Wind Farm Cybersecurity Through Demilitarized Zones

(Technical project)

Discrimination in the U.S. Criminal Justice System from Recidivism Score Algorithms

(STS project)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Grace Kisly

November 1, 2022

Technical Team Members: N/A

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.

ADVISORS

MC Forelle, Department of Engineering and Society

Raymond Petit, Department of Computer Science

Introduction:

The United States has the highest incarceration rate in the world, with incarceration rates significantly higher for blacks and Latinos than for whites (Population Reference Bureau). While the implementation of artificial intelligence strives to reduce crime by identifying suspects more efficiently and to remove prejudice by providing objective insight on criminal data, its application has had adverse effects. By using historical data, minority groups that have higher rates of incarceration are more likely to be identified as a criminal, perpetuating the disproportionate rates of recidivism (Nicol, & Lai, 2022). The criminal risk assessment tool, Correctional Offender Management Profiling for Alternative Sactions (COMPAS), has been found in various statistical analyses to be discriminatory toward people of color and women (Dressel & Farid, 2018 and Hamilton, 2019). Alarmingly, algorithms like COMPAS are highly popular, and in 2021 they were used in 46 states in the U.S. (Mesa, 2021). Utilizing artificial intelligence to score individuals on their risk of misconduct can be extremely detrimental, as erroneous assessments of defendants can cause undeserving suspects to face harsher sentences.

While the goal of risk assessment tools is to alleviate inconsistency and inaccuracy in judicial decisions, in practice this is not the case. In comparison, wind turbines serve as a technological solution to environmentally damaging energy production; however, the incorporation of wind farms onto the US electric grid introduces new vulnerabilities to energy security. Researchers have discovered that wind turbines across the US are relatively easy to attack, potentially enabling electricity disruptions if cybersecurity is not integrated (Greenberg, 2017). I explored the use of demilitarized zones as a network security approach for wind farms, where a demilitarized zone is “a network area (a subnetwork) that sits between an internal network and an external network” (Cybersecurity & Infrastructure Security Agency). This would prevent illegal traffic from entering the network, serving as a potential solution to cyberattacks.

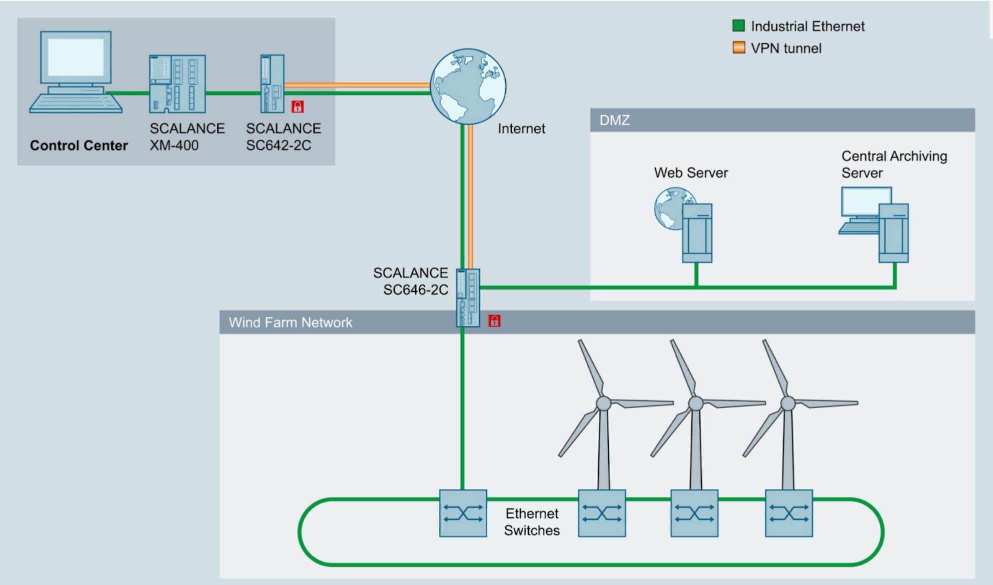
Efforts to improve the major issues within the criminal justice system and energy system introduce new, complex problems that must be resolved. It is the members in these sectors' duty to keep these systems secure and equitable. In terms of wind turbines, the full network required to operate this artifact must be secured, monitored, updated, and tested. Cybersecurity is not an addition, but an integral part of the development and deployment of the system, accounting for the whole lifecycle of the device (Van den Brink, 2022). Connecting a new technology into a system should require scrutiny on possible harmful and unexpected paths of use, especially for infrastructure that vast amounts of people rely on and are affected by its failure. Similarly, the ramifications of incorporating an automated scoring system for criminals needs to be thoroughly examined and tested prior to deployment. The COMPAS artificial intelligence technology has exacerbated existing discrimination, and without robust regulation could lead to its use in oppression, exploitation, and the increased discrimination of vulnerable groups (Altun & Humble, 2020). Although these technologies may seem like a perfect solution in isolation, once incorporated into practice unexpected issues arise, and its use may overcomplicate what it was attempting to reform, having potentially dire social repercussions. For wind turbines, I seek to explore technological solutions that secure wind farm networks from hacker-induced disruptions, as the entire energy system must be accounted for and protected. For protecting the criminal justice system from bias, I will evaluate if algorithmic risk assessments cause more harm than benefits, examining possible alternative paths that would achieve their intended purpose.

