TLLF Systems Redesign
ESE Senior Design
Spring Final Report
29th April 2021

Authored by Team 2 (MAJEC):
Ellie Cohen: SSE, ecohen19@seas.upenn.edu
Carsen Coggeshall: SSE, carsenc@seas.upenn.edu
Archit Dhar: SSE, architd@seas.upenn.edu
Julie Hoover: SSE, jhoove@seas.upenn.edu
Maher Abdel Samad: M&T: SSE, msamad@seas.upenn.edu

Advisor: Professor Hamed Hassani: ESE, hassani@seas.upenn.edu

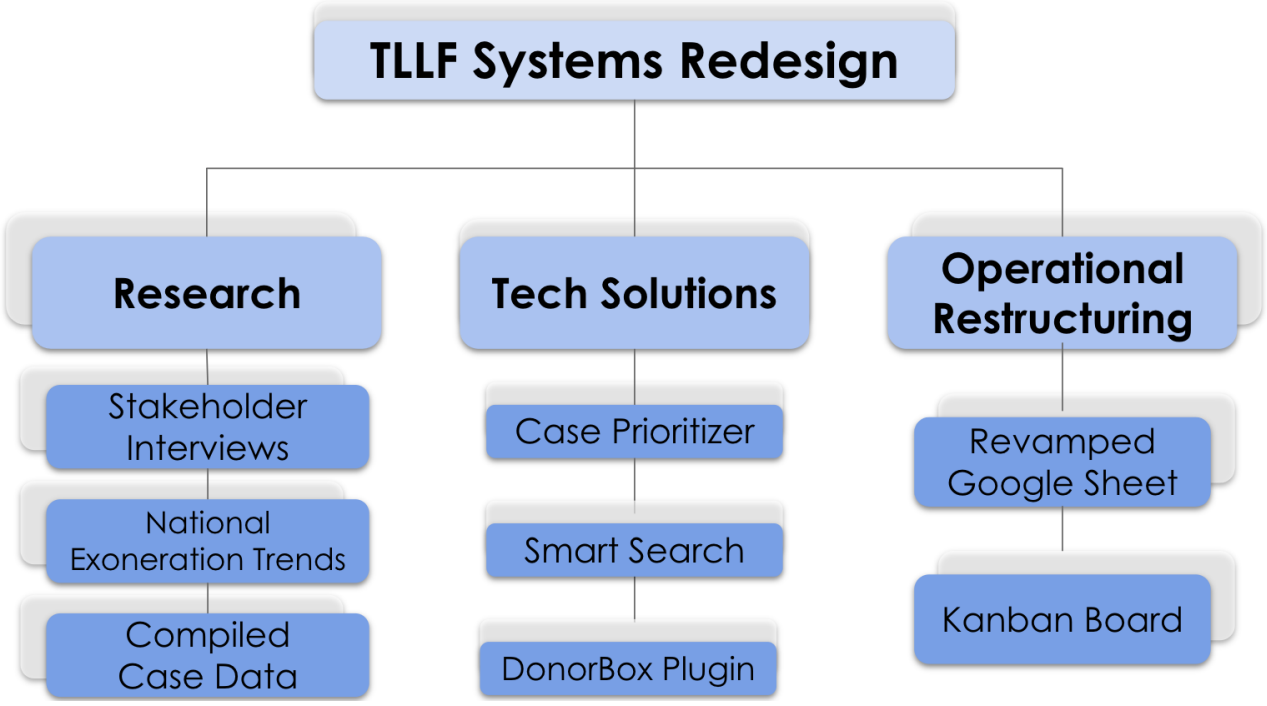


Table of Contents

II. Executive Summary

Our senior design project with the Terrance Lewis Liberation Foundation was focused on helping the organization optimize its workflows and operations while also preparing it to sustain this efficiency as they continue to scale. We helped create tools and procedures that facilitated the work of volunteers and members of the legal team. We chose a three-pronged approach. First was Research: researching and identifying trends in the cases, stakeholder interviews, and compiling case data. Next was Tech Solutions: Case Prioritizer, Smart Search, and Donorbox plugin. Lastly was Operational Restructuring: revamped Google sheet and Kanban Board. We were able to implement these solutions and improve the efficiency of TLLF and supply them with the tools to scale.

We know our project cannot solve institutional problems in criminal justice, but it can address microscale issues that organizations seeking to free wrongfully convicted prisoners have to deal with. These processes being improved makes a massive difference to these organizations.

Through the completion of our project, we found that the legal space is antiquated and hesitant to adapt technical solutions. As we gained the trust of our stakeholders, we were able to equip them with technical tools that make their processes more efficient while still ensuring that the incarcerated people are treated humanely and justly. Our solutions were customized to address the pain points of TLLF, and some components of our project have been adopted by law firms that work with the foundation.

III. Overview and Motivation of the Project

Our story begins with Terrance, a man sentenced to life in prison in 1996 for a crime he did not commit based only on one very shaky witness' testimony. He was released in 2017, 21 years later after being robbed of his freedom and two decades of his life at seventeen. In 2009, a federal judge found that Terrance was not guilty. However, since Terrance was part of the state prison system, he was not freed for another eight years. When he was finally freed, he founded the Terrance Lewis Liberation Foundation, a non-profit dedicated to securing justice for those wrongfully incarcerated individuals whose lack of resources and legal representation prevent them from regaining their freedom from an imperfect system

What motivated us to take on this project and drove our passion throughout the year was his story and getting to speak with him directly. We took the time to interview stakeholders and learn about the organization's problems that we could integrate tech solutions and systems engineering to solve. We strived to scale their ability to take on cases, process documents and do the tedious work required to petition the conviction integrity unit for people's freedom.

Our project specifically focused on creating tools and procedures that will facilitate the stakeholders' work, particularly for the legal team and members of Project RIC (Research, Investigate, Compile- a subgroup within TLLF). This team is focused on improving the intake, review, and management processes within TLLF. We worked with the organization to improve its questionnaire system so they can make a judgment on a case's potential more efficiently. Next, we aimed to reorganize the reviewal of information that allows the RIC members to understand all the important details of the cases. Finally, we aimed to improve TLLF's management of their clients and stakeholder's work. We know our project cannot solve institutional problems in

criminal justice, but it can address microscale issues that organizations seeking to free wrongfully convicted prisoners have to deal with. These processes being improved makes a massive difference to these organizations.

