



Use of Natural Language Processing (NLP) in Civil Case Management:

A Report on Three Proof of Concept Projects

By Paula Hannaford-Agor & Jannet Okazaki

May 2023



Contents

Acknowledgements..... **iv**

Introduction..... **1**

NLP Triage POC **4**

Quality Control POC..... **12**

Conclusions and Recommendations..... **15**

Use Cases **19**

Appendix A: POC 1–Civil Case Data Extraction and Case Matching POC..... **27**

Appendix B: POC 2 – Civil Case Triage POC **29**

Appendix C: POC 3 –Civil Consumer Debt Cases, Quality Control POC..... **30**

Appendix D: Civil Case Triage Criteria **34**

Acknowledgements

We would like to express our sincere gratitude to all those who have contributed to this proof of concept (POC) study on the use of Natural Language Processing (NLP) in civil case processing in state courts. First and foremost, we want to acknowledge the CCJ Civil Justice Improvements Committee for their recommendations to leverage technology to support effective case management. Their vision and dedication to improving the civil justice system have been instrumental in inspiring this project. We also are thankful for the attendees at the 2017 Court Technology Conference who suggested that the use of NLP to extract data directly from case filings might perform better than data extracted from court case management systems for a range of essential case processing tasks. Their insights and perspectives have been invaluable in shaping the direction of this study.

A great many individuals helped us throughout the study. We benefited greatly from the insights and suggestions of our project advisory committee members who spent two long days in a dark conference room helping us outline the requirements for the POC: Roberto Adelradi (Eleventh Judicial Circuit Court of Florida), IV Ashton (LegalServer), Judge Jennifer Bailey (Eleventh Judicial Circuit Court of Florida), Katherine Birchfield (McHenry County Circuit Court, Illinois), Chief Magistrate Gregory Clifford (Cleveland Municipal Court), Margaret Hagan (Stanford School of Design), Judge Steven Houran (Stafford County Superior Court, New Hampshire), Casey Kennedy (Texas Judicial Branch), and Kelly Steele (Ninth Judicial Court of Florida). We also owe a debt of gratitude to Judge Gina Beovides (Eleventh Judicial Circuit Court of Florida) who provided

feedback to the vendors during the machine learning phase of the project; to our research interns Camden Kelliher, Laura Acker, and Madeline Williams who spent many hours manually coding data from civil case filings; to our NCSC colleagues for their support and collaboration throughout the project, especially Jim Harris, Barbara Holmes, Allison Trochesett, Sarah Gibson, and Keeley Daye; and to Henry Sal, Jr. of Computing Systems Innovations and Abhinav Sonami of Leverton Intelligence, the commercial vendors who donated their time and talents to participate in the POC.

We want to express our heartfelt appreciation to the Superior Courts of Arizona in Maricopa and Pima Counties, the Fifteenth Circuit Court of Florida (Palm Beach), and the Cleveland Municipal Court, which provided exceptionally large troves of court documents for this study, and to Darren Dang, Karen Hernandez, and Brett Howard in the Superior Court of Orange County, California and to Richard McHattie of the Superior Court of Arizona in Maricopa County for showing us how NLP can work in real court environments. Finally, we are grateful to the State Justice Institute both for its financial support (SJI 18-P-020) and for its great patience as we struggled to complete this project in the midst of a global pandemic. We are confident that the lessons learned will benefit courts for many years to come.

The views expressed in this report are those of the authors and do not necessarily represent those of the State Justice Institute, the National Center for State Courts, or the individual courts, court staff, or vendors who participated in the project.

Introduction

Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics that employs predictive analytics and machine learning with a focus on the interaction between computers and both written and spoken language. First developed in the 1950s, NLP has become increasingly sophisticated over the past two decades as computational power has increased. Because legal language tends to be more structured in format than other linguistic forms, NLP applications have become particularly useful for a variety of law-related tasks. For example, NLP is the primary technique employed in e-discovery to identify documents related to a specific query based on keywords or phrases. This technology is also being used to extract information from multiple documents to assess variation in key data elements for risk management purposes.¹

Although courts are vast repositories of legal documents, they are only recently implementing predictive analytics and machine learning techniques, including NLP, to support court operations. For example, one area in which NLP has shown particular suitability is the task of redacting information disclosed in court documents to protect the privacy interests of litigants and vulnerable third

parties, including children.² More recently, courts have begun to explore the potential benefits of NLP and other tools such as data extraction and robotic process automation (RPA) for a variety of case processing tasks. Maricopa County Superior Court, for example, has used these techniques to extract information from both paper and electronic documents to enter onto the court's case management system (CMS). The Superior Court in Orange County, California is training these tools to recognize different subtypes of default judgment motions so that clerks do not have to open the electronic documents to verify the type of default sought by plaintiffs.

In 2016, the Conference of Chief Justices (CCJ) and the Conference of State Court Administrators (COSCA) endorsed recommendations to leverage technology to improve civil case management.³ In particular, NLP and related tools could be used to support two areas of civil case processing: sorting cases at filing based on the anticipated level of judicial involvement in case management, and confirming that essential procedural requirements have been satisfied before entering final judgments in cases.

1 Lahr Mahler, *What is NLP and Why Should Lawyers Care?* ABA LAW PRACTICE TODAY (Feb. 13, 2015) (<http://www.lawpracticetoday.org/article/nlp-lawyers/>).

2 See, e.g., TOM CLARKE et al., AUTOMATED REDACTION PROOF OF CONCEPT REPORT (NCSC Sept. 2017).

3 CIVIL JUSTICE IMPROVEMENTS COMMITTEE, A CALL TO ACTION: ENSURING CIVIL JUSTICE FOR ALL (NCSC 2016).

Previous efforts to automate civil case triage based on information extracted from CMS were only moderately successful in assigning cases to the correct track, in part because many of the data elements that experts believe are related to case complexity are not routinely captured in CMS. In addition, CMS data elements often lack sufficient precision to make meaningful distinctions between cases of varying complexity.⁴ NLP might overcome many of the limitations of CMS data in civil case triage by identifying and extracting data directly from case pleading documents. Indeed, NLP could capture a great deal more information than CMS data such as the number and nature of legal claims asserted and relief requested by the plaintiff, the defendant's response to each claim including the number and nature of affirmative defenses, counterclaims, crossclaims, and third-party claims. Collectively, this information could be used to determine the level of legal and interpersonal conflict between the parties and the anticipated volume of discovery, both of which are recognized as important factors in pathway assignment. The utility of these technologies for identifying factors related to case complexity might even be extended across

multiple cases, for example, by identifying individual litigants or attorneys who are more likely to require judicial direction or oversight. These technologies might also be able to identify external trends that contribute to individual case complexity, such as changes in case law, the regulatory environment or even the business practices of significant justice system stakeholders.

Courts also struggle to ensure quality decision-making in high-volume court dockets such as small claims, landlord/tenant, consumer debt collection, and mortgage foreclosure. The overwhelming majority of defendants on these dockets are self-represented and lack the legal expertise to challenge improper claims or raise legitimate defenses.⁵ NLP could be used to identify information in case documents that signal the need for additional scrutiny during in-court hearings or before entering default judgments. Such information could include inconsistent information (e.g., different defendant names or addresses on the complaint, the contract, and the service return affidavit), or the absence of essential information with the complaint (e.g., copy of

4 CIVIL JUSTICE INITIATIVE: CRITERIA FOR AUTOMATING PATHWAY TRIAGE IN CIVIL CASE PROCESSING (NCSC 2017).

5 During the mortgage foreclosure crisis in 2009-2010, many courts discovered widespread problems in court filings, including lack of standing to foreclose on the property, incomplete mortgage servicing records, and fraudulent certifications (e.g., robo-signing) of mandatory disclosures and documents. See, e.g., Maria Wang, *GMAC's 'Robo-Signers' Draw Concerns About Faulty Process, Mistaken Foreclosures*, PROPUBLICA (Sept. 29, 2010); Stacy Cowley & Jessica Silver-Greenberg, *Behind the Lucrative Assembly Line of Student Debt Lawsuits*, N.Y. TIMES (Nov. 13, 2017), available at <https://www.nytimes.com/2017/11/13/business/dealbook/student-debt-lawsuits.html>; MARY SPECTOR, *Default and Details Exploring the Impact of Debt Collection Litigation on Consumers and Courts*, 6 VA. L. & BUS. REV. 257, 285 (2011); Peter A. Holland, *Junk Justice: A Statistical Analysis of 4,400 Lawyers Filed by Debt Buyers*, U. MD. FRANCIS KING CAREY SCHOOL OF LAW LEGAL STUDIES RESEARCH PAPER, No. 2014-15; FEDERAL TRADE COMMISSION, *REPAIRING A BROKEN SYSTEM: PROTECTING CONSUMERS IN DEBT COLLECTION LITIGATION* (2010).

original contract, proof of standing, proof of timeliness, active military affidavit, or missing or incorrect documentation of damages and fees).

