Potential implications of using machine learning to supplement understaffed USCIS

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Sanha Kim

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Joshua Earle, Department of Engineering and Society

Introduction

The United States Citizen and Immigration Services (USCIS) is unable to process the number of cases being filed. It suffers from underfunding and understaffing, and its backlog of cases keeps growing with no signs of stopping. According to an independent office in the Department of Homeland Security, "there were 8.5 million pending applications at USCIS as of April [2022], and over 5 million of those were pending beyond their deadlines. By comparison, the backlog was around 2.7 million in July 2019" (Maurer, 2022). The pandemic worsened the pre-existing funding issues in the USCIS because of its "shutdown of in-person operations in March 2020". This, compounded with the decrease in travel during the pandemic resulted in a "40 percent reduction in immigration filings submitted between March and May 2020", which was devastating since "97 percent of the agency's operating budget is derived from the filing fees" (Maurer, 2022). This decrease in files submitted counterintuitively increased the backlog since "many employees left the agency anyway, worried about job security" (Maurer, 2022).

I propose introducing a machine learning algorithm to lower the number of cases in the backlog. While this paper will touch on how such a system might be constructed, this paper will largely focus on the potential effects of such a system on the parties involved. These parties being the immigrants themselves, the American economy, and the workers at the USCIS.

What is Machine Learning and How can it Solve this problem?

What is machine learning and how does it work? Machine learning is the process of training a program to learn how to solve a type of problem. Samuel A.L. designed one of the earliest machine learning algorithms which was trained to play checkers. Samuel explored two different types of machine learning in his paper, rote learning and generalization. Rote learning

involves a simple memorization of every single board state the program had seen. It is not useful for the purposes of this paper because it is only useful where "highly specialized techniques are required" (Samuel, 1959, Page 218). The generalization strategy had two versions of the program, Alpha and Beta, play a game against each other. "Alpha generalizes on its experience after each move by" slightly modifying parts of itself that seem to be unimportant (Samuel, 1959, Page 214). Beta, on the other hand, remains static throughout the game. Then, if Alpha wins an arbitrary number of games, Beta copies the current state of Alpha when it won the game. If Alpha loses enough times, "it is assumed to be on the wrong track, and a fairly drastic and arbitrary change is made" to it in order to ensure the program doesn't get stuck (Samuel, 1959, Page 214). After a few dozen games, Alpha was on the level of a better-than-average player. The games Alpha played as it improved is analogous to data used to train modern AI. Just as more games led to Alpha becoming a better player, generally, "more data… will lead to lower estimation variance (and thus lower error)" (Martens, 2011, Page 21).

Why is the reduction of the immigration backlog a problem which machine learning could potentially help solve? Although not every type of case filed in the USCIS can be automated effectively with a high accuracy, some can. Some types of cases involve highly personal materials, such as those involving asylum seekers, where there are "various factors for consideration when evaluating whether past threats made against an asylum or refugee applicant constitute persecution" (USCIS RAIO, 2019). These factors cannot be easily quantified into data, so it would be difficult to train a computer to make these decisions accurately. These cases also require human to human interaction to get the facts, and would not be suitable for automation by a machine learning algorithm. On the other hand, there are cases like the Employment Authorization Document (EAD), "the most frequently filed immigration benefit" (Maurer,

2022). Oftentimes, these cases are for simple renewals of "eligible, previously approved workers", and do not require any direct human to human interaction (Maurer, 2022). In fact, these cases now require such little human interaction that it, along with "twelve [other] USCIS forms, are now available for online filing, and the agency plans to offer 'end-to-end online filing and processing for all immigration forms by the end of fiscal year 2026'" (Maurer, 2022). Since all of the information is heading towards digitization already, it would make it easier to use that data to train the new AI. Additionally, with the millions of cases filed to the USCIS per year, there is no shortage of potential data that can be used for training.