This is a nationwide problem with nationwide trends. It's estimated that there are up to 120,000 wrongfully incarcerated individuals in the US, but the amount of people who clear the ridiculously rigorous exoneration process is an insignificant fraction of that number. The key factors contributing to this are limited access to legal resources, volunteers and lawyers as well as dependence on outdated, manual systems with no data whatsoever.

IV. Technical Description

Our technical contributions for TLLF were the case prioritizer, smart search, Donorbox plugin, and Trello board/Google sheet management systems. Each of these addressed specific problem areas that we identified during our initial research phase of the Fall semester: criterion for accepting/rejecting cases, case file review and comprehension, and stakeholder relationship management. After many iterations of development, they all work as easily operable, standalone tools that significantly improve workflows when used in conjunction. The following section will outline the technical aspects of each of these solutions.

Case Prioritizer

To capture the challenges faced in determining what new cases to take on, we created the following problem statement: A volunteer who reviews Pro Se Requests needs a way quickly determine if TLLF can accept a request with only preliminary information because it is important the individual is assisted by the right organization, not necessarily TLLF. The phrasing is very particular to clearly identify the requirements. The solution must be quickly and easily applied to a given case; it must work with the limited information provided in the Pro Se Request; and it must be transparent so as to provide understandable guidance for the best route for action (oftentimes that means TLLF will direct the case to another organization). Most of the volunteers have a non-technical background, so the user interface must be very intuitive. We also ensured the system would not take too much additional time to incorporate into the reviewal process. The change in process had to be simple yet effective to have long term adoption. Lastly, TLLF is a young organization with a limited budget, so we aimed to use free tools and platforms.

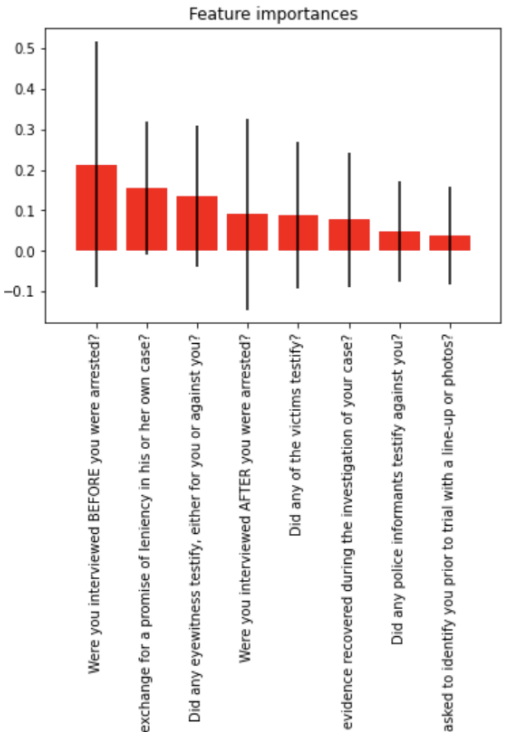
In the first iteration of the case prioritizer, we redesigned the Pro Se Questionnaire to only include binary questions. From an algorithmic perspective, the binary design would allow for easy featurization, particularly given that some questions had discrete accept/reject consequences. However, after talking to the volunteers and lawyers, we learned that many of the incarcerated people put key details in their open ended questions that lead to new trials. Omitting these would be a critical flaw. Rather than trying to reinvent their existing form, we thought about how we could design our algorithm around what they had. This had the added benefit of making the Foundation's dozens of existing responses usable as development data.

We reviewed the Pro Se Questionnaire and identified which questions should be included as features. We selected all the Yes/No questions as well as the questions that had common answers (i.e. What crimes were you convicted of? and What new evidence has come to surface?). To mitigate algorithmic bias, we opted to exclude demographic questions such as

race and highest level of education. While these could potentially be useful indicators, we felt that there could be underlying factors that would skew the trends it finds. The remaining questions focus on facts about the case, which are much more insightful regarding a case's validity.

In order to make the data accessible for our algorithm, we created a Google Form that served as a digital copy of the Pro Se Questionnaire. As a Project RIC volunteer reviewed the document, they would copy the responses into the form and send the data to a Google Sheet. We asked about the feasibility of sending the Google Form directly to incarcerated individuals rather than mailing the paper document, but prison rules and access to computers made that prohibitive. We found that copying the responses over was not a significant addition of work, and it greatly improved the availability of case data for the organization.

After sorting out the data, we were ready to apply a model. We discussed with Professor Hassani what the best approach would be, and we decided on a Random Forest classifier that would predict the binary decision for TLLF to accept or reject a case. We needed a transparent model that we could "look into" to understand the decision making factors. The final output was not necessarily the target of interest, but rather the features that drove them. A trained random forest classifier could calculate the importance of each feature, which would be used to highlight the crucial details of any given case.