Technical project:

Wind turbines have often been constructed with the focus on output reliability, and fail to account for cybersecurity, leaving them vulnerable to attacks (KK Wind Solutions). To understand the risk wind farms face, I examined a case study from researchers at the University of Tulsa who performed penetration tests on five different US wind farms across the Midwest and West Coast that use the hardware of five different wind power equipment manufacturers (Greenberg, 2017). With the permission of the wind energy companies, his team had been systematically hacking the farms to expose vulnerabilities among the increasingly popular form of energy production. His team accessed the turbine internals physically, as they often stood unprotected in open fields, and affixed $45 worth of computing equipment in order to launch their attack (Greenberg, 2017). The hackers had discovered all they needed was to break a lock to access the server closet and gain control of the computer that controlled the turbine, indicative of the lack of physical security as well as network authentication. Via an open port, they could send valid commands to the entire network of turbines, inserting three types of malware to disable turbines, jam on the break to damage them, and control them while undetected (Greenberg, 2017). Although the team would only shut off a single turbine at a time, their methods could easily dismantle an entire wind farm, shutting down as much as hundreds of megawatts of power. Causing expensive downtime leaves wind energy companies open to extortion and ransomware attacks, and these vulnerabilities must be addressed as we shift toward renewable energy (Teirstein, 2021).

Equipped with the knowledge of wind farm vulnerabilities, I strived to research the most effective ways to improve the current cybersecurity issues within wind turbines to make these systems more reliable. Wind farm operators need to employ authentication into the internal communications of their control system, as it has been proven that just isolating the systems from the internet does not suffice (Seals, 2019). In terms of the software solution I investigated, I sought to learn if demilitarized zones (DMZs) could be applied to wind farm networks. This adds a layer of security as connections from the DMZ are only permitted to the external network, and hosts cannot connect to the internal network, restricting access to sensitive servers and blocking untrusted traffic (Cybersecurity & Infrastructure Security Agency). Energy companies are developing technology to apply DMZs to wind farms, where firewalls regulate external access to a wind farm network and the DMZ prevents direct access from an external network (Siemens, 2022). Below is a diagram of the application of Siemens' SCALANCE S Industrial Security Appliance implementing a DMZ (Siemens, 2022).

Figure 1



Note: Diagram of a secure wind farm network with SCALANCE S technology. Adapted from Holistic Protection of Wind Turbines, In *Siemens*, Retrieved from https://new.siemens.com /global/en/markets/wind/equipment/security.html

My research project highlighted how the technological solution of demilitarized zones can alleviate security concerns with wind turbines, as at present there is a lack of security practices and standards in this realm of the energy sector (Department of Energy, 2020). I evaluated whether DMZs would provide sufficient protection for wind farms, and if this solution would be capable of widespread implementation. With the University of Tulsa case study cyberattack demonstrating how severe this problem can be, it is imperative to improve the cybersecurity of the control systems before a large fraction of Americans become reliant on wind energy. This connects to the overarching idea that technical actors must be cognizant of all potential consequences prior to inserting new technology into a system, as it could drastically impact the humans interacting with it.

STS topic:

The utilization of artificial intelligence to assess the risk criminals pose to society illustrates how we derive meaning from the output of an algorithm, as real humans' livelihood can depend on the score they receive. The results can persuade officials to sentence individuals and assign stricter or more lenient probation and parole requirements (Chohlas-Wood, 2020). Unless this system is perfect, the outcome of the assessment can propagate false information and uphold biased rulings. Along with the lack of transparency in the design of the algorithm and the bias, the COMPAS algorithm for predicting recidivism has been argued to be neither fair nor accurate enough for its applications (Dressel & Farid, 2018).

