

Tech for Humanity Case Studies

Automating Justice: AI and the Courts

The United States legal system like many similar legal systems around the world has a substantial case backlog. The Covid-19 pandemic exacerbated the case backlog and by November 2021 the Court Statistics Project estimated a national case backlog of 1.254 million cases. This combined with data indicating stagnant numbers of judges despite significantly increasing case load counts per judge from the Judge Information Center at Syracuse University sets the stage for the miscarriage of justice on a wide scale across, local, state, and federal judiciaries. These case backlogs impact everything from civil to criminal proceedings and trap millions of Americans in an ever-weakening legal system. Due process is enshrined in the US Constitution in both the Fifth and the Fourteenth Amendments. Yet at present the United States legal system is inundated and increasingly fails to meet these standards. Nowhere is the case burden greater than in the immigration system. In 2021 the U.S. Border Patrol reported 1.6 million encounters along the U.S.-Mexico border. At the end of 2021 there were an estimated 1.821 million pending immigration cases. These pending cases undermine the U.S. immigration system, a basic right to due process, and create a humanitarian challenge on the border.

Refugee, asylum, and parole cases are difficult because they impact people in an extremely vulnerable positions. Unlike individuals who cross the border without permission, refugees who seek parole must apply for refugee status prior to entering the country. Asylees by contrast must present themselves to Customs and Border Patrol officers at a valid port of entry or file for asylum status after entry. Humanitarian parolees are a special designation group who are admitted to the U.S. for protection in extraordinary circumstances. One of the most common contact points with the immigration system is a port of entry. A port of entry can be a land border crossing such as a border check point at the Texas – Mexican Border in El Paso, a water port of entry such as the Port of Los Angeles, or an airport port of entry such as John F. Kennedy Airport in New York. Often by the time individuals have arrived at these ports they have limited resources, including financial, but also it is not uncommon to have few documents, or other critical personal information identifiers. Frequently individuals who are seeking humanitarian parole do not speak English and many may not have friends or family already inside the United States.

The result is a complicated process that is difficult to navigate. The number of accepted refugees, asylees, and parolees varies wildly by year, based on global geopolitical, economic, environmental conditions, and domestic politics. In 1980, the United States welcomed 207,000 refugees. By 2020 that number had declined to 12,000. Under U.S. law there is a difference between asylees, refugees, and parolees. Whereas refugees are required to apply for lawful permanent residence in the U.S. 1-year after entering the country, Asylees may apply to become lawful permanent residents but are not required to do so. There are procedural differences as to where and when a person may enter the United States.

Tech for Humanity Case Studies

The complexity of the system has resulted in stagnation in an immigration system that is understaffed, overwhelmed, and slow. There are no readily apparent solutions to the challenges being faced by any of the actors involved. The trend towards increasing numbers of individuals waiting for their cases to be heard seems to be unstoppable.

Yet hope may be on the horizon. A recent study by Nu Wang highlights how China has turned to Artificial Intelligence to address issues of judicial reform. Wang's study identifies System 206, an AI system developed to aid judicial procedures case management, she also explores the use of AI judges in what are increasingly known as "Smart Courts (智慧法院)".¹ In some instances robot judges are even used under supervision to substitute for human judges in small claims disputes.² Although still in the test phases in 2020, these systems were credited with improving efficiency including concluding certain cases in as little as 30 minutes. The overall feedback to the use of AI in the Chinese judicial system has been positive. In many ways China, like the United States faces the same bureaucratic challenges. Addressing these challenges within the ethical, moral, and cultural constraints of each country is possible. In 2021, knowledge of use of Artificial Intelligence in the Chinese judicial system spurred AI research company Augmented Justice Incorporated (AJI), a law enforcement and legal AI software product development firm, to initiate research and development on the potential use of AI in the U.S. legal system. The CEO of AJI research, a first-generation daughter of Guatemalan immigrants who spent years stuck in the U.S. immigration system, tasked her R&D department to focus on developing an AI enabled product to help ease the immigration case burden. As a test product they sought to create a program that focused on the intake, processing, and analysis of documentation pertaining to refugee, asylum, and humanitarian parole seekers.