To explore the feasibility of NLP to support court operations in these two areas, the National Center for State Courts (NCSC) designed three distinct Proof of Concept (POC) projects. NCSC partnered with three general jurisdiction courts that participated in the CJI automated civil case triage project to use NLP techniques to identify and extract key terms and characteristics from the case pleadings for use in assigning cases to an appropriate civil case processing track.⁶ For quality control over high-volume dockets, the NCSC worked with the Cleveland Municipal Court on a POC to identify inaccurate or missing information from case documents in its consumer debt collection docket that would signal the need for increased judicial review. The NCSC partnered with two vendors that specialize in NLP technologies to control for variation in vendor quality. In addition, NCSC interviewed IT staff in the superior courts of Maricopa County, Arizona and Orange County, California about their experiences implementing these technologies for purposes similar to the POCs.

⁶ The courts that participated in the automated civil case triage project included the Arizona superior and justice courts; the Missouri circuit courts; and the Palm Beach, Florida circuit and county court.



NLP Triage POC

The previous study of automated civil case triage found that CMS data elements either lacked sufficient precision to make meaningful distinctions between cases of varying complexity or were not recorded in CMS at all.⁷ The most important data elements for triage purposes were those related to case type; the number of parties; the defendant's response, if any, to complaint allegations, including crossclaims, counterclaims, and third-party claims; and the defendant's representation status. The NLP Triage POC was designed to test whether NLP could extract those data elements from case pleading documents (complaints and answers) with sufficient accuracy and precision to employ the triage criteria developed in the automated civil case triage study.

In preparation for the Triage POC, NCSC assembled electronic copies of case pleadings from three of the general jurisdiction courts that participated in the automated civil triage

study.⁸ Case pleadings have both structured and unstructured elements. In all three courts, pleadings included a case heading on the first page featuring the name of the court in which the document was filed; the type of document (e.g., complaint, answer); the case number; the case title (plaintiff(s) name v. defendant(s) name; and the name, contact information, and bar number of the attorney filing the document. Figure 1 illustrates a typical case heading. A date stamp showing the date and time the case was filed generally appears on the upper right-hand corner of the document. The content of the documents following the case headings was a semi-structured narrative outlining the plaintiff's alleged facts of the case (complaint) or the defendant's responses (answer), the legal claims or defenses, and the relief sought, including demands for a jury trial.

7 CRITERIA FOR AUTOMATING PATHWAY TRIAGE IN CIVIL CASE PROCESSING, *supra* note 4.

8 Maricopa and Pima County Superior Courts in Arizona, and the Fifteenth Judicial Circuit Court of Florida.

Figure 1: Complaint Filed in Superior Court of Arizona, Maricopa County

18
A

1 CHANDLER, TULLAR, UDALL & REDHAIR, LLP
ATTORNEYS AT LAW
1700 BANK OF AMERICA PLAZA
2 33 NORTH STONE AVENUE
TUCSON, ARIZONA 85701-1415
3 TELEPHONE (520) 623-4353

4 Edward Moomjian II
5 State Bar No. 16667

6 LAVOY & CHERNOFF, PC
7 Christopher A. LaVoy
8 State Bar No. 16609
9 Two Renaissance Square
40 North Central Avenue, Suite 2800
10 Phoenix, Arizona 85004
Telephone: (602) 253-3337
Facsimile: (602) 253-3389

11 Attorneys for Plaintiff

12 **IN THE SUPERIOR COURT OF THE STATE OF ARIZONA**

13 **IN AND FOR THE COUNTY OF MARICOPA**

14

15 ESI ERGONOMIC SOLUTIONS, L.L.C., an
Arizona limited liability company,
16 individually and on behalf of all others
similarly situated,)
17)
18 Plaintiff,)
19 v.)
20)
21 UNTIED ARTISTS THEATER CIRCUIT,
INC., a Maryland corporation; AMERICAN
22 BLAST FAX INC., a Texas corporation;
JOHN AND JANE DOES I-V, BLACK
23 CORPORATIONS I-V, AND WHITE
PARTNERSHIPS I-V,)
24)
25 Defendants.)

MICHAEL K. JEANES
Clerk of the Superior Court

By DONNA PANKEY, Deputy
Date 11/18/1999 Time 10:51 AM

Description	Qty	Amount
CASE# CV9920649		
CIVIL NEW COMPLAINT 001		140.00
TOTAL AMOUNT		140.00

Receipt# 00003059753

NO. CV 99- 20649

COMPLAINT

[FEDERAL TELECOMMUNICATIONS ACT, 47 U.S.C. §227; CLASS ACTION]

Assigned to Judge _____

The Triage POC involved two components (Appendix A). The first component was purely a data extraction exercise to identify and extract case information from the pleadings that would permit judges or trained court staff to assign cases to a case processing pathway based on the formulas developed in the automated civil case triage project. Table 1 displays the key data elements.

The second component was a relational data test to match cases based on the court and case number, to compare the number of defendants named in the complaint and answer, and to identify differences in the number of parties, names, or litigant types. In terms of civil case processing, this information would indicate

whether a case was “fully joined” – this, that all named defendants had responded to the initial complaint – and the court should issue a case scheduling order or set a date for a case management conference to establish expectations for the litigation process.

A second Triage POC invited vendors to use AI tools either to review and triage cases based on the NCSC formulas or to develop and test a new model based on predictive analytics. This POC essentially asked the vendors to identify and do computational processes of key data to create information pertinent to case management processes such as counting the number of defendants. These computations were then used to triage the case into a specific path.

Table 1: Data Elements Extracted in NLP Triage POC

Complaint
Court in which the case was filed
Case number
Filing date
Names and types of first six plaintiffs
Names and types of first six defendants
Unknown defendants included in complaint
Case type
Bar number and law firm name of plaintiff attorneys
Plaintiff demand for jury trial
Amount of compensatory damages demanded
Injunctive relief, punitive damages, attorneys fees or declaratory judgment demanded
Answer
Answer date
Names and types of defendants in Answer
Bar number and law firm name of defendant attorneys
Defendant allegations of crossclaims, counterclaims or third-party claims
Affirmative defenses
Defendant demand for jury trial

NCSC assigned most of the assembled documents to a Learning Set that participating vendors could use in the machine learning phase to teach their software to extract the data elements needed for triage. In this process, an analyst works within the software to identify and label data elements within the documents. Through the iterative process, the machine learns the pattern and reaches a threshold where it can identify the data elements at a high level of accuracy. The learning set included 39,765 pleading documents for 34,796 civil cases filed in the Superior Court of Arizona in Maricopa County; 9,862 pleading documents for 5,004 civil cases filed in the Superior Court of Arizona in Pima County; and 16,632 pleading documents for 13,724 civil cases filed in the Fifteenth Judicial Circuit Court of Florida (Palm Beach County).

Although vendors had the opportunity to ask clarifying questions about the desired data extracts, the Triage POC was more complicated than previous POCs insofar that it required knowledge of civil procedure and terminology. In addition, the learning process was conducted in a static environment (documents saved on NCSC servers) and was based on computer algorithms with limited human review and feedback. Machine learning is an unavoidable and critical first step to train the software. A large volume of representative documents and human review time are required to achieve desired thresholds of accuracy. The level of structure within the documents may also influence machine learning time. For example,

structured forms are easier to learn than unstructured documents.

NCSC selected pleading documents for 250 cases as a Test Set that was released to the vendors at the end of the Learning Phase. Cases selected for the Test Set were weighted toward those with higher complexity index scores to assess the extent to which NLP methods could improve the accuracy of triage pathway assignment compared to the automated civil case triage algorithms.⁹ Twenty percent (20%) of the POC Test Set consisted of cases assigned to the complex pathway compared to 7% of the cases overall; 40% of the POC Test Set consisted assigned to the general pathway and 40% to the streamlined pathway compared to 19% and 75%, respectively, of the cases overall. Due to an error in assigning cases to the Test Set, 26 cases were not manually coded by the NCSC. Consequently, the vendor results reflect 224 usable cases.

Legally trained project staff reviewed the Test Set cases and documented data elements related to case complexity. Using the triage criteria developed in the previous study, project staff also assigned each case to a case processing pathway as well as indicated their recommendation for a different pathway if warranted based on their review of the pleadings. The vendor ran their data-extraction software on the Test Set and submitted it to NCSC project staff to be compared to the manually coded Test Set. The compiled results are reported in Table 2.

⁹ The algorithms developed as triage criteria for the automated civil case triage project assigned 74% of cases to the correct case processing pathway. For incorrectly assigned cases, however, the algorithms more often failed to elevate cases to a higher pathway (22%) than they were to elevate cases inappropriately (4%).

