatically refine its understanding of the data, the demonstrated intent of the review decisions, and from that provide its own insights about the relevance of a document. It would then feed those insights back into the review as an ongoing process of refinement.

Using machine learning in these areas has enabled AI to become a tool to support and enhance an eDiscovery professional's expertise in quickly locating and identifying relevant patterns and information in collected data sets. An excellent example of the recognition of these abilities by the courts is the Proctor v. Safeway, Inc. matter wherein a document production was a point of conflict, as it was deemed not to have met discovery requirements.[5] In this matter, the court mandated the use of a technology-assisted review to identify all relevant documents, rather than a human-based document-by-document review.

The opposite side of this pattern matching and relevance evaluation, moving into the future, is the ability to recognize broken patterns, such as unexpected shifts in the use of email. Gaps in sequence or shifts in frequency, changes in word use behavior, and other indicators of omission or deceit can be teased out and appraised. Finding these anomalous breaks are the red flags that can shift the definition of what information may be relevant in an investigation or review.

From its evolution as a tool to locate data, to then assisting in the identification of relevant documents, to revealing shifts in patterns, AI-based tools are being created to recognize more nuanced classifications of information, and identify inferences that would be missed in a human linear review. This is driven by the ultimate goal of locating facts from more tenuous elements that may be detectable across a larger range of sources. The next step toward this goal is the creation of applications that can identify tones in communication such as concern, satisfaction, or deceit. New AI tools can now take these patterns, whole and broken, word use and context, and inclusions and omissions to isolate these inferences for additional investigation. Companies, such as KeenCorp, have developed technology capable of assessing the overall morale or intent of an organization by the analysis of emails, while researchers at Florida State University have built a framework capable of detecting deception in text messages[6] with an 85% accuracy rate. The ability to assist in a search, to conduct a search, to logically aggregate and present data, to provide new types of context, and to then determine deceptions, intent, or mood defines the progression of AI in eDiscovery.

AI has continued to spread and become a strategic component at each stage of the eDiscovery process. Its ability to process and filter large unstructured data sets into discrete management groupings through email threading, intelligent batching, and near-deduplication allows a litigation team to rapidly assess the data forming the foundation of a potential case or to better prepare a complaint as part of an investigation. These AI-framed early case assessments, and the decisions humans make that are driven by them, can be fed back into the AI tool to provide additional learning data for it to use for future, similar projects, further improving subsequent assessments, and outcomes.

AI advances in natural language recognition bring concept-clustering forward as a tool to recognize nuances in the order or phrasing of terms, assembling them into a relationship or proximity sets that can refine the view of data beyond simple keyword patterns. These capabilities to learn and apply lessons gleaned, either from human interaction or provided example data, pave the way to deal with the onslaught of new data sources, and their relationships with each other and the end users who manage or interact with them.

These capabilities are being tested against the steadily rising tide of potentially discoverable data. People have become adapted to information that is created by the formal use of content production tools in the office, such as desktop or mobile applications. Now the courts and eDiscovery professionals are learning to cope with additional streams created by casual data creation tools such as social networks and texting, automated systems such as GPS tagging and IoT-based devices, along with the forest of audio, video, and images. The convergence of these classes of data creates a chaos of information from which an eDiscovery professional is expected to collect, tame, and produce relevant results for a client.

"AI eDiscovery tools have the ability to not only quickly sort and organize this chaos of data, but can then provide detailed pattern matching and relationship analytics to better understand it. These abilities allow an experienced eDiscovery professional to track threads of relevant information across all of the identified and collected generating sources."

This thread of review may start as a document, that is then referenced in an email that may generate some texts that may then need to be analyzed in the context of a video deposition or a comparable stream of media. At each stop along this path, meanings and intent may change, the language will shift to accommodate a given culture, and context may evolve. This is a prime example of the reach and analysis that AI can bring to the preparation of a legal argument.