AJI's focus on asylum cases stems from concern for the humanitarian needs of asylum seekers who now often wait an average of 1,621 days for a hearing. The period between temporary parole and formal asylum status can be emotionally difficult and places families and individuals under duress and uncertainty. Providing a pathway to clear up the backlog and alleviate humanitarian concerns is central to AJI's platform development strategy. To begin their program development, they spent time analyzing the laws pertaining to refugees, asylees, and humanitarian parolees, the procedures of Customs and Border Control (CBP), the United States Citizenship and Immigration Services, the Department of Justice guidelines to Immigration Judges, the experiences of asylum seekers, the NGOs, legal advocates, and other organizations that provide assistance.

The requirements for asylum are codified in 8 U.S. Code § 1158. The law requires that to establish that the applicant is a valid refugee, asylee, or parolee within the meaning of such section, the applicant must establish that race, religion, nationality, membership in a particular social group,

Tech for Humanity Case Studies

or political opinion was or will be at least one central reason for persecuting the applicant. In the case of humanitarian parolees, other conditions may apply based on the nature of the humanitarian crisis. The first task of the team was to decide on which AI model would best fit the problem they were attempting to solve. Namely, they needed to use a model that would take in large volumes of data and output decisions that met the legal standard above. The team had previously focused much of its AI work on Bayesian modeling for facial recognition. Bayesian modeling looked at the probability that any given image was probabilistically similar to other images. The team also used Deep Neural Networks to explore image recognition and had made some progress in law enforcement applications. The team understood the challenges associated with AI applications in image recognition, in particular, they had been dealing with issues of bias in their model that made their models unable to accurately identify persons of color with the same fidelity as white individuals.

The team wished to avoid many of these issues in their new research efforts. Exploring different AI model types, the team decided to settle on supervised decision-tree models. The decision to start with a supervised as opposed to an unsupervised model stems from both practical and ethical concerns. First the team did not have a robust initial sense of the corpus of potential asylum cases. As a result, they were uncertain as to what an unsupervised baseline would produce and whether such training would result in implicit or explicit bias to mitigate potential concerns about the fidelity of output from the AI the team decided that as a new entrant to the space they were best suited to build a model that required supervised training.

The team further tried to do their due diligence in developing a supervised training method. The team understood that in supervised training methods there was the possibility of both explicit and implicit bias. To avoid this the team set up a series of controls in an attempt to mitigate bias. The team set up a diverse competitive supervised training model that required three individuals of at least two genders and two ethnicities review each data point within the training sample. The trained data were subsequently reviewed by a third diverse team comprised of mixed gender and ethnicity individuals who evaluated the trained data evaluated for bias based across race, gender, sexual orientation, ethnicity, and several other identified factors. The team provided feedback to the initial data trainers and sought to engage in dialogue and discussion on issues pertaining to bias. After the initial process of data training a second round of data training was undertaken. The data from this second round of training leveraged the same diverse composition of trainers as the initial round. The supervised trained data was subsequently fed into the model.

The supervised decision tree model was run against one million historical immigration cases with known results in the courts. The model output demonstrated 75% decision fidelity with prior cases. Among those cases that did not match known court results, 70% suggested humanitarian parole be granted where courts declined to grant parole and 30% suggested declining

Tech for Humanity Case Studies

humanitarian parole where courts ruled humanitarian parole was warranted. Although the team was proud that their initial model had achieved a relatively high degree of fidelity they recognized that a 25% discrepancy between the model and real-world outcomes would result in tens of thousands of rulings impacting real people in ways that did not align with the existing system. In particular, the team was concerned most about the 30% of cases in which courts had ruled humanitarian parole was warranted and the model found it was not. It was at this point the team realized they had a problem with the existing decision-tree model. The model as developed did not provide sufficient information on how it arrived at a given decision.

The team architected a mechanism to understand the AI decision-process by providing an output log at every stage of the decision process. At each stage of the decision process the AI would output a section of text with the supporting documentation, pair it with the legal standard, and provide links to relevant prior case decisions. The result was an auditable machine “thought” process that provided justification for ultimate decisions derived by the machine. The amount of data generated in this process was large, but provided a unique, robust case file for review should any case be challenged by either the humanitarian parole applicant or the U.S. Government.