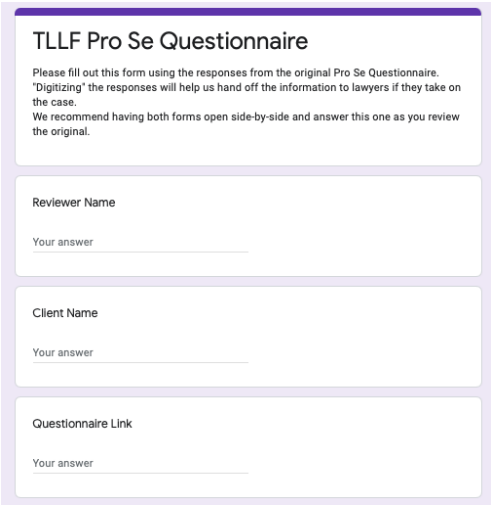


Random Forest Feature Importances

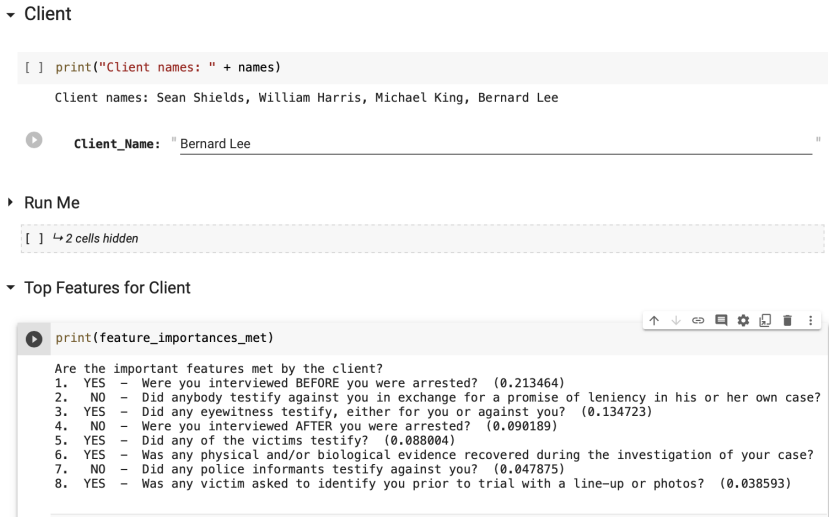
The final version of the case prioritizer consisted of scikit learn's random forest classifier trained on 40 Pro Se Questionnaire responses sent to TLLF. We used cross fold validation to calculate the optimal tree depth of three. One challenge we faced was the lack of "negative" cases- responses sent to TLLF that would be rejected. The classifier needed samples from both categories, so we augmented the dataset with negative examples in which every response was "No" or a similar equivalent. We considered up-sampling the negative examples or down-sampling the positive ones, but that would create problems due to the limited size of the dataset and fabricated nature of the negative examples. Because of these challenges, reporting the accuracy of the classifier would not be representative of its true performance. Instead, we focused on understanding which features played the largest role in determining the outcome.

The case priorizer was packaged into a user-friendly Colab notebook used in conjunction with a Google form, both made accessible to Project RIC. The original process flow was for a volunteer to read a Pro Se Request and summarize the details in a memo document to be later reviewed by RIC leaders, who would decide if the case should be further pursued. The new flow requires the volunteer to answer the Google Form as they review the request, and then log into the Colab notebook once the form is submitted. The volunteer will enter the name of the client and run the cells (no coding necessary). The result is a printed output listing the top eight most important features deemed by the prioritizer, as well as which of those features are met by the

client. This can be included in the Pro Se Writeup and used to guide Project RIC on where to investigate further if the case is indeed taken on by the organization.



Digital copy of TLLF Pro Se Questionnaire



Intuitive UX using Colab Notebook. No coding necessary for volunteers

<u>Smart Search</u>

The problem statement we created to address the challenges in the case file reviewing phase was the following: Project RIC members need a way to review and understand all of the case information and compile key points because the litigation lawyers require a concise memorandum of a case to present to the Criminal Integrity Unit. The problem here is more abstract, encompassing issues faced throughout the entire reviewal phase. Specific examples include obtaining and organizing the information from their various sources (public records, private documents, police memos, etc.); parsing through hundreds of pages of information; and bringing new volunteers up to speed on the case details and next steps. Rather than superficially addressing all of the challenges, we wanted to deep dive into one aspect that would create a significant improvement in workflow.

We brainstormed several potential solutions and looked into their feasibility. A promising initial idea was to create a “linking” tool within the Google suite that could tie comments on Docs, Sheets, and Drive files all together by consolidating the information and hyperlinks into a master spreadsheet. This would make it much easier for a RIC member to find the relevant details and navigate between documents when reviewing a case. We investigated the Google API documentation to check the feasibility of this solution, and quickly realized that we would run into challenges with the API’s authentication requirements. If we were to make it work, the solution would have opposed TLLF’s requirements of needing an easily integratable tool into their current workflow. We accepted this fact and moved on to a different idea: making the information easier to find. After confirming with TLLF that a tool that does that would be extremely useful, we began development of Smart Search.

Smart Search is a workflow we set up for TLLF used to make the text in any PDF document searchable with the standard “search” feature (i.e. CTRL+F). The processing is done by a Java application that uses AWS Textract to identify text and convert the document into a searchable copy. We adapted the code from a repository linked in the AWS Documentation and set up an automated workflow using the AWS environment.[1] It goes as follows: a volunteer signs into the AWS management console and navigates to the “smartsearch-tllf” bucket. They upload a PDF document in the “documents” folder, which automatically triggers an AWS Lambda function to run the Java application. After about five minutes the file is converted, and the searchable copy is uploaded to the “processed” folder of the same bucket. Finally, the volunteer just has to download the file and use that from now on. It is important to note that the workflow is incredibly simple for volunteers: they simply have to upload, wait, then download the new file, while everything else is automated. We also shared a step-by-step guide with screenshots that walks through the process, as well as provided contact information for any questions.

Smart Search is a well packaged tool that cleanly addresses one of TLLF’s difficulties. The reviewal phase is the most laborious and time consuming part of the process. Volunteers must read through every detail in hundreds of pages so as to not risk overlooking a crucial piece of evidence. Once these elements are found, however, it begins the search for every detail about them. Here is where Smart Search is most effective; volunteers can search specifically for occurrences of a term, like a witness name or “DNA”, and document the context surrounding them. Due to the volume of information and hundreds of hours put into the review process, this tool will save many volunteer-hours over time.

We chose an AWS Textract solution for the added benefit of having a highly available platform that will be indefinitely supported. We could have opted for a free and open source Optical Character Recognition (OCR) service, but it ran the risk of becoming out-dated as technology changes. As an AWS service, Textract is constantly improving and one of the best options for OCR with typeface as well as handwritten text.[2] Additionally, Smart Search is well within TLLF’s budget. Textract charges $1.50 per 1000 pages and S3 costs $0.0023 per GB stored. Even if the organization rapidly expands, they will only be paying tens of dollars per month.

One last consideration we made was to limit volunteers’ access to other AWS services. We used Identity Access Management (IAM) to create a user with basic permissions to only upload/download files from the S3 Smart Search bucket. We followed AWS best practices of “granting least privilege” to prevent the unlikely chance of users breaking the system or using unauthorized resources. Volunteers have access to the login credentials for the base user, and Project RIC leaders were given the Administrator credentials.