Table 2: Data Extraction Success Rate

	Total N	Correct	%Correct
1st Plaintiff Name	208	206	99.0%
Answer Filed	224	221	98.7%
1st Defendant Name	209	206	98.6%
1st Plaintiff Bar Number	192	189	98.4%
Defendant Jury Demand	112	110	98.2%
Plaintiff Law Firm Name	200	195	97.5%
Damages Unspecified	106	103	97.2%
Plaintiff Jury Demand	208	201	96.6%
Cross Claim	110	106	96.4%
1st Defendant Bar Number	100	96	96.0%
Third Party Claim	111	106	95.5%
1st Plaintiff Type	208	197	94.7%
Counter Claim	111	105	94.6%
Affirmative Defenses	108	102	94.4%
Punitive Damages	214	201	93.9%
2nd Defendant Name	147	138	93.9%
Defendant Law Firm	106	99	93.4%
Attorneys Fees	209	195	93.3%
Injunctive Relief	214	198	92.5%
1st Defendant Type	207	190	91.8%
Answer Date	104	105	91.4%
Declaratory Relief	206	187	90.8%
2nd Plaintiff Bar Number	74	64	86.5%
2nd Defendant Bar Number	36	26	72.2%
2nd Plaintiff Name	60	42	70.0%
3rd Plaintiff Bar Number	30	21	70.0%
Compensatory Damages	95	62	65.3%
Unknown Defendants	100	62	62.0%
Case Type	224	69	39.2%

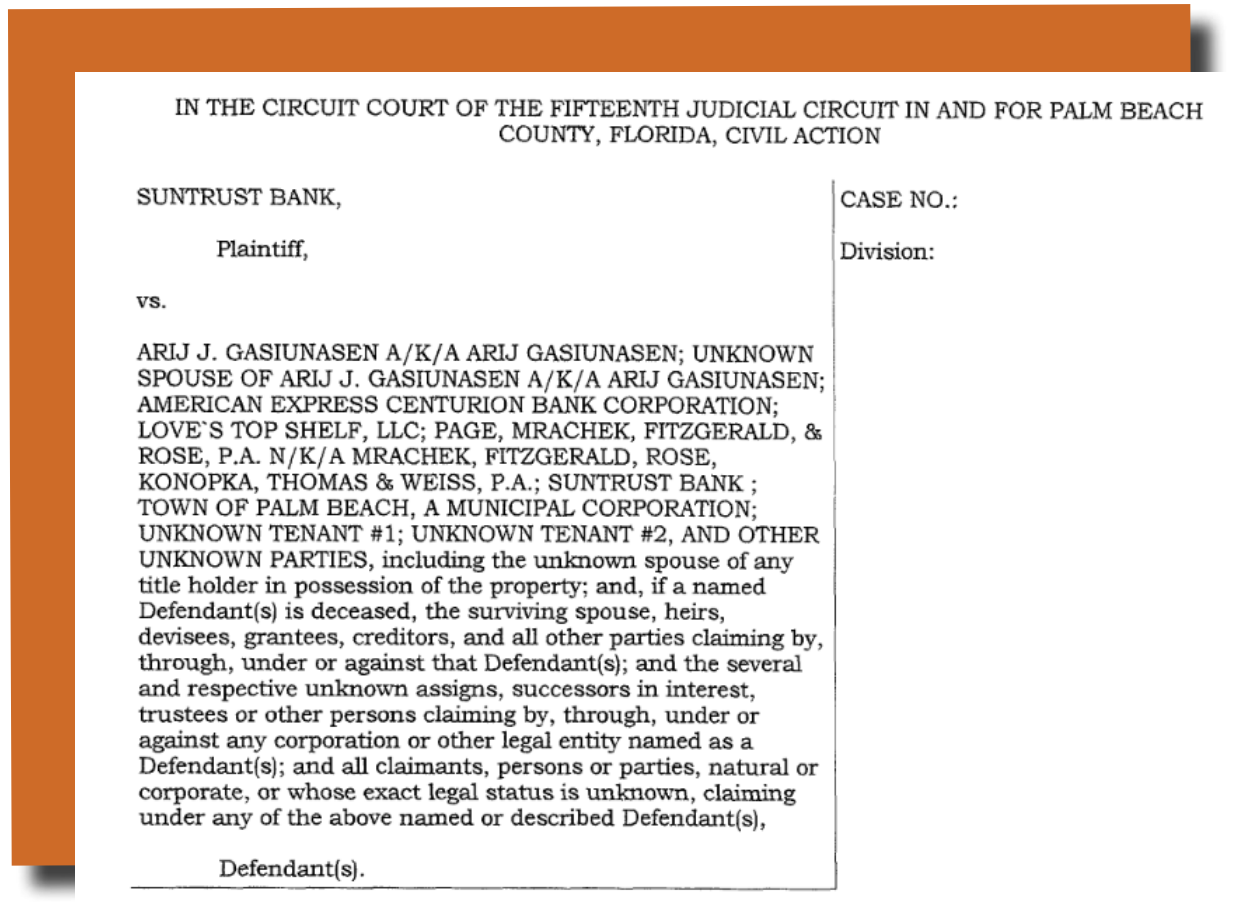
Overall, NLP performed quite well on the data extraction test, correctly identifying most of the requested data elements more than 90% of the time. Many of these data elements were structured or semi-structured data located in the document heading, making them relatively easy to identify and extract. Others, such as demands for jury trials, injunctive or declaratory relief, affirmative defenses and crossclaims, counterclaims, and third-party claims were often only found in the nonstructured narrative sections of the pleadings, but were sometimes set off as subheadings within the documents.

The few instances that NLP extracted incorrect information were most often due to incomplete machine learning concerning idiosyncratic formatting styles employed by lawyers in the participating jurisdictions. For example, many plaintiff lawyers named “John Doe,” Jane Doe,” and “XYZ Corporations I through X” as placeholders in the named defendants

in the event that additional defendants would be identified at a later time, but NLP did not recognize these as “unknown defendants.” Similarly, the use of DBA (doing business as) or AKA (also known as) to designate plaintiff and defendant pseudonyms was often misidentified as a second party rather than an alternate name for the original party. Finally, several smaller law firms filed pleading documents with the names and bar numbers of all licensed attorneys employed by the firm listed on the letterhead; the filing attorney record would then highlight or mark their name to indicate that they were counsel of record on the case. Additional direction during the machine learning phase would likely have corrected these errors over time. If uncorrected, however, those errors would have created additional errors involving calculations for the number of parties, which was a key factor in the triage algorithms.



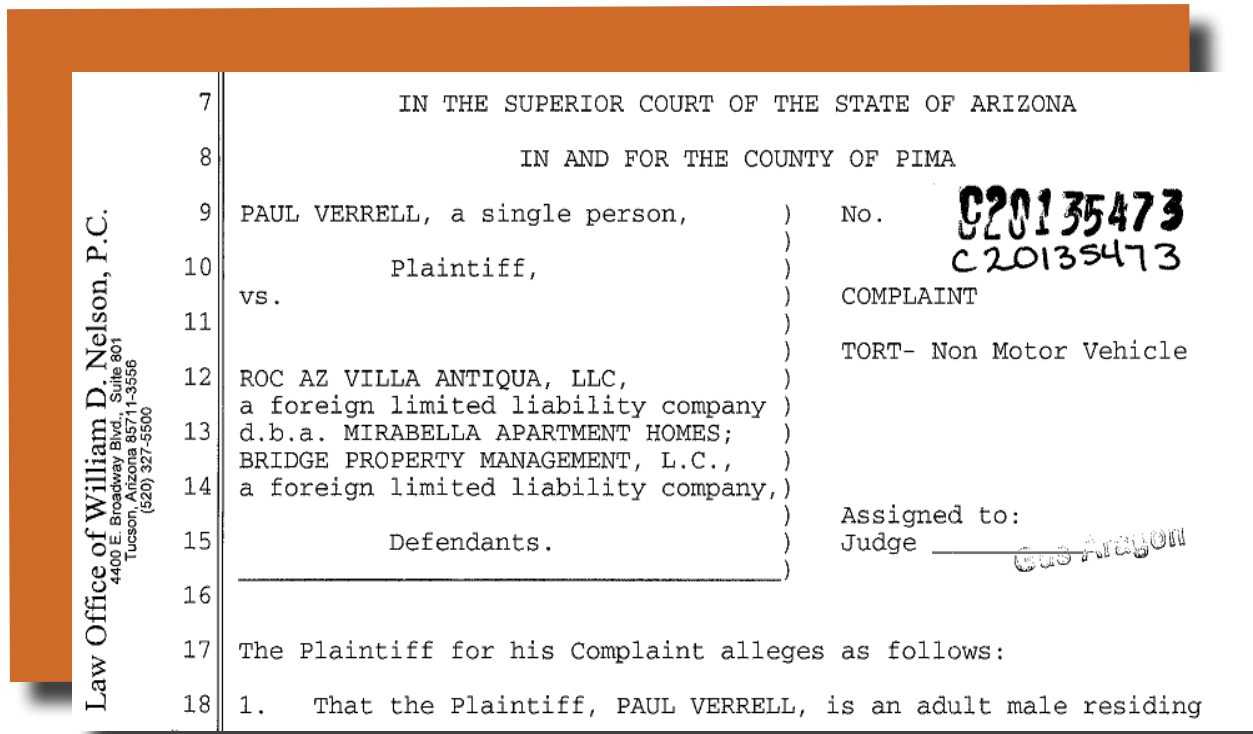
Figure 2: Example of Unknown Defendants Not Identified by NLP Technologies



The data element that posed the greatest difficulty for NLP was identification of the case type. NLP correctly identified the case type in only 39.2% of the cases. In those instances, it did so only because the case type was prominently included in the case heading with sufficient detail to be of use for case triage purposes. For example, “mortgage foreclosure” and “motor vehicle tort” were often identified correctly in case headings in all three participating courts. Other case types might be identified in the heading as “non-motor vehicle tort” or “breach of contract.”