This ability to collect, sort, prioritize, and present complex information and relationships for evaluation to experienced eDiscovery teams will drive AI adoption, going forward.

IMPLEMENTING AI TOOLS AND SERVICES IN LITIGATION

Discussions on the use and value of AI in litigation can create anxiety within some firms or offices of inside counsel. Visionaries such as Stephen Hawking and Elon Musk have voiced significant concerns on the misuse of AI. As mentioned previously, market analysts have

revealed patterns detailing the amount of impact on employment generally and the legal industry in particular. To add to the unrest, there are articles that detail how an AI-based system may have embedded and subtle biases, based on poor reference data or acquired by programming decisions implemented by their designers.[7] These concerns highlight the significant cultural hurdles to overcome when legal professionals look at bringing AI into a practice.

Discovery support professionals, recognizing the impact AI has on the discovery process with technology-assisted review, can, therefore, meet resistance when proposing it to assist in more analytic endeavors. Their role is to understand an issue, and to suggest a solution that is effective, efficient, and provides value to their clients. These systems, using frameworks built around cognitive intelligence models that are acquired learning, based on observation and experience, can suggest decisions or provide predictive analytics that seem to infringe on human judgment. These systems are often based around unfamiliar workflows, and can contain algorithms that can challenge or overwhelm a litigator, creating distrust of the provided information and avoidance of its use.

Another way to consider the implementation of AI in practice is to look through the lens of the art of Law. Litigators, while coming from an objective point of view, are often required to be creative or artistic in their approach to a legal matter. In this blending of perspectives, they bring experience and judgment, which AI could mimic, but also empathy, creativity and emotional intelligence that AI currently could not. In this context, AI becomes a tool that discovery professionals use to understand, respect, and support the creative power of the Law and its use in a matter. It then can offer possible solutions that fit within a litigator's overall strategy and vision, and in a manner that can be reliably controlled and executed.

Another path of entry for the use of AI in a law firm is the sheer magnitude of data that discovery professionals and litigators are required to manage in timelines that can be mandated by the courts. Forbes, in a recent article, detailed that 2.5 trillion bytes of data are created each day.[8] This data can be distilled to a large legal matter that may comprise 100 terabytes of data, containing millions of images, hours of media, and incredible volumes of text.[9] With corporations producing such high volumes of data that need to be analyzed, it is quickly surpassing the ability for timely human review.

AI offers the reluctant litigator a range of large data management and review assistance options to provide effective and efficient solutions, while remaining subject to their control:

- **Reducing initial data** set sizes through prioritization

- **Quickly parsing and analyzing** large volumes of data at unmatched speeds and accuracy

- **Categorizing documents and data** types to better stage a review or production

The control of outcomes from these options is placed in the hands of the lead litigator whose expertise is applied to review and assess the results. The time savings at the early case assessment stage can then provide a cushion in the decision to implement a TAR, human, or blended review process. This approach helps mitigate the valid concerns of legal professionals of missing those small or nuanced details that drive the abundance of caution in managing eDiscovery.

In developing a review strategy that leverages the speed, consistency, and accuracy of AI, along with the experienced control of the discovery lead and the litigation team, well-designed workflows become a critical component. Creating active learning workflows that are intuitive and less burdensome can be integrated with traditional linear reviews and supporting analytics technology.

These integrated workflows allow the litigation team to:

• **Better manage time and costs**

• **Work efficiently and effectively** on relevant areas of the review

• **Free up time and resources** to concentrate on legal strategy and trial preparation

• **Lastly, allow a controlled scenario** to develop a trust and understanding of the technology.

Not all cases are ideal candidates for the use of AI-based tools and services. The discovery professional becomes a critical resource to ensure that optimal solutions are provided, based on the nature of the matter and the results being sought. They are in a position to identify risk, address concerns, and propose workflows and supporting processes that can ensure a defensible outcome. Placing them at the beginning of litigation can help ensure that AI, if a valid tool for the matter, can aid in its execution.