---

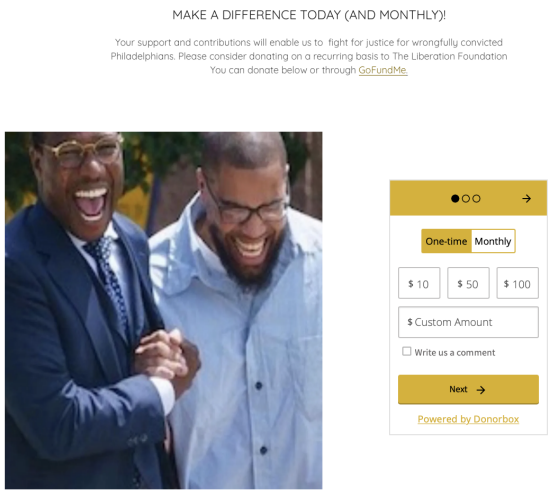
[1] aws-samples/amazon-textract-searchable-pdf: Generate searchable pdf documents from scanned documents with Amazon Textract

[2] https://aws.amazon.com/blogs/machine-learning/amazon-textract-recognizes-handwriting-and-adds-five-new-languages/

Donorbox Plugin

A small but impactful change to TLLF's website was changing their donation platform. The website, www.tlewisfoundation.com, originally used the prepackaged PayPal donation feature provided by GoDaddy, their hosting service. While this was an easy and effective solution, this feature did not allow for people to set up repeat monthly donations. TLLF wanted this ability because their fundraising rounds required a lot of resources, and having a steady monthly donation stream would give them a lot more financial stability.

We researched the options and settled on Donorbox, an easily integratable plugin that allows for one-time and monthly donations with a small fee of 1.5%.[3] After signing up the Foundation, we embedded the custom HTML into the GoDaddy site. Knowledge of CSS and HTML proved incredibly useful as we worked to find the best layout for the donation section. TLLF was very thankful for the timely assistance with this project, as we completed it just before they had a featured special on CBS This Morning.[4]



Donorbox featured on tlewisfoundation.com

Operational Restructuring

When working on technical solutions to improve TLLF's operations, we had a couple of constraints. Firstly, the solution had to prioritize ease of use, especially for people with limited technical experience. Secondly, the solution had to be low-cost since the foundation has limited financial resources. Thirdly, security was prioritized since a lot of the information contained on their Google Sheet is private information that should not be widely accessible.

Google Sheet:

Based on feedback from both the pre-surveys and interviews with Project RIC members, it was apparent that the primary Google Sheet they rely on was increasingly difficult to navigate as they took on more clients. It was confusing to figure out what the color coding represented

[3] https://donorbox.org

[4] https://www.cbs.com/shows/cbs_this_morning/video/Ogk4Ck0ut2GD_v93pwFJEhGSpYCNcsr_/philadelphia-district-attorney-larry-krasner-on-mission-to-reform-office-s-culture/

and deadlines were non existent. After completing one of the tasks in the case pipeline, the volunteers would often neglect to update the Google Sheet and relied on the student leader to inform them of the next steps. In order to increase the foundation's organization, we removed any unnecessary information from the Google Sheet. We readjusted the formatting in order for all of the columns to fit on a single screen view. We implemented formula driven checkboxes in which the cell color defaults to red when unchecked and switches to green when checked. We also devised a custom formatting function in which a cell remains green until the date in it has passed and then the cell turns red. This increases deadline awareness. Lastly, we placed the legend on its own page in the sheet. It now clearly outlines which items needed to be completed in order to move onto the next colored phase and the corresponding action items.

Kanban Board:

From discussions with new and old volunteers, it was apparent that the task assignment process needed improvement. Volunteers were often confused on their deadlines, and they felt that the work each week did not align well with how much time they were able to dedicate to TLLF that specific week. We conducted research on different project management systems, and we found that work breakdown structure would be best for them. Since all of the volunteers are students, their availability fluctuates on a weekly basis due to class assignments and exams. Work breakdown structure (WBS) allows for volunteers to view how much time each task will take, and then they can select units of work each week. We collected data from the volunteers in Google Forms to understand how much time common tasks take. So if a person has a little time to volunteer one week, he or she can select one unit of work that is estimated to take 20 minutes. Alternatively, if a volunteer has a lot of availability one week, he or she can select 6 units of work that total to 3 hours of work. Once we studied WBS, we thought about the best way to implement it. We researched different project management softwares. We tested different free softwares including Jira, Asana, and Trello. We found that Trello was the best software given the constraints. It is very easy to use and its free version is sufficient for TLLF currently.

The main goal of this project was to further a societal goal. This project is rooted in social impact and that definitely impacted the way our team approached the design. The main way this influenced our design is we were perpetually focused on actionable solutions that TLLF could realistically implement within our time horizon and available resources. Designing solutions under time and resource constraints is an essential part of engineering that is often overlooked during classes or design projects since those real world constraints don't exist. Additionally, we prioritized solutions that would have the greatest impact on TLLF's process. This gave us the guiding roadmap to select which solutions to focus on, design, and implement. Economic considerations were very important. The main value and competitive advantage behind our platform is that it's implemented without any tech solution costing over $10 to make it accessible for organizations such as TLLF. Since we were working in the non-profit sector, it was paramount that we optimize for limited resources and high throughput. Considering the low capacity of paralegal organizations and high throughput due to decades of blockage in the justice system; robustness under stress and economic efficiency were key considerations.

When it comes to the impact of the work we did over the past year for TLLF, we find it to be quite clear that our goals were met. Not only were the goals met, the impact we were able to have on the organization's operations and structure were incredibly well received. As a human systems focused project, we believed the best user testing results we could collect to validate the impact of our work would be user testimonials. Below are some of the testimonials we received:

- “The problem statements are exactly what we need help with” - Project RIC Student Volunteer Leader
- “The new spreadsheet has been a major help in making TLLF’s legal aid work much more efficient and user friendly” - Project RIC Student Volunteer Leader on Revamped Google Sheet” - Project RIC Student Volunteer Leader on Revamped Google Sheet
- “Tasks [have become] easier for volunteers to understand and keep track of, which overall increased productivity and maintenance of the litigation work” - Project RIC Student Volunteer on Kanban Board

On top of that, we heard from Terrance himself, who emphasized how the work that we did was critical to the organization. The video can be found here. Our project was social impact focused and implemented systems engineering to make a difference.

V. Self-Learning

As we approached the actual implementation of our ideas, there were subjects and areas we had to teach ourselves to move forward with the project. For Smart Search, none of the members of our team had experience with Amazon’s lambda functions and how they fit into the S3 ecosystem. Maher and Carsen had to parse through pages of Amazon documentation, online forums, and even email Amazon support to understand how to use the lambda function environment, to implement Apache-run code that plugs into Amazon S3 and can call Amazon Textract. While both had Java and Python programming experience, learning how to work with a new ecosystem and libraries is always a challenge.