These more general designations cannot differentiate a slip-and-fall premises liability case from a medical malpractice case or a credit card collection suit from a commercial contract dispute or partnership dissolution. As a general rule, medical malpractice, commercial contract disputes, and partnership dissolution cases are far more complex and require far more judicial involvement and oversight than premises liability or credit card collection cases.

Figure 3: Example of Incorrect Case Type Identification



Ultimately, none of the NLP vendors attempted the second or third components of the Triage POC, so the NCSC used their ability to correctly identify and extract information from the first component to assess the rate at which they could have done so. As Table 2 showed, NLP successfully identified and extracted 90% or more of most data elements other than case type. The relational data test required the NLP vendor to determine whether an answer was filed in response to the complaint, successfully count the number of plaintiffs in the complaint and defendants in the answer and determine whether all of the named defendants had responded to the complaint. It correctly determined that an answer was filed in 98.7% of the cases and correctly identified all plaintiffs and defendants in 83.9% of the cases. Consequently, it would have successfully

performed the relational data test for 87.6% of the cases in which an answer was filed.

Successfully completing the third POC component, however, was heavily dependent on correctly identifying the case type, the existence of an answer, the representation status of the parties, the number of plaintiffs and defendants, and in many instances, the relief sought including a jury demand by either or both parties. Although the success rate was acceptable for most of these items individually, NLP correctly identified all of the necessary information for triage in only 24 cases (10.7%). Incorrect case type was the most frequently occurring error.

Quality Control POC

The NLP Quality Control (QC) POC was an intentionally ambitious test of NLP ability to classify documents, extract information, and analyze and compare the extracted information to a checklist of case processing requirements for debt collection cases. See Appendix A for POC 3. The dataset consisted of 21,469 documents filed in 3,420 unique consumer debt collection cases disposed in the Cleveland Municipal Court. The Cleveland Municipal Court was specifically requested to participate in the POC because it had recently enacted Civil Practice Rule 6.13, requiring plaintiffs seeking default judgments to provide an affidavit of current military status, proof of assignment from the original creditor or original party in interest to the plaintiff, and the last billing statement from the original creditor sent to the defendant or an affidavit explaining why the required documents are not available. If Rule 6.13 is satisfied, the relevant documentation would include proof of the plaintiff's standing to bring suit, proof that the defendant received notice of the lawsuit, proof that the case was filed within the Ohio statute of limitations governing debt collection cases, and proof of the amount of damages sought.¹⁰ Documents related to 100 cases were selected for the QC POC Test Set while the remaining documents were made available to vendors as a Learning Set.¹¹

Like the Triage POC, data from the QC test cases were manually coded by project staff and entered into a dataset for analysis. In addition to documenting key information, the coders answered a series of relational questions related to standing, notice, timeliness, and proof of claims. Table 3 provides basic descriptive information about the QC Test Set cases. Of particular note, 59% of cases were filed by a plaintiff who purchased the debt from the original creditor, but only 88% of those cases included documentation showing the chain of custody for the debt. Sixteen percent (16%) of cases included proof that the defendant received notice of the claim and in an additional 80% of cases notice was presumed because nothing in the file indicated that the summons was not delivered. Three cases, however, had no summons documentation and in one case the summons was mailed to the *plaintiff's* address. In three cases, the name of the defendant did not match the debtor named in the contract on which the suit was predicated. Six cases did not indicate the date of default, which is necessary to determine whether the case was filed within the statute of limitations governing debt collection cases. Four cases failed to include proof of the amount claimed in the suit. Two cases indicated that the debtor had filed for bankruptcy, which should have stayed the proceeding in the municipal court. Each of these

10 The CCJ Civil Justice Improvements Committee identified proof of standing, notice, timeliness, and amount of damages as elements that are fundamental to procedural due process that had often not been observed in high-volume dockets. *Supra* note 3, at 33-34.

11 All cases selected for the NLP QC Test Set included at minimum the complaint, summons, proof of service return, and motion for default judgment with accompanying documentation.

inconsistencies should have triggered additional judicial scrutiny before a judgment was entered. The Quality Control POC was designed to identify those inconsistencies that might have been overlooked and bring them to the attention of a judicial officer.

Table 3: Description of QC Test Set Cases

Average number of documents	7.6
Average claim amount	\$2,938.70
Percent of cases served by certified mail	97%
Percent of cases with proof of service	16%
Percent of cases with presumed service	80%
Percent of cases with service date < 1 year	96%
Percent of contested cases	2%
Percent of cases with same defendant and debtor name	97%
Percent of cases filed by original creditor	41%
Percent of cases with proof of ownership by assigned plaintiff	88%
Percent of cases with default date included in documentation	94%
Percent of cases with proof of claims	96%

The electronic documents provided by the Cleveland Municipal Court included .pdf, .tif, and .xml formats and the image resolution for the documents varied from 200dpi to 400dpi. In addition, the case number was not always consistently marked on each filing. For example, case number 2018-CVF-06499 appeared variously as 18 CVF 6499, 18CVF 6499, and 2018 CVF 006499 in different documents. Finally, case filings often included duplicate copies of previous filings (e.g., affidavits included with both the complaint and the motion for judgment), which were subsequently scanned by court staff as part of the electronic file. Consequently, a significant challenge for the NLP vendors was correctly identifying the document type, associating the document with the correct case number, and then ignoring duplicate documents within the same electronic files.

The first task for the POC was to classify the type of document and count the number of unique documents associated with each case. Table 4 compares the number of unique documents identified by manual coding and the NLP process. It is clear from the analysis that poor image resolution and the duplication of documents within files greatly undermined the accuracy of the NLP document classification process. Variations in the format of the case number (truncation of year, extraneous leading zeros, and hyphenation or spaces between different sections of the case

Conclusions and Recommendations

A global movement towards digitalization is underway and the courts are included in this trend. With the public becoming more digitally savvy, there are greater expectations for courts to embrace digital technology and innovative approaches. Public interactions with the court system are a main driver of change as their demands for quality and speed of service are evolving both online and offline. New ways of working are also influencing the court's workforce. Technology provides opportunities for courts to work differently with new approaches to case processing, remote services, and public access to the courts.

The tools within Artificial Intelligence continue to grow and evolve. These proof of concept and use cases demonstrate that AI and NLP technology are capable of improving processes and delivering needed outcomes given the appropriate machine learning time

and attention to the quality of data. Courts that implement NLP technology usually start with areas that contain iterative tasks with low variability. Identifying iterative processes that are clear and easy are a common starting point, yet the benefits can be incredible. Reducing staff time by having technology deal with redundant tasks allows staff to shift attention to more complex tasks.

Data are at the core of successful digital transformation and one of the main benefits of AI technology is that data are no longer bound by traditional databases. Today data can be found in more diverse forms such as images, searchable text, handwriting, and even audio/spoken word. With the ever-increasing processing power in computing systems, large data storage capacities, and innovative tools, there are huge opportunities to harness the power of data.



Key Takeaways

Some key takeaways that should be considered before courts begin implementation of NLP and other innovative AI tools.

Data are Central to Innovation

As expected, the quality of the data greatly impacts future processes. If data is in a searchable format, such as a .PDF, it is easier for the software to fully understand the information. If the information is in a scanned document image such as .TIF or .JPG, then an Optical Character Recognition (OCR) process must be completed before the software can read and process the information within the document. The quality of the image resolution is critical for the OCR process to work effectively, so courts using scanned images should employ the minimum resolution standards necessary for effective OCR. Courts may need to improve existing document resolution if OCR minimum requirements are not met before starting the machine learning process.