One statistical analysis from the investigative journalism site ProPublica calculated that COMPAS is only accurate for two out of every three cases, and black defendants were often predicted to be at a higher risk than they actually were while white defendants were often predicted to be less risky than they were. (Angwin et al, 2016). The data encompassed more than 7000 individuals arrested in Broward County, Florida between 2013 and 2014 who had been scored on the COMPAS general recidivism risk scale. Another piece published by Melissa Hamilton in the Behavioral Sciences & the Law academic journal found that the COMPAS risk assessment tool is sexist, utilizing the same aforementioned dataset, as the results systematically over-classify women in higher risk groupings (2019). By over-classifying women, they are more likely to be unfairly treated in criminal justice decisions, perpetuating bias throughout the system

(Hamilton, 2019). Further, researchers Julia Dressel and Hany Farid at Dartmouth College issued an article for the Science Advances scientific journal published by the American Association for the Advancement of Science advocating against the use of COMPAS in courts throughout the U.S., as it is ineffective at being more accurate and unbiased than humans (2018). They conducted multiple studies to support this, one proving that the algorithm is no more accurate or fair than predictions made by people with little or no criminal justice expertise, and another that a simple linear predictor provided with only two features is nearly equivalent to COMPAS with its 137 features (Dressel & Farid, 2018).

From the lack of transparency on how the algorithm works and how much of a role it plays in influencing court decisions, to the plethora of sources disproving its legitimacy in terms of accuracy and fairness, the COMPAS algorithm infringes on the human rights clauses denouncing discrimination (Roa Avella et al., 2022). This racial and gender bias does not only exist in the case studies, its tangible applications permeate the US court and prison systems with its detrimental consequences. The implementation of this technology reshapes the network of actors involved in the criminal justice system, impacting criminals and redefining the roles of judiciaries alike. Evaluating how the introduction of a new actor into a system connects to other human and non-human actors is a valuable lens to determine whether a technology is just in its use. Michael Callon, Madeleine Akrich, Bruno Latour, and John Law's Actor-Network Theory (ANT) will be employed to understand how the use of COMPAS risk assessment critically impacts a variety of actors in the criminal justice network. ANT is summarized in Darryl Cressman's overview as "reducible neither to an actor nor to a network… An actor-network is simultaneously an actor whose activity is networking heterogeneous elements and a network that is able to redefine and transform what it is made of" (Callon 1987, p. 93, as cited in Cressman, 2009, p. 3). This view of the recidivism scoring algorithm serving as both an actor and network will not only highlight how small biases can manifest vastly throughout the criminal justice process, but also propose how systemic change is necessary to approach more ethical artificial intelligence use. Instead of solely scrutinizing the data quality and statistical models to reduce discrimination, one must examine the entire context of the organizational structures that form the ways where discriminatory practices may or may not be produced (Schwarting & Ulbricht, 2022). This way, technologists can mitigate biases and protect historically marginalized groups to move toward equity in the digital age (Nicol & Lai, 2022).

Methods:

The question I strive to answer is- How can the bias in the COMPAS risk assessment algorithm be mitigated to reduce the perpetuation of discrimination in the criminal justice system? I want to understand what systemic change must occur in order to facilitate the ethical use of recidivism scores in the criminal justice system, and if it is even plausible. To best approach researching this, I will examine existing literature through published papers as well as research articles in technology journalism. An important area to examine first would be to get a broad understanding of various discriminatory practices of artificial intelligence, then specifically hone in on COMPAS recidivism score applications in the criminal justice system. It is critical to establish a solid background of the ethics of AI and to grasp how harmful the consequences of its use can be. Delving into articles from the past 5 years in tech journalism will also reveal the current voices discussing this and who is advocating for change. Discourse analysis will be performed to better understand which scientific, technological, and investigative journals are covering automated recidivism scoring, as well as to better comprehend the motives behind writing their articles. A few statistical case studies on the data from COMPAS assessments in Broward County, Florida will be incorporated, illustrating the current racist and sexist effects produced by the tool's use. This will highlight the urgency necessary for addressing this issue and justify the problem's importance. The analysis of this research will also include what has already been explored as agents of change, whether it be better technological solutions, policy changes, diversified data, or shifts in the role of risk assessments, or a combination of solutions.

Conclusion:

Hopefully sparking enough discussion about the detriments of AI will enact change and create a movement toward resolving its discriminatory applications. It must be enforced that potential bias is carefully considered throughout the process of designing, building, and implementing risk assessment software. Introducing new technology to a system, particularly one already ridden with issues, requires thorough examination and regulation that currently is not present. This can also be seen with wind turbines, as wind farms require more stringent cybersecurity standards to protect the energy systems providing power to millions. I foresee that introducing network security in the form of demilitarized zones will be an adequate solution to incorporating protection from hackers. For more equitable recidivism scoring, I anticipate that the changes necessary will encompass strategies such as opening dialogue about national AI governance, outlining cases in need of oversight, incorporating anti-racist principles into every aspect of the design process, increasing diversity in design teams and in the data itself, enacting policy for regulation, and decoupling the recidivism score from the legal reasoning to emphasize the professional decision autonomy of the judges (Nicol & Lai, 2022, and Schwarting & Ulbricht, 2022). If there are too many complexities and tensions between political systems and risk assessment tools, then it would be illogical to rely on these tools as a means of criminal justice (Solow-Niederman et al., 2019). Minimizing bias in the system, particularly in increasingly prevalent artificial intelligence technologies, is the only way to drive criminal justice reform.