As matters are presented to a litigation team, the discovery lead will be assessing the potential scope of the project, and if AI can or will play a role. They will be reviewing the type and nature of the collections that will be needed, along with the data expected to be extracted from them. Based on a variety of factors that drive litigation, they will then design workflows that may incorporate a range of tools and human interaction. As the litigation team moves through the process, the discovery lead or project manager will continuously assist the team in the work. They will aid in the evaluation of the results and the performance of the tools as they relate to the desired outcome.

Should circumstances change, such as a recognized difference in an AI's understanding of the data, the discovery lead will be there to review the changes and see if they are driven by valid differences in the data, which would improve the review through that active learning process. This will in turn provide additional areas of focus for the review going forward. This process of AI-enhanced review, recommendations from the AI, evaluation of the recommendations by the litigation team, and then implementing them into the review, is cyclic. With each day, new information can be presented and create valuable review targets for the next cycle.

Should a change be created by a human-driven event, the discovery lead would assess the potential impact of the event and work with the litigation team to make adjustments, or provide the additional information to the AI as a training set to help guide the review going forward. Bringing the tools to bear at the outset of a change to assess its impact, can save time and reduce costs by keeping the reviewer focus on the documents, while the AI tool provides this new data into the review workflow.

The expertise and technical resources offered by Canon Discovery Services can support litigation teams with the creation of these AI-based workflows. These workflows will process the vast amount of data, and generate insights for the discovery lead and the litigation team to assess, as a review is being developed or implemented. This will ensure that AI remains a managed tool for discovery, and not become the engine of discovery without guidance from the litigation team. This balanced approach will build the trust and acceptance of AI into a practice and provide a foundation for continued innovation, going forward.

About Canon Discovery Services

Canon Discovery Services has a skilled, dedicated team of discovery professionals with a proven track record in solving complex discovery matters. Backed by over twenty years of experience, we help law firms and corporate legal departments develop practical, defensible eDiscovery response plans to support successful outcomes. Our services range from ESI processing, culling and analysis, document review, hosting, and production to implementing information governance and readiness response programs. Canon Discovery Services is a part of Canon Business Process Services, a subsidiary of Canon U.S.A. Visit us at cbps.canon.com/managed-services/discovery-services.

This paper is written by Lincoln Mead and Gisselle Singleton of Canon Discovery Services.

[1] Valentin Pivovarov, 713% Growth: Legal Tech Set an Investment Record in 2018 (Forbes Magazine January 2019)

[2] Thomson Reuters Report, Ready or Not: Artificial Intelligence & Corporate Legal Departments (2017)

[3] Deloitte Report, Developing Legal Talent Stepping Into the Future Law Firm (2016)

[4] Andrew Arruda, ESQ., The World's First AI Legal Assistant (TED Talk November 2016)

[5] THOMAS PROCTOR, Plaintiffs,v.SAFEWAY, INC., Defendant.No. 11-cv-3406 March 08, 2018: https://ediscovery.co/wp-content/uploads/2018/03/U.S.-ex-rel.-Proctor-v.-Safeway-Inc._2018-03-18-21_24_55-0400.pdf

[6] Shuyuan Mary Ho and Jeffery T. Hancock, Context in a bottle: Language-action cues in spontaneous computer-mediated deception, Computers in Human Behavior, February 2019

[7] Harvard Business Review - Addressing the Biases Plaguing Algorithms, Michael Li, MAY 13, 2019: https://hbr.org/2019/05/addressing-the-biases-plaguing-algorithms

[8] Marr, B. "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read", 11 Mar. 2019, www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#3611044160ba

[9] Borgström, Anna. "Investigations with Terabytes of Data and Millions of Images." NetClean.com, 27 Feb. 2018, www.netclean.com/2017/08/07/investigations-terabytes-data-millions-images/.

*Siri-provided definition

This whitepaper is not intended to provide any legal advice.