A machine learning algorithm also aligns with many of the guidelines laid out by USCIS's Refugee, Asylum, and International Operations (RAIO) Directorate officer training. Although this officer training document publicized by a Freedom of Information Act (FOIA) request is specific to officers working in a specific subsection of the USCIS, the general nature of the guidelines should still be applicable to the organization as a whole. The guidelines state that officers "should strive for consistency in applying the law from one case to another" (USCIS RAIO, 2019). A single entity processing each of the mundane forms would be far more consistent than any group of humans could ever be. The guidelines also have to specify not to take into consideration "[Their] personal opinions and beliefs" (USCIS RAIO, 2019) since it is extremely hard for humans to cast aside preconceived notions. The personal opinions and beliefs of the system cannot corrupt its decision making since it has no opinions and beliefs. At least not in the traditional sense. There is a chance of human beliefs corrupting the system's decision making, but that will be discussed later in the paper.

How a Machine Learning System Might be Implemented in the USCIS.

Although machine learning cannot be used for many legal proceedings due to their complexity, it can be beneficial for cases "involving a high volume of similar types of case and relatively clear-cut legal rules" (Miller, 2019, Page 5). As shown before, the USCIS has no shortage of cases to use as training data for the AI, with millions of cases in its backlog alone. By its nature, machine learning depends on "firstly, a large dataset of past cases and, secondly, that new cases have similar features to past ones" (Miller, 2019, Pages 5-6). This large amount of data may seem beneficial at first glance, but if the dataset of past cases is contaminated by prejudice or human error, the machine's predictions would also be flawed. A machine learning tool cannot make a "correct" decision. The only thing it can do is predict how the USCIS would make decisions based on prior decisions.

A machine learning algorithm trained on all past cases without care will carry over any previous faults in the decision-making process to the new system. Although an automated system would be faster than the current manual review process, the goal should be to improve the fairness of the decisions that are made as well. It is important to take into consideration that a profile of a case to be approved or denied should not be the sole determining factor. The government has a history of profiling certain peoples, so there is a real danger of a machine learning algorithm being used to perpetuate discrimination if it is trained on government data. If such a system were to be implemented, there would need to be strong oversight by a regulatory committee in order to ensure racial, gender, or age discrimination isn't built into the system. And ideally, a culture of virtues would be cultivated in that regulatory body. "Those virtues are supposed to raise the likelihood of ethical decision-making practices in organizations that develop and deploy AI applications" (Hagendorff, 2020). At the very least, there should be some sort of auditing, and evaluation of the training sets before they are used.

The current machine learning business's standard "ethical guidelines postulate very broad, overarching principles which are then supposed to be implemented in a widely diversified set of scientific, technical and economic practices" (Hagendorff, 2020). Hagendorff argues that instilling virtue ethics into companies and the broader technological community is the better way to guide how AIs should be used and made. Virtue ethics is a type of ethics focused on cultivating virtues in people. A person who has virtues would act in a good way without having to adhere to strict guidelines. Technology does not just spring out of the ground. It is created by the people of a society to influence that society. It does not matter how beneficial a machine learning algorithm could be in theory if it is constructed haphazardly in practice. Therefore, the people chosen to work on this system should be carefully chosen, and diverse in order to ensure no group is overly advantaged or disadvantaged by the prejudices of the people behind the project.

The purpose of this paper is not to determine the specifics of how to implement a machine learning algorithm to solve the problems of the USCIS, but rather what the potential effects could be if one were used. As such, the paper will go no further into the specifics of how such a system might be implemented.