For the algorithmic case prioritizer, Julie and Ellie were both familiar with the concept of computer vision and how to utilize it, but neither had implemented it in code before. They are both confident in their Python programming skills, but learning how to use the Python OpenCV library was a new topic they had to teach themselves. This required reading online documentation and working through other code available online that was related to the goals of the algorithmic case prioritizer, but not exactly the same. Overall, Julie’s and Ellie’s understanding and application of the OpenCV library and boxdetect package built on top of it were much improved.

For this semester’s development of the case prioritizer, Carsen had to deep dive into random forest classification and learn how to “featurize” the Pro Se data so that it could be interpretable for the algorithm. While he had used sklearn before in ESE 305, this was his first time applying it to a novel dataset that required pre-processing. He leaned on Professor to help him understand how to fill the holes of the data, such as by imputing N/A values and adapting to skewed label distributions. Carsen feels much more comfortable handling end-to-end ML applications that require data acquisition/cleaning, creating an intuitive UI for non-technical users, and all steps in between.

There were specific classes that were incredibly useful in assisting with the various components of our project, including both the problem identification and solution implementation phases. These courses are listed below.

A. ESE 402/542: As a core Systems Engineering course, Professor Hassani's Statistics for Data Science course taught all of us the importance of upholding statistical engineering standards and validating our algorithms appropriately. It also provided us with the foundational knowledge for the scale of training data required to execute industry-accepted data analytics methods.
B. ESE 545: Professor Hassani's Data Mining course has been really useful in providing us with the skills needed to analyze large data sets and determine the best algorithm(s) to implement given the nature of the data and our objectives. The class also strengthened our Python skills which has been crucial for writing code that can efficiently process the massive amount of data.
C. CIS 545: Big Data Analytics gave a high level overview AWS ecosystem as well as provided hands on experience working with some of its resources. This familiarity bolstered our confidence when we developed the Smart Search with Textract and S3.
D. ESE 444/544: Project Management has been an incredibly useful course for developing a more structured and logical approach to how we divide work between team members and how we communicate with relevant stakeholders. The course emphasized the importance of understanding different team members' work styles and preferences so we can optimize productivity and enjoyment while working towards our goals. We chose to implement the Kanban Board for TLLF because we learned in ESE 44 about its potential to drastically improve productivity and organization.
E. IPD 509: The course provided lots of practical experience conducting Needfinding, including conducting interviews and doing sticky note analysis. We drew on these concepts when identifying TLLF's problems and determining which ones we should address.
F. EAS 545/546: Engineering Entrepreneurship I and II proved extremely helpful when we put together our pitch deck for the final round of the M&T Senior Design Summit. The courses taught us how best to structure a start-up pitch and present effectively to business experts.
G. OIDD 236: This course provided a great set of guidelines for implementing a project with scaling in mind. All too often engineers are so busy taking a project from 0 to 1 that they forget that it needs to be taken from 1 to N. This is why we took a platform based approach that encouraged reusing data (with differential privacy) throughout multiple cases. This way, we took advantage of the network effects that arise from the platform
H. MGMT 237: This course gave us the understanding of competitive dynamics in legacy industries (specifically legal tech) and empowered us with the tools to construct efficient ways of disrupting these dynamics. We used the teachings in this course to evaluate alternative solutions to this platform and form our UVP (unique value proposition) such that our project would provide distinct value from already-existing solutions in the market.

VI. Ethical and Professional Responsibilities

The core of the project is focused on improving the processes for liberating wrongfully convicted incarcerated peoples. This inherently is tied to the larger societal issue of criminal justice reform and mass incarceration that plagues the United States. The US has the highest rate of imprisonment compared to any other country in the world and mass incarceration disproportionately impacts people of color and poor people. With recent events in the US surrounding police brutality, the prevalent issue has come to light in the mainstream. The fact of the matter is that people in the US criminal justice system are charged with crimes they did not commit, but just happen to not have access to the proper resources to prove their innocence, leading to incredibly long sentences for a crime they did not commit. We know our project cannot solve institutional problems in criminal justice, but it can address microscale issues that organizations seeking to free wrongfully convicted prisoners have to deal with.

Additionally, there were ethical issues and professional responsibilities we were aware of as we approached developing our solutions. Ethically, we used computer algorithms to make major decisions when it comes to the future of someone's chances at being considered for a reopening of their case and future representation. We made sure that we were humane with how we discuss people as data and the creation of the algorithm. Finally, the foundation currently takes cases in the order of most advocacy. This is considered to be "noise" from a technical standpoint, but there is something to be said that certain cases would be prioritized due to higher demand for representation as it occurs now. Finding a balance is our responsibility as professionals on this project to ensure fairness in their processes.

To address any potential ethical issues we faced, each of these solutions was made with constant communication and input from TLLF volunteers. Throughout our research and development stage we thought about other ways to address their problems such as creating a fully automated process to read intake forms. After talking to volunteers and experts in the space we learned that it is important to keep some questions open ended rather than binary. Throughout our entire process, we focused on ensuring that the incarcerated people are being treated with dignity and given a fair shot at a new trial. Because of this focus, we pivoted our designs as we learned more about the legal space and algorithmic biases that can arise.

VII. **Meetings**

We met with our advisor, Professor Hassani, during the beginning of the summer and had a short meeting with him to touch base at the start of the semester. We were able to establish a weekly meeting from 1:00-1:30p ET on Fridays and have successfully met every week since this was scheduled. We also had meetings with Dr. Vohra for the discussion of the "business side" of the project for the M&T requirements, and ultimately more meetings to discuss M&T finalist presentations. We also attended the bi-weekly Project RIC volunteer meetings to keep in the loop with internal operations at the organization.

Mentors and Advisors Consulted this Semester:

- Professor Hamed Hassani (consistent advisor)- hassani@seas.upenn.edu
- Dr. Sangeeta Vohra- vohras@wharton.upenn.edu

VIII. **Proposed Schedule with Milestones**

There were many spring semester milestones in all aspects of the project. Many of these milestones revolved around implementing the basic designs into the organization and converting design ideas into a real, functioning product.