The AJI team re-ran the entire dataset of cases back through the decision tree model and achieved a nearly identical result as it did the first time. Yet this time each decision also included a subfile information explaining how the ultimate decision was derived by the AI. Because the discrepancy was still present between AI results and actual court results the team wanted to better understand why and where divergences occurred and if the model needed better training through supervision or whether there was bias imbedded somewhere in the model.

The AJI team wanted to understand the discrepancy between their model and the actual case outcomes. The number of cases where discrepancies occurred constituted a sizeable number of approximately 250,000 cases. The team took a representative random sample of the cases ~1000 in all and hired a team of immigration lawyers to review the AI decision and compare it to the actual court outcomes. The lawyers spend three months reviewing the documents generated by both the AI and by the court in making a decision on humanitarian parole. In the end the legal review team found bias in the court decisions that discounted the claims Latin American Male applicants between the ages of 17 and 34 years of age. After removing this demographic group, the model’s fidelity exceeded 95%.

Nine months after initiating the process of building and training a decision tree model AJI believed it had a product ready for beta market release. AJI presented their product to the Department of Homeland Security (DHS) as a tool to augment judicial rulings and expedite the hearing timelines of prospective parolees. DHS agreed to beta test the software as a recommendation engine and as a way to identify bias in rulings. DHS and AJI began instituting

Tech for Humanity Case Studies

the program in fall of 2024 and within 1 year the software had been used to analyze more than 400,000 cases. The data derived from the new cases and subsequent court rulings improved the fidelity of the system to greater than 90%. By January 2026 DHS had purchased a license for the system and required all humanitarian parole applicants to file paperwork through a web-based interface. Applicants were given an immediate probability of parole score based on automated adjudication. This score had no legal bearing but could be used by the applicant to determine whether their case had merit or not. The system also provided the applicant with the complete decision-tree of auditable information. This information could be provided to a representing attorney. The court would also receive a copy of the automated analysis of merit.

By 2027 AJI's system was having a dual effect. It was providing information to individuals who were subsequently self-selecting not to complete the humanitarian parole process. This reduced the number overall cases entering the system by about 15%. While the overall volume of applications submitted to the system, the continuance of applications beyond assessment by the AI showed a decline. AJI and DHS speculated that the system was serving as a deterrent to individuals with weak cases. AJI's system also increased the speed of processing to get a hearing by about 75% and reduced the time spent in court on each case by about 60%. Because all parties entered the court room with a common understanding of the facts of a given case, less time was spent on deliberating over the specifics for each case. As a result, the case backlog of refugee, asylee, and parole cases was slowly declining. AJI was extremely pleased with the outcome of its program and was by late 2027 beta-testing systems for other applications within the judicial system beyond immigration.

Discussion Question #1

Waiting for a case to be heard in court can be stressful but is it a constitutional violation? What is the human impact of long case processing times? Understanding these impacts and their relationship to broader humanitarian and legal considerations is an important first step to thinking through how to address and solve these problems. It is important to remember that each immigration case out of the millions listed above impacts a real person and often entire families. Long case processing times can leave individuals in precarious positions in their country of origin, in a third-party country, or even within the United States. Without formal legal permission to travel to and enter the United States, legal status within the United States individuals may be prevented from achieving basic levels of human or economic security.

Discussion Question #2

While long processing times for cases may be a bad thing for individuals, there are some politicians who believe that these long processing times serve as a natural deterrent to immigration. What are the implications of deterring immigration for humanitarian reasons? In particular, what are the implications of AJI's system providing a probability of parole score? Does

Tech for Humanity Case Studies

such a score violate human rights? Does its deterrent effect potentially increase the risk posed to certain groups or individuals? Or, does having a probability of parole score improve the likelihood that the individuals most deserving of parole in some form will be granted that status more efficiently?