Benefits and costs to immigrants

It is easy to forget sometimes when reading all of these statistics, but those numbers represent millions of human lives which are dramatically altered by these decisions. These are time critical decisions which have an effect even if they are just delayed. "Biraj Nepal, a Nepali asylum seeker… went on unpaid administrative leave starting on January 26 [2022] because his work permit expired… If Nepal isn't issued a new work permit within 90 days of taking administrative leave, his company will, by law, no longer be able to hold his job for him" (Narea,

2022). Even if everything is happening as it should, just the fact that nothing is changing can be extremely stressful, as comedian John Oliver reminisces his own wait for a green card in his late-night show. "Oliver also recalled the stress and expense of living on a visa, having to file for extensions every year and physically leave the country every 11 months to get his passport stamped at the American Embassy in London, all while fearing he wouldn't be allowed to come back" (Martinelli, 2019). If nothing changes, that backlog is 8.5 million John Olivers & Biraj Nepals living with that chronic stress hanging over their heads. Not choosing to do anything about the issue is a choice as well.

If the machine learning system would only be as good as the current USCIS, that would still be a success. In this case, the only thing that would change is the time people have to wait for their case to finish processing, not the outcome of the case. The shortened wait times alone represent an incalculable amount of relief and economic opportunities for applicants. In some cases, it could even be the difference between life and death. "Even though asylum-seekers can stay in the United States while awaiting a decision, breaking through the delays can yield significant and potentially life-saving consequences" (Grossman, 2022). There would also be significant savings from having to pay fees to file certain cases or even travel to do so such as in John Oliver's case.

The machine learning algorithm might in fact harm immigrants, as evidenced by the contemporary problems of the governments' use of predictive policing and profiling. Law enforcement does not solely use the existence of past crimes to monitor people and communities, but the specific features of people who are likely to commit crimes (Miller, 2019, Pages 2-3). It used to be the case that profiling was only reserved for rare and serious crimes such as rape and murder, but the advent of machine learning makes it easier to profile for more common crimes

such as car theft. People train AI on data to get certain behaviors out of it. As the AI is trained on larger and larger data sets, it gets more accurate in its predictions. As AI made it easier to profile, police could use that data to more closely monitor people or locations that are statistically more susceptible to crime (Miller, 2019, Page 3). However, profiling is not evidence. Police who use predictive policing in such a way are monitoring these people and communities which have not shown probable cause to be criminal just because they have similarities to previous offenders. If the building of the system is not handled carefully, it is clear how these same types of problems could leak into it and harm immigrants.

Even if the process of building the AI is perfect, it could still harm the very people it is trying to aid. No machine learning system can be perfect even if the system would be fairer than humans and output decisions consistent with the USCIS goals and standards. 100% accuracy is only possible when solving problems with a limited range of possibilities which can be memorized through rote learning, such as finding the optimal move in Tic-Tac-Toe, but not for the vast majority of problems. "100% accuracy means that you've built a system that has memorized the training set, and such a system is unlikely to identify anything that it hasn't memorized"(Loukides, 2017). The decisions would be able to be appealed, as almost all USCIS decisions are, but "filing an appeal does not delay any decision in [their] case from going into the effect or extend a previously set departure date", so these people could still suffer irreversible harm (USCIS. Questions, 2021).

Potential Benefits and Costs to American Economy

Before jumping into how everything could change once the system is put in place, it is important to first take a look at how the current system is affecting the country. Even if we don't look at it from an emotional level, and strictly from a practical level, the current system is

harming America. "An application to obtain permanent residency (a green card), form I-485, for a high-skilled immigrant used to take four months and a couple of weeks in 2012; now, it takes nearly 11 months" (Martino, 2023). It is getting increasingly difficult for even the most productive immigrants to get into and stay in the country. If obstacles keep being raised to their immigration, they will take their skills and put them to use in another country. "Highly skilled foreigners, in particular, are less interested in immigrating to the U.S.: only 9% of interested migrants to the U.S. are college-educated, while 40% didn't finish high school, compared with 19% who are college graduates and 22% who are high school dropouts for Canada" (Martino). There may be other factors besides difficulties immigrating which contribute to that statistic, but it is undeniable that it is an important factor. Foreign nationals who earn a doctorate at a US university have the option to either stay in the US to find a job, or leave the country and use their doctorate elsewhere. If the goal is to have the domestic economy grow, the country should be trying to retain highly skilled, highly productive workers. However, "current (as of January 2016) predicted GC wait times for immigrant doctoral recipients from India and China are 10 and 6 years, respectively. This may decrease the probability of retention of fresh graduates (0–2 years since graduation) from these two countries by one-half and one-third, respectively" (Khosla, 2018). The current system has the country producing highly skilled workers and then chasing them away.