- Analyze trends in exoneration cases and consolidate different datasets (completed)
- Increase number of negative cases by 2x (completed 10x)
- Narrow down the most significant variables in the case prioritizer and extract top ⅔ (completed)
- Increase the number of categorical (non-binary) features and include new types of evidence (completed)
- Automate connection between Pro Se Google Form and Colab Case Prioritizer so data can be retrieved instantly (completed)
- Raise the number of project RIC volunteers who can use Smart search to 100% (completed)
- Implement donation plugin to increase recurring donations (completed)
- Implement Kanban board according to the systems design principles (completed)
- Save head volunteer organizer 90+ minutes per week of assigning tasks to volunteers (completed, 100 minutes+)
- Reduce cost of most expensive solution $10 or less (completed)

IX. Discussion of Teamwork

Our team was formed back in April of 2020 and we immediately created an imessage chat together and a Google Drive folder where we created a centralized location to upload relevant documents and meeting minutes. We met to brainstorm different project ideas and throughout the summer discussed next steps in how we wanted to approach the solutions for TLLF. We uploaded all our assignments and different objectives for each person on Asana. Here we could update our objectives and what each member is in charge of/has completed. Ultimately, Asana became too much of a roadblock to productivity, so we switched to a simple Google Doc, where info about each “prong” of our solution had its subsolutions listed out and the current status of them to keep us on track and increase accountability. Additionally, in the documents we wrote up or created, we would assign individual tasks to different team members, as discussed in our weekly team meetings.

<u>Subdivision of Tasks</u>

Questionnaire Solution- Julie, Ellie, Carsen

- Ellie and Julie worked together in the first semester to create the groundwork for case priorizer. They met frequently on zoom to work on different aspects of the solution.
- Ellie focused more on the questionnaire redesign and studying which algorithms will be most suitable for our problem.
- Julie focused more on finding the optimal way to code our intake program and ensuring the program will be easy to use for TLLF volunteers.
- This semester, Carsen took a larger role with the case prioritizer. He worked on identifying the relevant features for the algorithm; implementing the Random Forest classifier; and creating the digital Pro Se Google Form.

Smart Search Solution- Carsen and Maher

- Carsen initially looked into the feasibility of the Comment Organizer tool and found it wouldn't be doable. He discussed this with Maher and Archit and we decided to look into text extraction tools. We settled on AWS Textract and delegated the work. Maher and Carsen worked together on developing Smart Search. One person would work on it alone, then they would both meet together to discuss the progress and hand it off to the other. We found this much more effective than working simultaneously (since we were remote, it was annoying alternating sharing screens). Maher took charge of setting up the AWS resources and access permission, and Carsen predominantly worked on the code for PDF-searchability. Carsen also created the "PDF Searchability Guide" that instructs volunteers on how to use the tool.

Communication and Check-ins with Project RIC Volunteers and Deadline Manager- Archit

- Archit focused on the relationship management between our senior design team and the Project RIC team to ensure that they were updated with the project status on our end and make sure that the work we were doing was on track and matched their goals. This has been incredibly helpful when it comes to keeping our group on track and focused on the needs of the organization.

Operational restructuring - Julie and Ellie

- Ellie has focused on doing a complete redesign of the organization's intricate Google Sheet. They rely on this sheet to keep track of which phase each client is in and the assigned volunteer's workload. To improve the Google Sheet, Ellie worked closely with the head student volunteer to determine which data was the most pertinent to display and which could be deleted to reduce complexity. Next, Ellie made the Google Sheet more formula driven with color coordinated check boxes. She implemented a custom function to change the color of a cell after the deadline in it had passed. Finally, Ellie created a separate legend for the Google Sheet that highlighted explicit action items after each step was completed.
- Julie's focus in operational restructuring was centered around creating a Kanban Board to aid in the process of assigning volunteers work. Julie researched Work Breakdown Structure, and she implemented the fundamentals of WBS in the Kanban board to allow for self assignment of tasks.

At the start of the semester, it was difficult to keep in contact with TLLF and Project RIC and make sure we were meeting the needs and expectations of the organization with our solution. This has been recently addressed, as seen through the fact that we were just added to the Project RIC listserv, which will help us keep in touch with the frequent emails and updates that are sent out by the volunteers. Additionally, we were also added to the TLLF Slack, which will allow us to identify when they have meetings and any bottlenecks that arise we may not have been aware of in our previous meetings with them. This also will allow for faster communication with members of the organization that is far easier and more efficient than email. Finally, these new changes will allow for us to more effectively schedule meetings and demos.

X. Budget and Justification

Original Budget:

Part	Estimate unit cost	Quantity	Source of funding	Subtotal
AWS S3	$.023/GB/month	100 GB for 6 months	ESE	$13.80
AWS Textract	$0.0015/page	6000 pages fo 6 months	ESE	$9.00
TLLF Donation	$100	1	ESE	$100
Total				**$119.80**

Table 1

New Budget:

We planned to use the TLLF donation to ensure that the programs we implement continue to be functional after our graduation, however the donation was not approved. We had an AWS credit, so we no longer need funding for that. Ultimately, the new budget became no budget.

XI Standards and Compliance

The IEEE standards we comply to:

- IEEE P7003 Algorithmic Bias: The algorithm should be consistent with protected characteristics such as race, age, and gender
- IEEE ead1e-Law which defines the relationship between AI and law: AI used for judicial purposes should be fair, free from bias, consistent, and accurate,
- IEEE P7002 Data Privacy: The data and personal information of incarcerated individuals should be protected and have restricted accessibility.

As outlined in the IEEE standards for artificial intelligence, we remained cautious and diligent throughout our project to ensure our AI is just. As we handed off our solutions to the Terrence Lewis Liberation Foundation, we supplied them with TLLF with guides on how to best use our algorithmic case prioritizer to ensure efficacy. The IEEE has found that the legal space is resistant to technology because of the absence of trust. Since we were given the opportunity to penetrate the legal sphere, we wanted to ensure that we were comprehensive, transparent, and effective. In the legal space, the IEEE believes that AI can be beneficial if it is speedy, free from undesirable bias, fair, accurate, adaptable, and consistent. The standards are still in development, but the IEEE recommends they should be formulated with input from legal and engineering professionals. To comply with these standards, we worked with Professor Hassani as well as lawyers to ensure that the algorithms we created were compliant with engineering and legal standards. To comply with the IEEE standards of data privacy, we ensured that the data we were given access remained confidential. Beyond IEEE standards of privacy and algorithmic bias, we also complied with AWS best practices, Python's Style Guide and the

Privacy Act of 1974.

