How AI Transforms the Way Law Firms Profile Documents

An emerging technology called artificial intelligence (commonly referred to as AI), is fundamentally changing the way law firms profile and manage the documents they receive.

At a time when law firms are under pressure to rethink their business models, using AI for profiling will help them:

- Increase staff productivity
- Improve business outcomes
- Accelerate cycle times
- Reduce errors
- Enhance service

Automating routine tasks with AI also frees law firm staff to focus on complex, higher-value work.

Artificial intelligence enables software to understand tasks by producing for itself the rules for performing a task that programmers often cannot specify. Artificial intelligence was especially developed for:

- Automating repetitive tasks
- Analyzing large amounts of data
- Identifying features of interest in data

This white paper describes how law firms are struggling to capture and manage their information, how AI can address one aspect of the information management challenge, an example of how a law firm can apply the technology, the components of an AI solution, and the benefits law firms can achieve by managing information with AI technology.

> Insisting on the necessity of receiving more value for their "legal spend," [since the onset of the Great Recession in 2007, corporate] clients increasingly emphasized the need for greater efficiency, predictability, and cost-effectiveness in the legal services they received.
> - Georgetown Law report on the "State of the Legal Market"

Canon

CANON INFORMATION & IMAGING SOLUTIONS

The Problem

Like many businesses, law firms are struggling to manage the increasing amounts of information pouring in everyday.

37% of organizations report that their offices are "piled high" with paper documents, per the Association for Intelligent Information Management (AIIM). In fact, despite the push towards electronic documents, **19%** of organizations are receiving more paper documents.

Organizations also face soaring volumes of digital data. Some 2.5 quintillion bytes of data are created every day, IBM reports. Ninety percent of all data was created within the last two years.

The combination of the surprising resilience of paper documents and skyrocketing data volumes is making it hard for law firms and other organizations to effectively manage mission-critical information as it arrives. For instance, only **32%** of organizations can automatically analyze their text data, Forrester Research reports. And even fewer organizations can analyze their unstructured data.

This is a "speed bump" in the legal industry's move towards digital transformation.

Incoming documents – whether they arrive as paper or as an e-mail attachment – typically require a human operator to determine the type of document, manually enter important data from the document into a downstream system, notify the appropriate individuals that the document has arrived, assign the correct matter number to the document, and appropriately store the document within the enterprise.

This process is called profiling.

A document's "profile" needs to describe its type, the important data it contains, and its contextual relevance within an information ecosystem. Associating these informational attributes with a document allows it to be easily and comprehensively identified, understood, and managed.

Profiling creates an associated set of metadata "tags" that describe the document's type and indexes its key information and context. Metadata "tags" at a law firm may include information such as:

- The document author/source
- The type of document
- The client involved
- The date of the document

Profiling improves the "findability" of documents in the same way web pages are tagged so that search engines like Google can easily find them, and it also facilitates automated processes.

Leading law firms and other organizations may already be using software that extracts a document's key metadata using pre-defined rules. Software with this approach is effective mostly with structured documents (e.g. invoices, questionnaires, job applications) with clearly labeled data fields.

The problem is that law firms primarily receive unstructured documents (e.g. company bylaws, meeting minutes, memorandum of understanding, business plans, contracts) that do not have clearly labeled fields; critical information can appear anywhere in the text of an unstructured document.

Complicating matters, a full profile for each document has to also include contextual information that's not contained within the document.

Contextual information may include:

- The case/matter number
- The lawyers involved in that matter
- The document storage location
- Billing codes

Because of the domain knowledge it requires, adding contextual information is a manual process at most law firms – a laborious and error-prone task.

Metadata is essentially the 'who,' 'what,' 'when,' and 'where' of a document

AI can automate tasks traditionally undertaken by paralegals or legal secretaries

The Solution

This is where AI comes in.

AI can be taught to classify incoming digital documents and paper documents that have been scanned and processed using optical character recognition (OCR) technology. The AI technology trains itself to classify documents by their type, based on samples of each document and/or by observing how a human operator classifies documents. Over time, an operator's acceptance or correction of the software's predictions increases its accuracy at predicting a document type (e.g. contract, correspondence, non-disclosure agreement, stipulation, resume). Each document classification prediction is based on all the words and phrases contained in the document, not just on the appearance of a few keywords, or on its physical layout. Analyzing all the words and phrases in a document enables very similarly structured documents to be uniquely identified and classified.