Discussion Question #3

Immigrants and their reasons for immigration are extremely diverse. While the laws outlining who does and does not qualify for refugee, asylee, or parolee status provide a framework in which to identify individuals warranting such status the specifics of each case are extremely unique. Based on your knowledge of bias in AI systems what types of bias are most likely to occur in AI systems such as AJI's? Did AJI do a good job of thinking through bias? Could it have done better? If so, what could it have done?

Discussion Question #4

AI's often obscure how decisions are derived through complex algorithms involving neural networks, decision trees, or related methods. This obscurity is known as the black box. Within this black box of algorithms, the AI makes a series of "choices" and then outputs an answer. In legal cases the answer is important, but it is often equally important to know how an answer was arrived at by a judge or jury. Cases that reach the Supreme Court are heard by the justices and followed by both a decision (the adjudicated outcome) and an opinion, an often lengthy document that explains how a decision was attained. Often the court will issue multiple opinions to accompany a singly decision. These opinions include *majority* or *plurality*, *conurrence* or *dissent*. These opinions help establish what is known as precedent. While dissenting opinions are not binding, they can be seen as a persuasive authority to guide future decisions. Does AJI's logging system for its immigration AI sufficiently open the black box on its automated decisions? Is there a difference between a human rendered and a machine rendered justification for a decision? Could AJI's software one day obviate the need for judges and lawyers as is select cases within the Chinese Judicial system? What else could AJI have done to increase decision transparency in its AI system?

Discussion Question #5

AI's are becoming increasingly pervasive in human interactions. Often, we do not notice their involvement in our day-to-day decision processes. As we as citizens and individuals in a dynamic democratic society become more comfortable with AI's should we trust them in situations to include judicial settings or are there certain areas where humans should either remain in the loop of decisions or where AI's should be excluded? Justify why or why not AI's should be included or prohibited from certain aspects of life.

Reflecting on Automating Justice

Tech for Humanity Case Studies

Als are still in the early stages of development. As computational power and algorithms continue to improve the applications of Als are likely to expand. We already see the expansion of AI into new spheres of everyday life. This expansion can often be explicit in the form of virtual private assistants, but it can also be subtle and hidden in applications, services, and systems that we use on a regular basis. The expansion of AI whether explicit or implicit comes with challenges. Along with the many positive benefits that Als can bring, they can also result in negative outcomes as a resulting of alterations in the behavior of individuals and organizations. One increasingly explored area of Als finds that they can often result foster outcomes rooted in biases that impact certain demographic groups more than others.

Technologists often seek to resolve perceived inefficiencies in everyday life. As a result, many Als are designed to increase efficiency. As is the case of AI's immigration system the goal is to increase the efficiency of a system that over time became inefficient due to multiple concurrent factors. Yet the goal of the justice system is not efficiency by justice. Understanding that some systems are designed to privilege other attributes than efficiency is important when assessing whether an AI will be useful or not. Likewise, Als that are well-designed and take into account issues such as bias might not emphasize efficiency gains, but equity or other values that are equally important. Implementing Als in legal and immigration settings requires the consideration of challenges including accountability, transparency, dehumanization, legitimacy, and inequality.

Accountability: The U.S. has accountability within the immigration system that spanning all three branches of government. Very often the enforcement of immigration rules and laws falls to the Executive Branch through the Department of Homeland Security and its sub-agencies including CBP, United States Citizenship and Immigration Services (USCIS), and Immigration and Customs Enforcement (ICE) and the Department of Justice through Immigration Courts. Both departments monitor and evaluate the implementation of executive orders from the Office of the President and laws passed by congress. Each agency has an inspector general and must report to both the executive branch and to the legislative branch (Congress). Where there are questions on the constitutionality of enforcement or adjudication cases can be brought before the judicial branch. Historically, accountability has been an entirely human endeavor that resides in human decisions. The inclusion of AI into the mix raises questions about where accountability resides when an algorithm "decides." Does accountability remain with those tasked under U.S. law with enforcement, or could accountability also in some part fall on the company that produces a technology? To whom should an immigrant or organization advocating on behalf of immigrants see redress? While these questions are not explicitly addressed, the U.S. system tends to resolve responsibility back to the agencies and departments tasked with enforcement and adjudication. As algorithms substitute for human decisions, understanding who is responsible and how they can be held accountable is an issue of concern.