Other markings such as time stamps over text and handwriting on forms may offer additional challenges in accuracy. Software recognition of handwriting and the ability to ignore markings such as stamps (noise) has improved and will continue to improve. However, it is still best to work towards the cleanest documents possible for scanned images. Ideally, information submitted into the court case file should be in a fully digital format. Most information today is created within a computer, so printing and scanning information back in as an image should be avoided. Processes should keep information

“born digital” to be retained in a fully digital format throughout the process. Digital time stamps, digital signatures and digital notarization process help make this possible. Ultimately, the courts should focus on collecting “information” contained in documents.

Data should follow standards to provide continuity to the software. Initiatives like the National Open Data Standards (NODS)¹³ are useful for providing courts with standard data definitions and structures. The more courts can agree on and use standards, the more easily software can learn. Standards make sharing and understanding information between disparate courts much easier. Standards at the local level such as standard form structures, standard data collection methods (portals, guided forms assembly), and well-designed cover sheets can help business analysts utilize software tools such as NLP to a greater potential as these efforts provide consistent learning. Having to learn multiple possible terms related to Dissolution of Marriage for example is possible, but the more “variety” that exists, the more learning must take place. Variability also impacts the continuous learning process and courts will have to maintain a growing catalog of learned terminology with various degrees of clarity as to what is occurring in the case.

13 See www.ncsc.org/NODS.

Rethink Processes

Fundamental to moving into digitization and using tools such as NLP requires courts to ask the fundamental question “Why are we doing this process this way?” and “How should we organize our work?” Courts should also consider how they can create an environment where they can be fit for the future and adaptable to changing needs. Implementing new innovative tools provide the perfect opportunity to look at the entire process and make changes that support current innovative improvements as well as setup future opportunities. It is a time to transform not just technology, but also human processes, policies, and experiences.

Document Intelligence

Machine learning allows software to read, understand, and identify key data elements. Then the software can be directed to take actions such as redaction, data extraction, assessment of data, and assignment into work queues or workflows. Whether scanned paper or a natively digital document, a lot of information is contained in the case record. Finding new ways to tap into that information is the goal of developing document intelligence strategies.

Traditional Databases

Courts still rely on traditional case management systems with data defined and stored within databases. Extracted data using document intelligence may be integrated or placed into databases more easily without relying on manual data entry.

Robotic Process Automation (RPA)

When direct data integration is not feasible or is complex, many courts are using RPA. RPA makes use of machine learning to identify and extract key elements off the digital court case file, and then replicate human data entry steps to populate a database. RPA is also used to randomly select case records for quality control tests as well as other simple iterative tasks that can be learned.

Data Warehouses

During early computing days when data was centrally stored on a mainframe, storage was limited and highly managed. Now with storage and processing capabilities becoming more robust, it is possible to collect data from various sources to create a combined data repository in a data warehouse. This reduces time to conduct analyses from multiple sources because much of the data has already been combined and placed into a storage space that is a single source of query. Data warehouses store current data from multiple databases as well as historical data for purposes of in-depth data analytics.

Advanced Digital Assistants – Chatbots

Courts are making use of NLP and machine learning to create advanced digital assistants and Chatbots. These assistants and bots help the public with information, guide them to resources such as standard court forms, provide language access, and connect them to the appropriate court staff for one-on-one assistance, if needed. These tools also help internal staff with data analytics, staff education, and assistance with internal resources such as human resources.

Business Intelligence

When courts put the effort into machine learning, this catalog of learned information may be applied to multiple levels of court case processing. When used at multiple points, the key benefit is the development of business intelligence (BI). Business intelligence leverages technology-driven processes that collect and store data. Then data analytics can be more rapidly and comprehensively completed to inform decisions and process improvements. Business intelligence provides greater capabilities for benchmarking, metrics, and analysis.

Use Cases

The use cases described below make use of NLP as well as other AI tools to perform functions similar and separate from the Proof of Concepts in the grant. They are great examples of the flexibility and variety of uses in the court environment. These use cases focus on improving internal processes as well as public facing processes and services to improve overall customer experience (CX).

ARIZONA MARICOPA COUNTY CLERK OF THE SUPERIOR COURT

The Clerk of Court for Maricopa County Superior Court is the record keeper and fiduciary for the Superior Court of Maricopa County, the fourth largest county in terms of population. The clerk handles records, documents, and money. Maricopa is an all-electronic court record court, but filings are submitted both electronically and in paper. Paper is digitized by scanning.

- An average of 36,291 pieces of paper are filed daily.
- The Clerk processes an average of 14,500 documents daily.
- More than 155,000 new cases are filed annually.
- The document image repository holds 78 million scanned images; paper filings are still scanned.
- The Clerk operates nine geographic locations with multiple filing counters.
- The Clerk processes an average of \$563,414 in monies daily.

The main driver for Maricopa's AI initiatives stemmed from the internal question of "how can we improve our traditional document processing?" In addition to filings, the Clerk's office also received approximately 30,000 calls per month with questions ranging from case information questions, e-filing support, payments, and licensing. The Clerk of Court wanted to do more with technology than configure off-the-shelf systems or develop applications in-house. Instead, the IT office sought to be "future ready" to take advantage of tools like Artificial Intelligence and Robotic Process Automation (RPA) and apply them to the environment. The Clerk strategized and prioritized leveraging emerging technology to transform service delivery and to improve customer experience. Bold, but calculated.

Strategies used involved:

- Artificial Intelligence:
- Robotic Process Automation (RPA)
- Business Intelligence – Data Warehouse

It was also important to invest in talent before taking the journey. Maricopa hired a Chief of Innovation and AI. It takes a team to configure, train, test and support the AI. Customer Experience Engineers were put into place and are similar to business analysts, but focus is more on AI conversations to monitor and improve the customer experience.

Operational Efficiency – Transformation with AI

Many courts still have document management systems especially in the early days of scanning paper case files. Even with e-filing, paper filings still occur. Document imaging or “intelligent capture” is done by scanning the document and putting it through an OCR process to convert the image into readable data. For documents that are scanned or received natively in a fully readable format, once received the focus then shifts to data within the digital documents. Data is automatically identified, classified, and data types and classification are trained to trigger placement into workflows. Previously this was a manual process, but now has been automated.

Intelligent capture was customized to fit the needs of the clerk. The Clerk required not only the document title, but also the case type and docket code. Once those elements are identified, the case then is routed to be auto docketed.

Once the intelligent capture process reached the high 90% accuracy confidence threshold, the Clerk moved to implement Robotic Process Automation (RPA). By enhancing their workforce with a digital workforce (RPA), the organization improved further with timeliness and efficiency. With this complement of AI tools and measures there has already been an over **50% improvement in the turnover of paper documents from processing filings into electronic court records and docketing, and a 40% efficiency improvement in staff time.** This process has allowed for 24/7/365 processing both attended and unattended.

EXAMPLE OF INTELLIGENT CAPTURE, REDACTION, CONFIDENCE THRESHOLD

**SUPERIOR COURT OF ARIZONA
IN MARICOPA COUNTY**
Court Administration

**** FILED ****
7/27/2022
Clerk of Court

Case Number: PB 2019-070241
Date: 7/27/2022
**In the Matter of the Guardianship of and
 Conservatorship for:**
 [REDACTED] [REDACTED] #022281
 A Protected Adult. [REDACTED] #031478

COURT ACCOUNTANT'S REPORT

Initial Review - Second Annual Account

The Court Accountant has reviewed the following:

- The Second Annual Account for the period 7/1/2020 through 6/30/2021.

The Third Annual Account for the period 7/1/2021 through 6/30/2022 should be filed on or before 9/30/2022.

- A hearing will be set to consider the approval of this account, and a **Notice of Hearing** for the date and time that this matter is scheduled will be sent to the Petitioner on this day.
- Based on the information submitted, the Account reconciles.
- Based on the amount of unrestricted assets and estimated annual income reported in this account, the fiduciary's bond of \$100,000.00 could be decreased by \$50,000.00 to \$50,000.00.

If the petitioner desires to comment on the bond recommendation, a response must be filed, with a copy lodged with the above-named judicial officer's division, at least seven (7) calendar days in advance of the hearing date.

- It is noted that the Conservator has not filed the financial statements with Court. Instead, the

PB 2019-070241

PB 2019-070241

CaseType

PB

PB

DocumentTitle

COURT ACCOUNTANT'S REPORT

COURT ACCOUNTANT'S RE

DocketCode

IN MARICOPA COUNTY

435

FilingDate

7/27/2022

07/27/2022

Generic Sort Code

01W

	<i>Case Number:</i> CV2012-004939
--	--------------------------------------

- CaseNumber (1 item)		
Field	Number	Confidence
CV2012-004939	Case	98.60%

RPA – How RPA Robots were used in Maricopa

Robotic process automation (RPA) is a business process automation technology based on metaphorical software robots (bots) or an artificial intelligence (AI) digital worker. This involves developing an action list by having the bot watch a human perform the task within a software interface and then learning to perform the automation through repeated observations. This is an alternative to using an Application Programming Interface (API) to exchange information. A common use for RPA is to train it to identify data from case documents and perform data entry functions through an automated process. This use case for RPA helps with gaps in the workforce in areas where staff may be performing iterative tasks that can be learned and replicated by software.