References:

Altun, D. & Humble, K. P. (2020). Artificial intelligence and the threat to human rights. *Journal of Internet Law*, 24(3), 1-18. Retrieved from https://gala.gre.ac.uk/id/eprint/30040/

Angwin, J., Larson, J., Kirchner, L., & Mattu, S. (2016, May 23). Machine bias risk assessments in criminal sentencing. *ProPublica.* Retrieved from https://www.propublica.org/article /ma chine-bias-risk-assessments-in-criminal-sentencing

Brink, H. van den. (2022, July 6). Hacking wind turbines- explained*.* *Medium*. Retrieved from https://harmvandenbrink.medium.com/hacking-wind-turbines-explained-230997db62f6

Cressman, D. (2009, April). A brief overview of Actor-Network Theory: punctualization, heterogeneous engineering & translation. *CT Lab/Centre for Policy Research on Science & Technology (CPROST)*, 1-17. Retrieved from https://summit.sfu.ca/item/13593

Chohlas-Wood, A. (2020, June 19). Understanding risk assessment instruments in criminal justice. *The Brookings Institution*. Retrieved from https://www.brookings.edu/research/ understanding-risk-assessment-instruments-in-criminal-justice/

Cybersecurity & Infrastructure Security Agency (CISA). (n.d.). Control system security DMZ. Retrieved from https://www.cisa.gov/uscert/ics/Control_System_Security_DMZ-Definiti on.html

Department of Energy. (2020, July). Roadmap for wind cybersecurity - energy. *Office of Energy Efficiency & Renewable Energy.* Retrieved from https://www.energy.gov/sites/prod/files/202 0/07/f76/wind-energy-cybersecurityroadmap-2020v2.pdf

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/

PMC57 77393/

Greenberg, A. (2017, June 28). Researchers found they could hack entire wind farms*.* *Wired.* Retrieved from https://www.wired.com/story/wind-turbine-hack/

Hamilton, M. (2019) The sexist algorithm. *Behav Sci Law,* 37(1), 145– 157. Retrieved from https://pubmed.ncbi.nlm.nih.gov/30931534/

KK Wind Solutions. (n.d.). How to increase your wind Farm's cyber security. *Turbine Condition Monitoring.* Retrieved from https://tcm.kkwindsolutions.com/how-to-increase-yourwind-far ms-cyber-security

Mesa, N. (2021, May 13.). Can the criminal justice system's artificial intelligence ever be truly fair? *Massive Science.* Retrieved from https://massivesci.com/articles/machine-learning-compas-racism-policing-fairness/

Nicol, T. L., & Lai, S. (2022). The U.S. can improve its AI governance strategy by addressing online biases. *The Brookings Institution.* Retrieved from https://www.brookings.edu/blog/tec htank/2022/05/17/the-u-s-can-improve-its-aigovernance-strategy-by-addressing-online-biases/

Population Reference Bureau. (n.d.). U.S. has world's highest incarceration rate. Retrieved from https://www.prb.org/resources/u-s-has-worlds-highest-incarceration-rate/

Roa Avella, M. del P., Sanabria-Moyano, JE, & Dinas-Hurtado, K. (2022). Use of the COMPAS algorithm in criminal proceedings and risks to human rights. *Brazilian Journal of Criminal Procedural Law* , 8 (1). Retrieved from https://www.scielo.br/j/rbdpp/a/6W 9b8CHYbXcsc6 qczDxCSfr/abstract/?lang=en

Schwarting, R., Ulbricht, L. (2022) Why organization matters in "algorithmic discrimination". *Köln Z Soziol* 74(Suppl 1), 307–330. Retrieved from https://ideas.repec.org/a/zbw/espost/261200.html

Seals, T. (2019, November 1). Solar, Wind Power Utility Disrupted in Rare Cyberattack. *Threatpost*. Retrieved from https://threatpost.com/solar-wind-power-utility-cyberattack /149816/

Siemens. (2022). Holistic protection of wind turbines. *Network Security.* Retrieved from https://new.siemens.com/global/en/markets/wind/equipment/security.html

Solow-Niederman, A., YooJung Choi, & Van den Broeck, G. (2019). The institutional life of algorithmic risk assessment. *Berkeley Technology Law Journal*, 34(3), 705–744. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3405716

Teirstein, Z. (2021, May 10). Hackers found America's energy weak spot. *Grist Magazine*. Retrieved from https://grist.org/politics/hackers-found-americas-energy-weak-spot/