Tech for Humanity Case Studies

Transparency: The present immigration system has been plagued with issues pertaining to transparency. Often individuals seeking refugee, asylee, or parolee status do not know why wait times are so long, or how decisions will be made. Some improvements have been made. Improvements including expected processing times and online systems where individuals can check on the status of an application are a good first step. Yet as AIs enter the mix ensuring that transparency improves and the black box decisions of AIs are understandable and tied directly to legal or policy requirements will be increasingly important.

Dehumanization: Presently the immigration process for refugees, asylees, and parolees can feel quite dehumanizing. Long wait times, unclear adjudication, and the general complexity of the processes can leave applicants feeling trapped by bureaucracy. Yet, the reality is that while it feels dehumanizing every step in the current system is examined by human eyes and decisions are made by humans for other humans often working within legal or policy constraints established by the executive or legislative branches of government. If these processes were to shift to AIs that make “decisions” based off algorithms either supervised or unsupervised by humans then the process becomes intrinsically dehumanized. Algorithms do not have empathy. Algorithms do not and cannot understand the human condition. While the ultimate decisions might appear to be based on empathy, rationality, or other “human” factors, the reality is far different. While their inclusion in decisions might foster efficiencies that might make the process feel less dehumanizing, they are, in reality, shifting the burden away from human actors to machines making “decisions” based on programmed algorithmic structures that can and often do have biases, coding failures, or other bugs that undermine their effectiveness. It is critical not to lose sight of the why an AI is being utilized and it is equally important not to forget who this utilization impacts and how.

Legitimacy: The legitimacy of the U.S. immigration system is derived through the constitutional power assigned to the executive branch, and the legislative branch both of which are checked by the judicial branch. This system of checks and balances pits the branches against one another to ensure that rights are protected. An algorithm can be programmed to fit within the laws and policies established but its decisions only derive legitimacy if they are in concert with the constitutional powers assigned. An AI system that is either too lenient or too strict in making “decisions” is likely to undermine the legitimacy of the institutions it is standing in for in the decision process. It is also likely to be an easy target of politicians on both sides of the political spectrum. AI’s immigration AI needs to retain its accountability and transparency as a bulwark for maintaining its legitimacy. If either of these other two facets falters the system is likely to lose legitimacy. Shifting away from human to machines for certain decisions is also likely, at least initially to be viewed as illegitimate. The Chinese approach to AI inclusion in its courts is instructive. A slow and deliberate inclusion over time will shift the perception of its legitimacy. These points aside we must necessarily return to the question of whether we want AI to be a

Tech for Humanity Case Studies

legitimate substitution for human decisions on matters such as immigration and the adjudication of court cases. It might be that morally and ethically it is better to retain inefficiencies in the systems than it is to cede power over such decisions to an AI.

Inequality: There is noted inequality within the U.S. legal system. The immigration system is no different. Individuals with financial resources can hire lawyers and influence politicians to shape outcomes. The history of U.S. legal and immigration rulings is rife with incidents of bias and discrimination. Repeated efforts to reform the systems have made progress but inequality remains. Individuals are frequently discriminated against on the basis of sex, ethnicity, national origin, religion, and many other factors. Recent history has shown that AIs often “bake-in” biases of programmers, AI supervisors, and existing procedural outcomes. AI’s AI is not working with perfect data on immigration. Rather, the system has learned based on prior case outcomes which contain the biases of the system accumulated over years of cases. A failure to recognize bias and discrimination in the system will result in an AI that reflects our failings. Such an outcome would be extremely damaging and likely enhance inequality rather than alleviate it. If, however, the AI is built and honed to avoid biases it might inadvertently result in a more equitable system and improved rule of law.

Wang, Nu. “‘Black Box Justice’: Robot Judges and AI-Based Judgement Processes in China’s Court System.” In *2020 IEEE International Symposium on Technology and Society*. IEEE Xplore, 2020.

¹ Nu Wang, “‘Black Box Justice’: Robot Judges and AI-Based Judgement Processes in China’s Court System,” in *2020 IEEE International Symposium on Technology and Society* (IEEE Xplore, 2020).

² Ibid.