In Maricopa County, each of the bots was given a name, including “Ron Burgundy,” “World News Agent,” “Yoda,” “Alfred,” and “CLEO”. Each bot uses NLP to identify information and is given instructions on steps to perform via a learning/training process. RPA mimics human steps such as data entry or launching a search query on the Internet so these steps may be automated.

Ron Burgundy is an Internal Testing BOT that searches websites for new information about courts and technology and presents it back to the internal team. World News Agent assists employees to find information on external websites.

Yoda is an Internal Slack BOT that assists employees to find information about administrative and resources, such as signing up for benefits. (Assist Employees)

Alfred is an Internal Slack BOT that assists the technology division with monitoring and with managing technology requests. Alfred has some help desk assistance functions, including classifying the assistance request and automatically creating and assigning the help desk ticket.

CLEO (English) and CLEO (Spanish) is a customer-facing BOT Virtual Assistant that focuses on the customer experience. IBM Watson is used for voice conversations and Twilio to connect to Omnichannel. Using NLP, CLEO appears as a chat bot on the Clerk’s website and allows customers to engage 24/7 in both English and Spanish. CLEO averages 3,700 chats per month in includes the ability to seamlessly manage a warm hand off to a human conversation with a customer experience (CX) representative. Watson is used as a knowledge base for human conversations to help ensure information is consistent and evolves as it is exposed to new information. Thus far, 80% customers rate their experience as satisfactory 80% of the time. Maricopa will be moving from Chatbots to conversational AI as the next iteration in their transformation. Maricopa County Superior Court is working with the vendor Computing Systems Innovation (CSISoft) to implement AI, machine learning, data extraction, and RPA.

ORANGE COUNTY SUPERIOR COURT OF CALIFORNIA

Project Theme: Data is our Killer App. Orange County viewed this opportunity with the slogan “Data is our killer app”. To understand the existing process to transform the area of document intelligence, areas of workload, capacity, backlog, jury response rate, and fiscal impact of policies were reviewed in depth.

Orange County Superior Court of California was challenged with a high volume of unique forms entering the court. There is an investment of time to review these forms which is a highly procedural process. Information contained within the forms trigger placement into workflows. This process was using an incredible amount of human processing time and staffing was not sufficient to keep up. Many of the forms are paper files scanned and digitized as an image .PDF rather than having a native fully digital searchable .PDF. Faced with this challenge, Orange County looked at opportunities to transform and digitize the process.

Even in e-filing scenarios there was a high rejection rate. The Family Division had a 20% rejection rate of e-filed forms, and 40% of the time the reason was incomplete information. Each form is manually reviewed by a clerk regardless of entry method, scanned paper or e-filing. This takes a lot of time.

Transforming this process was accomplished by starting small and branching out. AI tools are now mature and “big” because there are many components to AI that work in various combinations to address specific processes. Technology using AI on forms was logical as forms have structure which makes it easier to train AI on repeatable steps since data is located at defined locations on the form. Machine learning is a process where AI is trained to locate data, identify it, and then process the data as per instructions. As the number of forms increases that AI processes and learns from, the more accurate it becomes over time. The civil division of court was selected first since there was mandatory e-filing using standard forms in place.

Three use cases are in play in Orange County.

1. Document Intelligence and Data Extraction.
2. Redaction – due to legalization of cannabis, many court records required redaction of past offenses.
3. Default Judgements

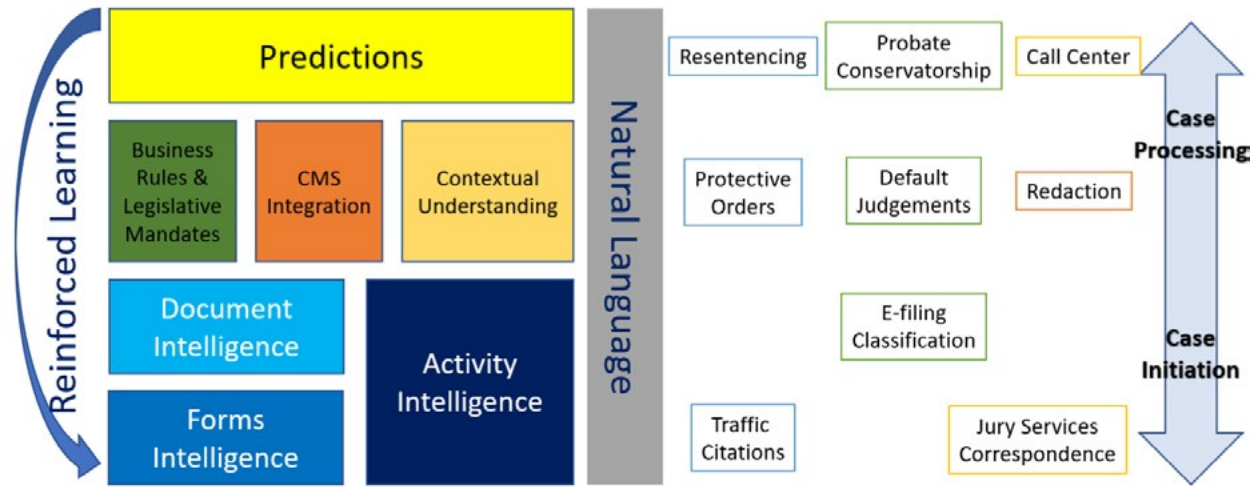
USE CASE 1 – Document Intelligence and Data Extraction

Document Intelligence is about unlocking the data within the case file or forms. The courts have lots of documents and untapped information that could be available for query and other actionable processes and automation scenarios. Document intelligence complements business intelligence by supplementing data extracted from documents with data from databases and data warehouses. Document classification is the first step in the process and in Orange County this is the Magic Classifier process. Document classification is a manual process to drill down from the high level to the sub classification levels needed to properly docket and place the case into a workflow queue. There is a lot of work being done now using data analytics to determine the key indicators for classification and then using the iterative machine learning process to train AI to perform the classification process.

There are 3 case management systems in Orange County: 1) Tyler Odyssey for Family and Juvenile (SQL); 2) V3 for Civil, Probate, Small Claims (Oracle); and 3) Vision for Criminal (Oracle). There was already in place an established method of unlocking the data from these sources and putting them into a data warehouse (Snowflake). There was also an established method to visualize the data using Power BI, Tableau, SharePoint Online, and MS Excel. The layer that was added was the AI and Machine Learning layer. It was placed after the data warehouse, so the presentation tools had more information available Orange County is using these tools in the AI and machine learning swim lane: Databricks (data analytics), Azure DevOps, and Azure Forms Recognizer (Azure DevOps and Forms Recognizer are completing the data extraction and forward actions).

The building blocks below take information from the AI and Machine learning through the document intelligence process and adds to the business intelligence. The activity intelligence integrations, contextual understating, business rules along are combined with Natural Language Processing (NLP) to support processes to the right. These processes are simple such as case initiation, document classification of e-filed case information to more complex processes supporting redaction, default judgements, protection orders to name a few. These building blocks and automation help the clerk and courts with case processing. Predictive Analytics are used for such things as case filing levels and workload predictions.

BUILDING BLOCKS



DATA ROADMAP

Legend:

Black: Completed

Blue: In progress



Data 1.0 Data Warehouse

- i. CAVE (CMS data) – CIG grant
- ii. Expand CAVE to include non-CMS data (Jury or Self Help)
- iii. Finance / HR data mart



Data 2.0 Data Exchanges & Smartbots

- i. Grand central data exchange
- ii. Smartbot – Dialogue flow: jury & collections
- iii. JMS – Jail data exchange
- iv. CUP – proof of corrections
- v. Self Help – editable court forms
- vi. CAP – attorney exchange
- vii. Online Records access



Data 3.0 Predictions

- Build predictions engine using CAVE input data:
- 1) Predict filings
 - 2) Predict # of trials
 - 3) Predict # of hearings
 - 4) Predict # of court reporters
 - 5) Predict number of jurors to summons



Data 4.0 Doc AI/ML (Document meta data)

- i. POC – evaluate viability of extracting meta data from court filing documents for case processing
- ii. Document Classification
- iii. Marijuana Redaction
- iv. General purpose fuzzy match (demurer ‘met and conferred’)

USE CASE 3 – Default Judgments

In Orange County Superior Court, all default judgments are filed electronically. The courts received meta data and PDFs. As these filings go into a review queue for default judgments, the clerks would have to view each one and determine the correct subtype. There are 9 subtypes for default judgments. Making the subtype determination may require the clerk to find information from other sources such as a lookup in the case management system. Once the subtype was identified, it was added to the notes section in the CMS. Then the clerk assigned to the work the specific subtype for default judgments would have to search the notes to “find” these cases assigned to them. This was a time consuming and inefficient process.