A reduction of the backlog by means of automation would have significant impacts on the American economy. Immigrants will need to work after being accepted into the country, which means an influx of immigrants also means an influx of new workers . It is hard to calculate the exact number of new workers, but economist Gordon Gray estimates that "eliminating the backlog by FY2026 would add over 938,000 new workers to the U.S. economy"

(Gray, 2022). These approved one-time immigration benefits would lead to life-long contributors to the economy who would collectively "add an average of nearly $110 billion per year in real terms to U.S. gross domestic product (GDP), or over $1 trillion over a decade" (Gray, 2022). It is also important to note that immigrants do not work in all industries equally. While immigrants with less than a 4-year college degree only make up about 11% of the US workforce, they make up over a third of the workforce for the farming, fishing, forestry, building & grounds cleaning, and maintenance industries, as well as almost 30% of the food manufacturing, textile & apparel manufacturing, and hotel industries (Sherman, 2019). Therefore, an influx of new immigrants by the clearing of the backlog should impact these labor-intensive industries more than the average numbers would suggest.

Immigrants can bolster the US economy in ways that cannot be seen by solely considering the increased number of workers. The nation is aging. By 2034, The US Census Bureau projects that people over 65 will outnumber people under 18 due to a diminishing fertility rate and higher life expectancy (Vespa, 2019). This aging population results in a growing group of seniors who are less able to work, and "immigrants are vital to helping us improve our ratio of workers to retirees and support the baby boom generation in their retirement years" (Sherman, 2019). Approving new immigrants would improve this ratio since "about 78 percent of the foreign-born population is of working age, between 18 and 64 years, compared with just 59 percent of the native born" (Sherman, 2019). This increase in the work force by foreign born workers helps support aging Americans by feeding into the Social Security System. "The program's trustees estimate that for every 100,000-person increase in immigration the long-range actuarial balance of Social Security improves by .08 percent of taxable payroll" (Sherman, 2019). This economic growth may make it seem like there are only positives for the American

economy if more immigrants' cases are completed and approved, but there is another side of the story to consider as well.

The total wealth of the country may increase, but some worry about which groups the wealth benefits. Economist George J. Borjas agrees that additional immigrants help grow the economy, estimating "the current 'immigration surplus'—the net increase in the total wealth of the native population—to be about $50 billion annually" (Borjas, 2016). However, Borjas argues that an influx of immigrants leads to a wealth redistribution from the employees to the rich employers. The basic law of supply and demand states that when supply of a good or service goes up without a proportional increase in demand, the price will fall. The hours an employee works for an employer can be seen as a product that is 'sold' to the employer. So, "when the supply of workers goes up, the price that firms have to pay to hire workers goes down" (Borjas, 2016). Historical trends "suggest that a 10 percent increase in the number of workers with a particular set of skills probably lowers the wage of that group by at least 3 percent", and after Crider Inc., a chicken processing plant, lost about 75% of its workforce in a raid by immigration agents in 2006, "Crider placed an ad in the local newspaper announcing job openings at higher wages" (Borjas, 2016). These saved costs from lower wages don't just disappear, they stay in the pockets of business owners and executives, who are often already some of the richest Americans. "The total wealth redistribution from the native losers to the native winners is enormous, roughly a half-trillion dollars a year" (Borjas, 2016).