After AI learns to classify document types, the software can extract the relevant information and facts (metadata) for that type of document, and this informational metadata can be identified within both structured and unstructured documents, regardless of where it is located within the text.

Next, the AI needs to determine a document's contextual relevance within a firm's information ecosystem by learning to associate the document with specific clients, lawyers, cases and other information. This is a task that is still performed manually in nearly all legal firms since it typically requires human judgment based on domain knowledge of the firm, its activities, and information systems. The technology accomplishes this by reviewing and analyzing the profiles of all the documents already stored on the law firm's servers by examining both their internal metadata and additional contextual metadata that has been added to them, or implied for them based on the names of the hierarchy of folders in which each document is stored.

> A lot of legal documents look alike. That's why traditional document classification solutions that analyze the visual layout of a document – but not all its content – can recognize a structured invoice but cannot easily distinguish between a Stipulation and a Motion in Limine

> Nearly all legal firms still utilize humans to manually add the metadata to a newly arrived document that provides its context to the firm – its matter number, a client name, the lawyers involved, etc.

Importantly, AI can leverage a law firm's existing index of metadata about its documents as created by its document management system, eliminating the need for the law firm to provide a vendor with permission to comb through confidential information to create new indexes.

Once the AI has analyzed all the metadata about the documents a law firm has stored on its servers, the technology simply searches for existing documents with similar metadata as is contained in the new document, and recommends the same contextual metadata as those existing similar documents. It's usually this contextual metadata that is used to initiate automated workflows, such as how to name the document file, automating its storage in an appropriate location or folder, and even to notify the appropriate lawyers and staff members about a document's arrival.

A Use Case Example

To understand the power of AI in the legal profession, consider the case of a law firm that receives paper and e-mail attachments in its mailroom. The paper documents get scanned and electronically routed for processing. The AI solution first extracts the metadata from within the paper and electronic documents and also classifies the documents by type. The AI solution then uses the extracted metadata to find similar documents in the law firm's document management system to make predictions on the contextual metadata that ought to be associated with the documents. An operator would then approve the contextual metadata, with any low confidence predictions about metadata to be added being highlighted or flagged to ensure they are reviewed. The document is then stored with all its associated metadata in the law firm's document management system, while any automated notifications or workflows are triggered.

Elements of an AI Solution

AI solutions combine five critical components to automate document profiling:

COMPONENT #1: NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) analyzes each word by examining its relationships within a block of text (e.g. neighboring words, punctuation, etc.) to identify words and conceptual phrases that are key data (e.g. people, places, locations, dates, numbers, etc.). Unlike software that extracts metadata from structured documents, NLP neither requires that critical information be labeled (e.g. "Name:" or "Case Number:") nor relies on its position on a page. This ability to find important information anywhere within a document enables law firms to more accurately categorize and extract relevant information from the highly unstructured documents normally found in law offices.

COMPONENT #2: PREDICTIVE ANALYTICS

Predictive analytics is another key part of an AI framework for law firms. Predictive analytics are statistical techniques used to predict the likelihood of an outcome based on known facts or historical data. For instance, the technology is making predictions about which contextual metadata to associate with a document based on the historical associations made for previous documents. "Confidence" levels are assigned to each prediction based on how well aligned a new document is with existing documents, enabling any low confidence predictions to be flagged for verification by a human.

COMPONENT #3: MACHINE LEARNING

Machine learning enables AI to learn and improve upon its ability to correctly make predictions on document classifications and metadata associations. When a user corrects or accepts a predicition, it creates a machine learning feedback loop that enables the technology to automatically adjust its prediction-making model and improve accuracy over time. As a result, profiling accuracy can quickly increase from between 60% and 65% for a newly installed solution to between 90% and 93% as users operate the software.

COMPONENT #4: PRE-TRAINED ON A RELEVANT DOCUMENT LIBRARY

Of course, any technology that needs to learn, must be trained on the types of documents it will process. That's why some AI solutions have been pre-trained on a comprehensive set of example documents for the industry they will serve. From that baseline, the software can be quickly trained to recognize and classify a law firm's unique document types and be trained to recognize/extract internal company acronyms.