To transform this into a more efficient digital process, the AI will scrape the pertinent data from the default judgment filing, rules will be applied to the data, there will be 9 specific sub-queues and the rules engine will 1) determine the appropriate subtype and 2) place the filing into the correct queue. Automating this part will free up clerk time from the heavily manual process of determining subtype and allow them to work on the queues. No jobs are lost in this process, but the repeatable steps have been automated to allow the clerks to work more timely on cases. This will help reduce backlogs.

SAMPLE DEFAULT JUDGMENT

Case information must match complaint

“Entry of Default” marked

Date of complaint/cross-complaint Name of plaintiff/cross-complainant

Name of defendant

Date and printed name

Signature

Clerk enters or denies default; signs, dates, and files form

FOR COURT USE ONLY

Adapted from a product of the Superior Court of Napa County
Center for Judicial Education and Research (CJER) • 2017

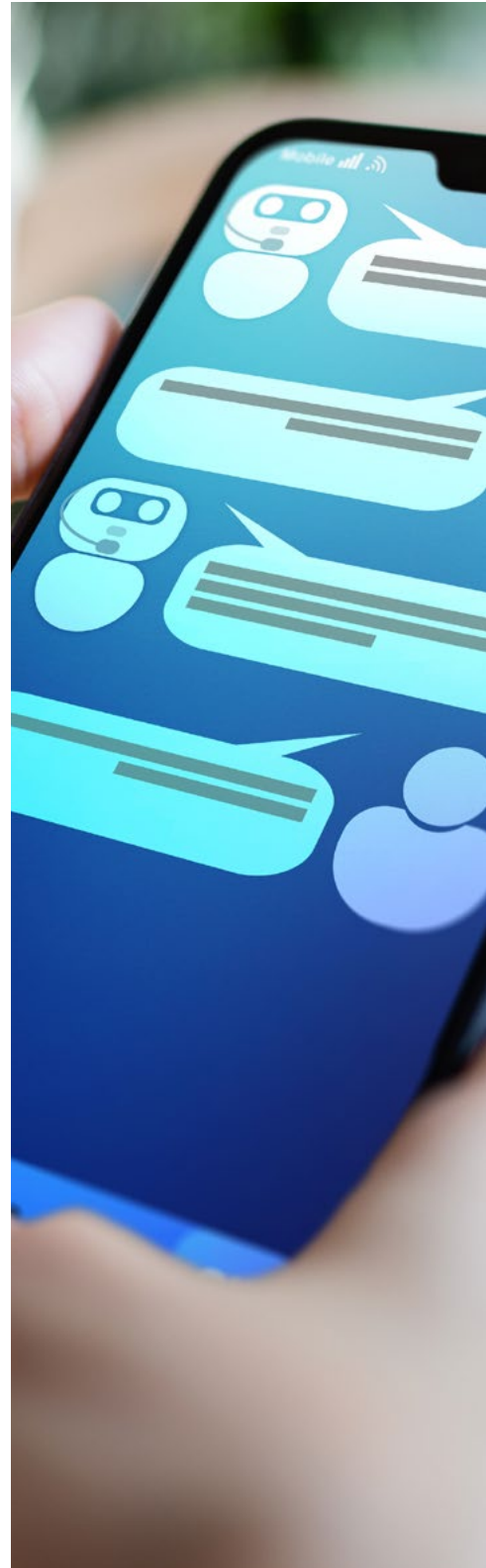
Lessons Learned

1. Start with a relevant business question.
(What problem needs to be solved?)
2. Leverage an integrated technology stack. (Buy and build can be combined, look at what works best for the court's environment).
3. Be agile. (start small, iterate, learn, repeat)

Other Uses of AI

Other uses are AI in Orange County includes Chatbots using Google Contact Center AI in areas of the Collections Group and of Jury Group since those are high volume areas where the court receives a lot of questions. The BOT is used to answer the common questions coming in. Collections has a team of 2 people working part time to work on the Q/A to refine parameters around "intent", or "What are you trying to find?". Business analysts look at the questions coming in and help refine the ChatBot's ability to answer incoming questions. Special emphasis on new questions. This is known as intent mapping. Orange County is evolving from Chatbots to conversational AI as their next step in their digital transformation.

Orange County is using other tools than RPA, but sees the benefits of this technology. The term robotic may be misunderstood and make employees concerned about being replaced by a robot. Perhaps the , the "R" should be viewed as "Repeatable" since this technology is a great fit for repeatable tasks that the software can learn by mimicking the pattern through repeated observations of the steps. RPA is an excellent fit for older systems where direct integration through an API may be difficult or unavailable.



Appendix A:

POC 1—Civil Case Data Extraction and Case Matching POC

Background:

The National Center for State Courts has already completed proof of concepts on data redaction and would like to look at the technology to complete data extraction from civil cases. Data extraction would include initial document classification and capture of data.

POC Purpose:

The purpose of this POC is to determine the effectiveness and accuracy of extracting specific targets from civil documents. These extracted data will be critical for use in population of other application's databases. It is anticipated that the software will be more effective in finding and extracting data from the document that will lead to more complete and accurate data sets. To demonstrate some potential use in an outside application component, extracted data will have some relational comparisons.

Data Set:

The Civil Case Triage dataset consists of approximately 65,000 pleading documents (Complaints ≈ 37,000; Answers ≈ 28,000) from the Maricopa County (AZ) Superior Court, the Pima County (AZ) Superior Court, and the Palm Beach County (FL) Circuit Court.

Data Extraction:

For each document, extract the following information:

- Extract the name of the court in which the document was filed;
- Extract the case number assigned to the document;
- Identify the type of document (e.g., complaint, answer)
- Extract the date the document was filed;
- Is this document written in a language other than English? Y/N
- Is this document written in plain English? Y/N
- Indicate the number of pages in the document.

If the document is a Complaint

- Extract the bar number of plaintiff's lawyer; and the name of the law firm; OR
- Indicate that the plaintiff is self-represented.

- How many plaintiffs are named in the Complaint?
- Extract the name of each plaintiff and indicate whether the plaintiff is a person or an organizational party.
- How many defendants are named in the Complaint?
- Extract the name of each defendant and indicate whether the defendant is a person or an organizational party.
- Indicate if the plaintiff(s) seeks class action certification? Y/N

Indicate the subject matter of the lawsuit:

- Automobile negligence (Pima, 3,425; Maricopa, 8,177; Palm Beach, 3,425.
- Premises liability (Pima, Maricopa, 754; Palm Beach, 1,042;
- Medical malpractice (Maricopa, 440)
- Legal malpractice (Maricopa, 180)
- Other professional malpractice (Maricopa, 52);
- Product liability (Maricopa, 6)
- Slander/Libel/Defamation (Maricopa, 172)
- Intentional tort – Assault/Battery
- Intentional tort – Vandalism
- Pet attack
- Breach of contract – plaintiff buyer (Maricopa, 35)
- Breach of contract – credit card debt collection (
- Breach of contract – student loan debt
- Breach of contract – other consumer debt collection
- Breach of contract – commercial debt collection
- Landlord/tenant – residential eviction
- Landlord/tenant – past due rent collection
- Landlord/tenant – tenant plaintiff (housing violation, deposit collection)
- Landlord/tenant – commercial lease

Outcomes:

Extraction Test

- Capture data in a structured dataset;
- Capture document content for future search capability;
- Generate summary of extracted data.

Relational Data Test

- Match cases based on identical court and case number.
- Compare number of parties in Complaint(s) and Answer(s).
- Identify difference in the number of parties, names, or litigant types.

Appendix B: POC 2 – Civil Case Triage POC

Background:

The National Center for State Courts captured a diverse data set of civil cases and their outcomes to develop a case triage model. This model placed cases into one of three categories: 1) simple, 2) standard, and 3) complex. This model was based on experience from subject matter experts.

POC Purpose:

The purpose of this POC is to determine the effectiveness and viability of using AI tools to place triage civil cases into the three categories. These categories assist clerks/courts with workflow. The vendor may approach this POC to use the existing model for triage or to use AI tools to conduct analytics to determine a more effective model.

Outcomes:

Depending on the approach of the vendor for this POC the anticipated outcomes may fit into one of two categories:

1. Use AI tools within the software to triage cases based on the NCSC model. Compare POC results to actual results outcomes in the model.
2. Use AI tools to review and analyze the same civil case types and determine the appropriate case management pathway using a new model based on predictive analytics. Compare POC results to actual results outcomes in the model.

Dataset:

The Civil Case Triage dataset consists of approximately 65,000 pleading documents (Complaints ≈ 37,000; Answers ≈ 28,000) from the Maricopa County (AZ) Superior Court, the Pima County (AZ) Superior Court, and the Palm Beach County (FL) Circuit Court.