Benefits and Costs to USCIS Workers

There is a possibility that by attempting to help immigrants by automating away the backlog at the USCIS, we may inadvertently automate away the jobs of the workers at the agency. However, I do not believe that my proposed machine learning system would take away

many jobs. With any new automation of a role, there is the fear that those jobs would be completely automated away, leaving people jobless. There are projections that in the "United States, automation could displace 20 to 25 percent of the workforce by 2030" (Manyika, 2018). While that statistic is scary at first glance, the same study predicts that "Nearly all occupations will be affected by automation, but only about 5 percent of occupations could be fully automated by currently demonstrated technologies" (Manyika, 2018).

This system will likely only help boost USCIS's output without jeopardizing many, if any, current jobs. The USCIS is currently, and has a history of, being overrun by too many tasks, causing a backlog to pile up. Even before the recent pandemic skyrocketed the backlog to 8.5 million cases, there were about "2.7 million in July 2019" (Maurer, 2022). A job can be broken down into a variety of tasks which need to be completed. The AI system would merely be processing menial cases to lessen the burden of these tasks on workers so they can spend their time on those tasks which AI cannot automate. "A key reason why it is hard to automate jobs with AI is that virtually all AI is 'narrow AI', focused on doing one thing really well" (Atkinson, 2018). A specialized AI that was trained to automatically process EADs would have no idea how to process green cards. It would be worse than asking a clerk to design the architecture of the hospital they work in. This inherent barrier to automating away jobs does not disappear by just training a new AI to cover those cases which the previous AI does not know how to process.

Even in an AI dominated world, there are certain "Social, emotional, and higher cognitive skills, such as creativity, critical thinking, and complex information processing," which will grow in demand in the wake of mass adoption of AI (Manyika, 2018). AI cannot yet emulate these skills which are essential for a plethora of jobs (Manyika, 2018). Many of the tasks done at USCIS require certain social and critical thinking skills which cannot be automated. For

example, an official at the USCIS needs to determine whether this requirement of "there has been a reasonable effort to locate [both parents] as determined by a competent authority in accordance with the laws of the foreign-sending country" has been fulfilled for law "r 8 C.F.R. 204.3(b) for orphan cases" (USCIS RAIO, 2019). How would an AI be able to determine what a reasonable effort is for finding a lost parent? Especially when the standards for what's reasonable can be limited by the economic factors of each individual country. It would be nigh impossible and highly impractical to try to automate these types of tasks. Although it is hard to say with a high degree of certainty with such a cutting-edge technology, the jobs of the USCIS workers should still exist after the adoption of this automated system.

Conclusion

There are definitely many potential costs and benefits to consider before making such a big change. However, after I laid them out on this paper, I believe it is still worth taking the chance. There is a high chance of benefitting the millions of immigrants stuck in the backlog. The workers at the agency are likely to still keep their jobs at the agency, and focus their efforts on more complex cases. The only group which has a legitimate reason to fear the proposed system are the current American workers who would have to compete with the around 1 million new immigrant workers. However, if that is an issue, they should debate in the public square and convince others to vote for what they believe should happen, as is the American ideal. The tardiness of a government agency should not be a weapon to get what they want. The only weapon fit for that is the ballot box. The people in the backlog have not been legally approved nor denied. They are stuck in a bureaucratic limbo through no fault of their own.

References:

Atkinson, R. D. (2018). "It Is Going to Kill Us!" and Other Myths About the Future of Artificial
 Intelligence. IUP Journal of Computer Sciences, 12(4), 7-56. Retrieved April 7, 2023
 from https://ssrn.com/abstract=3066182

Borjas, George J. (2016). Yes, immigration hurts American workers. POLITICO Magazine.
 Retrieved October 9, 2022, from
 https://www.politico.com/magazine/story/2016/09/trump-clinton-immigration-economy-
 unemployment-jobs-214216/

Gray, Gordon. (2022). The Budgetary and Economic Tradeoffs of Reducing the Immigration
 Backlog. Retrieved April 5, 2023 from
 https://www.americanactionforum.org/research/the-budgetary-and-economic-tradeoffs-
 of-reducing-the-immigration-backlog/