XII. **Work Done Since Last Semester**

Operational Restructuring

After talking with the TLLF volunteers over winter break, we determined that we needed to pivot our project to more accurately address the organization's pain points. In order to accomplish this, a large focus of our work this semester was operational restructuring. This semester, we conducted research on different project management strategies, and we found that the Work Breakdown structure would work best for TLLF's needs. We conducted interviews with new volunteers, and it was clear that many people had trouble working with the foundation's Google Sheet which contains assignments for the volunteers. During the spring semester, we revamped their Google Shee, and we have received very positive feedback on the implementation. The last piece of our operation restructuring was implementing a Kanban board for TLLF. The Kanban board is held on Trello, and it allows the volunteers to assign themselves tasks. We had multiple Zoom meetings teaching the volunteers how to use all of the new Operational Restructuring solutions.

Smart Search

We finalized the technical aspects of Smart Search last semester to have it ready for the Fall Demo Day. However, it was created on Carsen's personal AWS account. This semester we set up an AWS account for TLLF and migrated the code and architecture over to it. This would give the Foundation long term access and control over the tool. We also created and shared the "PDF Searchability Guide" with Project RIC so that all members are able to use our solution.

Case Prioritizer

Most of the work on the case prioritizer was done this semester, as we focused on Smart Search in the Fall. Julie and Ellie developed a baseline prototype last semester that used a revised checkbox-based Pro Se Questionnaire with strictly binary questions. We focused on automating the intake using OpenCV to identify the responses from an uploaded picture of the form. The responses were used to check required answers: reject if the case was not in Philadelphia, the person is not claiming innocence, and the person is not currently serving a sentence. This early version did not use a machine learning algorithm, and we ultimately moved onwards when we realized it would be better to work with TLLF's existing Pro Se Questionnaire rather than making a new one. With this experience from the Fall semester, we restarted the work on the case prioritizer with the new vision.

M&T

This semester, we were fortunate to be selected as finalists for the M&T Integration Lab competition. For our presentation, we created a business plan of how we would scale our solutions to help freedom foundations across the country. We conducted research on the industry, the competitors, and the current solutions. We found that our low-cost and highly customizable solutions created a competitive advantage. The full business model can be found in the appendix.

XIII. Discussion and Conclusion

Ultimately, our experience working with TLLF has been absolutely incredible. When it comes to our solutions themselves, we first have the Case Prioritizer; a supervised learning algorithm that uses past case data and outcomes to let the organization know which cases to take on first. This solution relieves the volunteers of a lot of tedious pre-work of trying to determine which cases the lawyers will be most likely to accept. Additionally, it eliminates bias that could arise from a human prioritizing the cases. In order to create the case prioritizer, we went through many iterations. Next, we have Smart Search, which allows even the most non-tech savvy lawyer to upload hundreds of pages of scanned case file documents with typed or handwritten text and receive easily searchable interactive PDFs. Prior to the smart search solution, TLLF had to manually read hundreds of pages. If they were looking for a specific name or date, they had to flip through all of the pages, wasting a lot of time. Another technological implementation we executed upon was that we helped switch TLLF's donation platform to Donorbox, an html plugin that supports recurring donations. Their godaddy website originally supported one time donations, but they wanted recurring donations to give a more consistent income flow. We replaced their existing donation service with the Donorbox plugin and custom html to support this feature, which would still maintain pci compliance and secure encryption, without taking a cut of TLLF's donation revenues

Finally, we focused on operational restructuring for TLLF. TLLF heavily relies upon a very intricate, expansive Google Sheet to keep track of all of their clients and which volunteers are working on them. We revamped their Google Sheet to be formula driven and more clearly emphasize the action items for each stage of the legal process. To top it off, we utilized the principles of Work Breakdown Structure to implement a Kanban Board on Trello that enables students to assign themselves tasks and save the student leader 90+ minutes of work a week.

One of the biggest roadblocks for our project was access to case data of previously exonerated individuals. We were not able to gain access to the Pennsylvania Innocence Project's resources due to privacy reasons. We overcame this challenge by finding a substantial publicly available data set. However, we had to convert the data to binary by filling out the yes/no questions of TLLF's Pro Se Questionnaire from the perspective of the freed individual. We were able to amass over 100 data points to train our Random Forest classification algorithm. This enabled us to tease out the relative weight of each case feature.

It was an honor to be able to meet Terrance and work with the organization for two full semesters. Our team's individual sill sets complimented each other so well and helped us take the project to the next level. As we wrap up our last few weeks of our undergraduate experience at the University of Pennsylvania, it serves to be a time of reflection. When it comes to the lessons we've learned throughout ESE450 and ESE451 while working with TLLF, we learned that it's not always the biggest, grandest solutions or most technologically advanced solutions that work the best. Instead of focusing on "beautifying" the technology and focusing on buzzwords to make the solutions sound more "grand," we learned that the most impact comes from understanding the needs of the consumer/customer that needs help. That will lead to the most impact in your work. We also learned that using design thinking and problem identifying methods, such as Sticky Note Analysis, were the core to making our project function and address the correct problems. Without taking a few weeks to execute needfinding and truly understand what the dynamics of the organization were and how certain problems have been

incredibly relevant in preventing their optimized efficiency and growth, we would have totally missed the correct issues to focus on and would have been taken on the wrong path. Ultimately, we also learned awful truths about the United States prison and legal systems. Although the sympathy we express for those individuals who have been wrongly incarcerated will never be enough to right the wrongdoings brought against them, it has consistently driven us to put our all into this project for the past year.

XIV. **Appendix**

Business Analysis

Value proposition:

In an extremely antiquated and inefficient legal space, where 81% organizations report difficulty accessing information across multiple platforms[5] and 22.7% of Paralegal/Lawyer time is wasted on document management and searching challenges[6], we propose common-sense solutions that empower legal firms to do more with less time and access the keys to scalability.

Stakeholders:

The Terrance Lewis Liberation Foundation, wrongly incarcerated individuals (and family), paralegal/law firms, public defenders, District Attorneys.