COMPONENT #5: CONFIDENTIALITY

As with other enterprise applications, AI solutions must safeguard confidential information by synchronizing with the secure user logins and permissions specified by the law firm's Microsoft Active Directory or perhaps as specified for something like their Box cloud storage. To ensure confidentiality, these AI solutions are usually deployed on-premise behind a law firm's firewall or in a private cloud.

The Benefits

Automating the profiling of scanned and digital documents using AI provides significant benefits:

- Improved staff productivity: AI automates tasks typically performed by law firm staff, such as classifying legal documents by type (e.g. stipulation, contract, non-disclosure agreement), and associating documents with corporate contextual information (e.g. matter numbers, client names, file names, etc.).

- Better business outcomes: The technology also is pre-trained to recognize a variety of litigation and court document types and can be trained to recognize company-specific document types. An AI solution can predict document types based on all identified words and phrases, not just keywords so that unlike visual layout-based recognition technologies, the AI can identify very similarly structured documents (e.g. can distinguish a "Motion in Limine" vs. a "Stipulation").

- Faster cycle times: Automatically identifying a document's type and intelligently associating relevant contextual information that may not be contained within the document itself makes it easier for documents and the information they contain to rapidly enter workflows. AI solutions can also accelerate workflows by automatically notifying the appropriate individuals about new documents of interest, and by electronically routing documents to a docketing application.

- Fewer processing errors: AI solutions typically allow users to verify the accuracy of a document's associated metadata or add additional information through a Web-based user interface or sometimes within their Outlook email window using an add-in before delivering a new document to the law firm's document management system. Automating metadata solutions also ensures a consistency across an organization that makes organizing and finding documents easier.

- Enhanced service: A continuous "machine learning" feedback loop enables AI solutions to automatically incorporate new client names and matter numbers. This is especially useful for dynamic law office environments with ever-changing client/matter lists.

Some Final Points About AI in Business

Most CEOs worldwide agree that AI will ultimately impact every facet of business, offering an unprecedented opportunity to innovate and grow companies, PwC reports.

Organizations that embrace digital transformation earn 26 percent higher profits, per Accenture.

AI already is streamlining the way that law firms profile their incoming documents as they enter digital storage and workflow solutions to deliver improved staff productivity, better business outcomes, faster cycle times, fewer errors and enhanced service.

The stakes have never been higher for firms to embrace digital technologies such as AI. Fifty-four percent of digital transformation leaders believe if they are not successful, their company will be out of business or absorbed by a competitor within four years, per Couchbase/Vanson Bourne.

AI can help ensure that law firms can meet their information challenges today and in the future.

> AI will drive gross domestic product gains of $15.7 trillion by 2030, per PwC

> "Those firms that are able to adjust to the new market realities… by engaging in a thoughtful review and (where necessary) redesign of their approaches to client service, pricing, legal work processes, talent management, and overall structure will enjoy an enormous competitive advantage."
> - Georgetown Law report on the "State of the Legal Market"

Canon
CANON INFORMATION & IMAGING SOLUTIONS

About Canon Information and Imaging Solutions, Inc.

Canon Information and Imaging Solutions, Inc. (CIIS), a wholly owned subsidiary of Canon U.S.A., Inc., brings together Canon's world-class imaging technologies and information management expertise to assist organizations in achieving their digital transformation objectives. With a focus on innovation, CIIS's software development and solutions delivery capabilities scale across several practice areas: Business Process Automation - including Procure-to-Pay & Order-to-Cash automation, Document Solutions, Information Management Services with a focus on content capture, management and collaboration, and Security and Infrastructure Management. With expertise in emerging technologies such as artificial intelligence, machine learning, and big data analytics, CIIS deploys its solutions in partnership with leading technology providers and offers comprehensive consulting and professional services that are trusted by organizations of all sizes. To learn how CIIS can transform your business processes today and in the future, visit www.ciis.canon.com.

For more information, please contact us at info@ciis.canon.com or visit www.ciis.canon.com

// LOOK FORWARD