NCSC will provide complexity scores and raw data for each case based on actual case activity reported in CMS and will provide complexity thresholds for pathway assignments in each court.

Appendix C:

POC 3 –Civil Consumer Debt Cases, Quality Control POC

Background:

The National Center for State Courts would like to explore the use of AI tools to assist with quality control in civil cases, specifically the consumer debt collection case type. There is a need to check completeness of information and other critical indicators to determine if a case is ready to move forward or requires additional case management.

POC Purpose:

There are a host of requirements to process civil cases in debt collection. This POC will utilize document classification and data extraction tools to match documents in cases and extract various required elements. Then these information points will be further analyzed and compared to a quality control requirements checklist.

Dataset:

The Quality Control dataset consists of 21,469 documents filed in 3,420 unique consumer debt collection cases disposed in the Cleveland Municipal Court. **The image resolution varies from 200dpi to 400dpi. This particular jurisdiction will recopy the entire court file upon each filing, and you will find duplicate documents within the image. Software will need to be able to identify and ignore duplicates.

For each document:

- Identify the document type; ** In the data set, there are duplicate copies in subsequent filing, so document identification will be important to this POC.
- Extract the case number.

If the document type is a Complaint, extract:

- Case number
- Filing date
- Name of Plaintiff
- Number of Defendants
- Name of each Defendant(s)
- Address of each Defendant(s)

- Amount of debt claimed
- Date of default
- Amount of principle claimed
- Amount of interest claimed
- Amount of fees claimed
- Attorney signature Y/N

If the document type is a Return of Service document, extract:

- Case number
- Service date
- Filing date of return
- Who served the notice? (USPS, Sheriff, private process server)
 - Name of private process server
 - Image of signature on USPS return Y/N
 - Failure of service (undeliverable, unclaimed, refused, not served)
- Name of Defendant
- Address of Defendant on summons
- Address of Defendant where served
- Type of service (personal, residency, publication, certified mail, first class mail)

If the document type is an Answer, extract:

- Case number
- Filing date
- Number of defendants
- Name of defendant(s)
- Address of defendant(s)
- Bar number of lawyers, if any
- Is the debt admitted or contested?
- Indicate defenses alleged in Answer:
 - Debt Satisfied
 - Debt discharged/bankruptcy
 - Not me
 - Not my debt
 - Amount in dispute
 - Statute of limitations
 - Debt invalid
 - Identity theft
- Attorney/Party Signature Y/N

If the document includes Supporting Documentation:

- Indicate in which document type the supporting documentation was appended;
- Indicate the page number in document where the supporting documentation was appended
- Indicate whether the supporting documentation is a billing statement or statement of debt owed.

If so, extract:

- Case number
 - Filing date
 - Name of Plaintiff
 - Name of Defendant
 - Date of original contract/application
 - Date of statement
 - Date of last payment
 - Date of default
 - Amount of principle
 - Amount of fees
 - Amount of interest
 - Signature on Affidavit N/A
 - Affidavits (attorney or other source)
- Indicate whether the supporting document is an affidavit.

If so:

- Indicate the page number in the document where the affidavit was appended
- Indicate if the Plaintiff is the original creditor Y/N
 - If the plaintiff is not the original creditor, indicate whether a statement describing the chain of ownership/custody is included.
- Extract:
 - Case number
 - Filing date
 - Attorney or creditor affidavit
 - Signature on Affidavit

If the document type is a Motion for Judgment, extract:

- Case number
- Filing date
- Plaintiff name
- Number of defendants
- Defendant name(s)
- Defendant address(es)
- Amount claimed

- Statement describing proof of standing (original creditor or chain of ownership/custody)
- Military affidavit
- Amount of attorneys' fees
- Supporting documentation
- Attorney Signature

Outcomes:

Extraction Test

- Capture data in a structured dataset;
- Capture document content for future search capability.

Relational Data Test

The output will be a checklist that will summarize key indicators in a case to assist the court in determining the quality of the case, identifying issues requiring additional action, and determining readiness of the case to move forward.

1. Show chain of ownership of the debt if the debt has been sold.
2. Show evidence of debt (contract, billing statement, other documentation)
3. Motion for default judgment – must show supporting documentation and financial accounts
4. Military service check has been conducted. (military receive special exemptions/ accommodations).

Appendix D: Civil Case Triage Criteria

CIVIL TRIAGE CRITERIA FOR MARICOPA COUNTY SUPERIOR COURT		
Case Type	Assign to General Pathway if all conditions are met	Assign to Complex Pathway if all conditions are met
Debt Collection	Not applicable	Plaintiff and defendant are represented, 2 or more defendants, answer or responsive pleading filed, and jury demand filed by either party
Landlord/Tenant	All cases	Not applicable
Other Contract	Plaintiff represented, 2+ defendants, answer or responsive pleading filed	Plaintiff represented, 2+ defendants AND 2+ plaintiffs, and answer or responsive pleading filed
Automobile Tort	Plaintiff and defendant represented, 2+ defendants AND 2+ plaintiffs, answer or responsive pleading filed, and jury demand filed by either party	Not applicable
Intentional Tort	Plaintiff and defendant represented, 2+ defendants	Plaintiff and defendant represented, 2+ defendants, answer or responsive pleading filed
Medical malpractice	Not applicable	All cases
Other malpractice	Not applicable	Plaintiff and defendant represented, 2+ defendants, answer or responsive pleading filed
Product liability	Plaintiff and defendant represented, 2+ defendants, answer or responsive pleading filed	Plaintiff and defendant represented, 2+ defendants AND 2+ plaintiffs, answer or responsive pleading filed
Premises liability	Plaintiff and defendant represented, 2+ defendants, answer or responsive pleading filed	Not applicable
Other tort	Plaintiff and defendant represented, 2+ plaintiffs, answer or responsive pleading filed	Not applicable
Real property	Plaintiff represented, 2+ defendants, answer or responsive pleading filed	Not applicable
Other civil	Plaintiff and defendant represented, 2+ defendants, answer or responsive pleading filed	Not applicable

Appendix D (con't): Civil Case Triage Criteria

CIVIL TRIAGE CRITERIA FOR FIFTEENTH JUDICIAL CIRCUIT COURT OF FLORIDA		
Case Type	Assign to General Pathway if all conditions are met	Assign to Complex Pathway if all conditions are met
Debt Collection	Plaintiff and defendant are represented, more than 2 defendants, answer or responsive pleading filed	Plaintiff and defendant are represented, counterclaim or third party claim filed, answer or responsive pleading filed, and jury demand filed by either party
Landlord/Tenant	Not applicable	Not applicable
Other Contract	Not applicable	Not applicable
Automobile Tort	Plaintiff and defendant are represented, more than 2 defendants, answer or responsive pleading filed	Not applicable
Intentional Tort	Not applicable	Not applicable
Medical malpractice	Not applicable	Plaintiff and defendant are represented, more than 2 defendants and 2 or more plaintiffs, answer or responsive pleading filed, and jury demand filed by either party
Other malpractice	Not applicable	Not applicable
Product liability	Not applicable	Plaintiff and defendant represented, more than 3 defendants, answer or responsive pleading filed, and jury demand filed by either party
Premises liability	Not applicable	Not applicable
Other tort	Not applicable	Plaintiff and defendant represented, more than 2 defendants, answer or responsive pleading filed, and jury demand filed by either party
Real property	Not applicable	Not applicable
Other civil	Plaintiff and defendant represented, 2 or more defendants, answer or responsive pleading filed, and jury demand filed by either party	Not applicable

Appendix D (con't): Civil Case Triage Criteria

CIVIL TRIAGE CRITERIA FOR PIMA COUNTY SUPERIOR COURT		
Case Type	Assign to General Pathway if all conditions are met	Assign to Complex Pathway if all conditions are met
Debt Collection	Not applicable	Not applicable
Landlord/Tenant	Plaintiff and defendant represented	Not applicable
Other Contract	Not applicable	Plaintiff and defendant represented, answer or responsive pleading filed, and jury demand filed by either party
Automobile Tort	Not applicable	Plaintiff and defendant represented, answer or responsive pleading filed, and jury demand filed by either party
Intentional Tort	Not applicable	Not applicable
Medical malpractice	Not applicable	Plaintiff and defendant represented, 3 or more defendants, and answer or responsive pleading filed
Other malpractice	Not applicable	Not applicable
Product liability	Not applicable	Not applicable
Premises liability	Not applicable	Not applicable
Other tort	Not applicable	Plaintiff and defendant represented, answer or responsive pleading filed, and jury demand filed by either party
Real property	Plaintiff and defendant represented, organizational defendant, 3 or more defendants, answer or responsive pleading filed	Not applicable
Other civil	Plaintiff and defendant represented, answer or responsive pleading filed, jury demand filed by either party	Plaintiff and defendant represented, no organizational parties



ncsc.org/cji

ISBN: 978-0-89656-328-5

© 2023

0 1 1 0 1