Competitors:

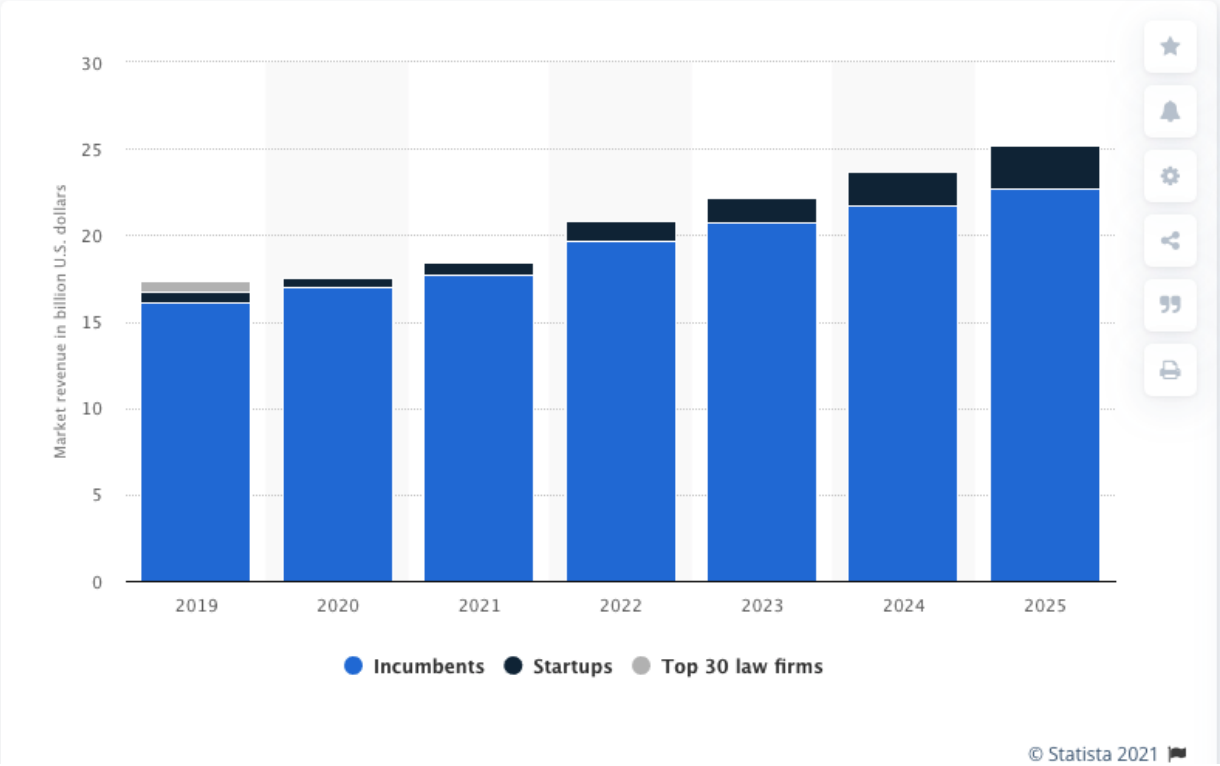
Name	Pro	Con
Luminance 	- Uses AI to make document analysis more efficient - Saves time during document search	- Structured product can only work if organizational structure is already existent and can handle additional case load
Ravel Law 	- Uses AI to offer research, analytics, and KPI on case load and performance - Uses past judges decisions to predict potential future decisions	- Expensive, making it within reach for corporate clients or big firms - Only effective at scale
Onit 	- Provides for operational and process improvements, combining business process management, project management, and information management into one tool.	- Learning curve is steep, requiring familiarity with Onit's App Builder program. Someone with tech expertise needed to set it up .

Market opportunity:

[5] Kaplan, Ari. "Are Law Firms Prepared for a Portable Future?" *Mobility Metrics*, 2015, resources.relativity.com/rs/447-YBT-249/images/AnalystReport_Kaplan_MobilityMetrics.pdf.
[6] Webster, Melissa. *Bridging the Information Worker Productivity Gap in Western Europe: New Challenges and Opportunities for IT*. IDC, Sept. 2012, warekennis.nl/wp-content/uploads/2013/11/bridging-the-information-worker-productivity-gap.pdf.

The core market identified for our products and services is legal firms and paralegal services. Our products and serhere are many trends in this. The US legal services industry is estimated to be a $364.25B market, with the global market being almost twice as big. Law firms across the world are now continuously being forced to consider ways of reducing their costs, while creating new methods that are different from the competition is a further challenge. The introduction of legal tech solutions is one such way that a firm can save time and streamline work processes.[7]

The legal tech market generated $17.32B in 2019, the vast majority of which (97%) was generated by incumbents. However, According to many meta-studies[8,9], legal tech startups are expected to fill the many gaps in the industry, and legal tech startups are expected to have a 27.82% CAGR throughout 2025. This will grow a $0.57B market into $2.49B by 2025.



Cost:

Fixed- SG&A: coding environments and general overhead, employees, marketing, customer support, software licenses

Variable- Consultant billable hours, miscellaneous COGS

[7] https://www.statista.com/statistics/1168096/legal-tech-market-revenue-by-business-type-worldwide/

[8] Kaplan, Ari. “Are Law Firms Prepared for a Portable Future?” *Mobility Metrics*, 2015, resources.relativity.com/rs/447-YBT-249/images/AnalystReport_Kaplan_MobilityMetrics.pdf.

[9] Webster, Melissa. *Bridging the Information Worker Productivity Gap in Western Europe: New Challenges and Opportunities for IT*. IDC, Sept. 2012, warekennis.nl/wp-content/uploads/2013/11/bridging-the-information-worker-productivity-gap.pdf.

Revenue Model:

We've proposed three separate (or potentially concurrent) revenues models to bring our products to market:

- Research Model: Use a mix of prepackaged and personalized data and AI techniques to analyze a firm's operations, processing, case history, connections in the industry and sell a report with recommendations
    - Revenue generated per hour billed or per research recommendation package

- Technology Solutions Model: Build a portfolio of tools personalized to the firm's needs to accelerate their processes and alleviate their pain points. Many of these products can be replicated from firm to firm.
    - Revenue generated per product delivered

- Technology Consulting Model: This approach is similar to the full scope of our senior design project. Using in-depth research to inform the development of tech solutions that integrate operations together holistically.
    - Revenue generated using the consulting fee model