Grossman, Sandra and Keeler, Christopher. Updated by Chen, Rex and Goettel, Kate M. Delay.
 (2022). Actions In The Asylum Context: Avoiding Dismissal And Proving The Case.
 American Immigration Council, Retrieved April 6, 2023 from
 https://www.americanimmigrationcouncil.org/practice_advisory/delay-actions-asylum-
 context-avoiding-dismissal-and-proving-case

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds &
 Machines 30, 99–120. https://doi.org/10.1007/s11023-020-09517-8

Khosla, Pooja. (2018). Wait time for permanent residency and the retention of immigrant
 doctoral recipients in the U.S., https://doi.org/10.1016/j.eap.2017.11.002.

Loukides, Mike (2017). The machine learning paradox. Retrieved April 5, 2023 from

    https://www.oreilly.com/radar/the-machine-learning-paradox/

Manyika, James and Sneader, Kevin. (2018). AI, automation, and the future of work: Ten things

    to solve for. McKinsey & Company. Retrieved April 7, 2023, from

    https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-

    future-of-work-ten-things-to-solve-for

Martens, David and Provost, Foster. (2011, Sep 26). Pseudo-social network targeting from

    consumer transaction data. Retrieved February 14, 2023, from

    http://hdl.handle.net/2451/31253

Martinelli, Marissa. (2019, Sep 16). John Oliver Breaks Down the Four Major Paths to Legal

    U.S. Immigration, Including His Own. Retrieved February 13, 2023 from

    https://slate.com/culture/2019/09/john-oliver-immigration-trump-last-week-tonight.html

Martino, Daniel Di. (2023, Dec 15). Reducing the Immigration Backlog. Manhattan Institute for

    Policy Research, Inc. Retrieved February 15, 2023, from https://www.manhattan-

    institute.org/reducing-immigration-backlog#notes

Maurer, Roy. (2022, Aug 3). USCIS Must Address Ballooning Backlogs, Ombudsman Says.

    SHRM. Retrieved October 11, 2022, from https://www.shrm.org/resourcesandtools/hr-

    topics/talent-acquisition/pages/uscis-must-address-ballooning-backlogs-ombudsman-

    says.aspx

Miller, S. (2019). Machine Learning, Ethics and Law. Australasian Journal of Information

    Systems, 23 https://doi.org/10.3127/ajis.v23i0.1893

Narea, Nicole. (2022). Immigrants could help the US labor shortage — if the government would

    let them. Vox Media. https://www.vox.com/policy-and-politics/22933223/work-permit-

    uscis-backlog-immigration-labor-shortage

Samuel, A.L. (1959) Some Studies in Machine Learning Using the Game of Checkers. IBM

    Journal of Research and Development. vol. 3, no. 3, pp. 210-229.

    https://doi.org/10.1147/rd.33.0210.

Sherman, Arloc et al. (2019). Immigrants Contribute Greatly to U.S. Economy, Despite

    Administration's "Public Charge" Rule Rationale. Center on Budget and Policy

    Priorities. Retrieved April 6, 2023 from https://www.cbpp.org/research/poverty-and-

    inequality/immigrants-contribute-greatly-to-us-economy-despite-administrations

USCIS. (2021) Questions and Answers: Appeals and Motions. Retrieved October 8, from

    https://www.uscis.gov/forms/all-forms/questions-and-answers-appeals-and-motions

USCIS. (2019) RAIO DIRECTORATE – OFFICER TRAINING: RAIO Combined Training

    Program DECISION MAKING - TRAINING MODULE. Retrieved October 8, from

    https://www.uscis.gov/sites/default/files/document/foia/Decision_Making_LP_RAIO.pdf

Vespa, Jonathan. (2019). The U.S. Joins Other Countries With Large Aging Populations. United

    States Census Bureau. Retrieved April 6, 2023.

    https://www.census.gov/library/stories/2018/03/graying-america.html