

An Interdisciplinary Approach to the Impact of Algorithmic Bias on Cybercrime Detection

Madison R. Baughman

Department of Interdisciplinary Studies, Old Dominion University

IDS 300W: Interdisciplinary Theory and Concepts

Dr. Kat LaFever

November 26, 2025

Introduction

When criminals weaponize digital tools to their benefit, criminologists and law enforcement need new methods of detecting and predicting these crimes to stay one step ahead. The solution to this challenge is artificial intelligence and crime prediction software. While this appears to be the best solution to a rising demand, it comes with a complication. The complication of algorithmic bias. To improve the efficacy of these programs, a question must be answered. What is algorithmic bias, how does it impact cybercrime detection, and how can it be avoided? A question as complicated as this cannot be fully addressed through conventional research approaches. Utilizing an interdisciplinary lens allows a more comprehensive and collaborative investigation. Without this approach, certain perspectives would be ignored and only a partial explanation would be reached. An interdisciplinary approach demands that we identify relevant disciplines. To fully examine the scope of algorithmic bias and cybercrime detection, the disciplines of data science, machine learning, and criminology are employed. The discipline of data science is necessary to investigate because of the large impacts it can have if the data produced is bad. If an AI machine is trained off of unprocessed or biased data, the likelihood of algorithmic bias is far higher. Machine learning must be included because this discipline is responsible for training the machines used in cybercrime detection. Human error in the coding process is a massive factor in the outcome of the machine's efficacy. Lastly, we must include criminology because this is the discipline that utilizes these tools. Crime detection algorithms are produced with this discipline and usage in mind, and therefore influences the impacts of algorithmic bias. Examination of this research question and discovery of the answer has the potential to encourage deeper consideration of the responsibility humans hold in creating and utilizing artificial intelligence. If used or developed irresponsibly, the consequences can be

massive. This discussion and discourse is especially relevant to my major and minor in cybersecurity and cybercrime, respectively. My vested interest in these fields demand I explore the implications of this topic.

Algorithmic bias, *artificial intelligence (AI)* and *impact parity* are among the key terms used in this paper. Shin and Shin (2023) define *algorithmic bias* as “referring to systematic and structured errors in an artificial intelligence system that generate unfair results and inequalities.” *Artificial intelligence (AI)*, as defined by Sheikh et al. (2023), “is a technology that enables machines to imitate various complex human skills.” *Impact parity* is defined as a requirement of machine learning algorithms to avoid disparity in minorities or sensitive groups to ensure fairness by Fu et al. (2022).

Data Science

Before any computing, research, or analysis can occur, data must be cleaned and processed so that is usable. This is critical to the production of good, usable data. Were these steps to be skipped, data would contain holes, errors, include corrupted data and likely be unreadable. Without the proper sensitivity to potential discrepancies in data, biases within data can go unnoticed and when sent out into the world to be used, bias leaks into other domains. “If we cannot guarantee we have fully addressed bias in the data pipeline, the overall harm in a system is a product of the interactions between the data and our model design choices,” (Hooker, 2021). It is a desire to protect the ‘system’ from harm that encourages scrutiny of data processes. A large part of this protection involves ensuring data is inclusive and varied, representing a large data pool. This is especially important when we consider the implications with algorithmic bias. If, for example, our data set processed in a machine learning algorithm contains a small demographic of crime, the machine will not operate accurately and will continue to predict that

small demographic. “If the data analyzed by algorithms or used to train ML do not reflect the various parameters of users correctly, the results would be biased,” (Shin & Shin, 2023). It is exactly this bias that we are trying to avoid. To safely and responsibly utilize artificial intelligence without the threat of algorithmic bias, it all starts with the data used. If we can’t trust the data, we can’t trust the machine.

Machine Learning

In machine learning, it’s important to consider the ethical implications of biased algorithms in decision making. Algorithmic bias can lead to unfair outcomes in many areas such as hiring, lending, and law enforcement. Scraping the internet of open source data and using it to train a model can seem like a good idea on the surface. However, borrowing unprocessed data, for example from a social media forum like Reddit or Twitter, can lead to bias in a majorly unethical way. “It is possible that such biases emerge from the operation of seemingly innocuous code paired with statistical regularities of the training data,” (Johnson, 2021). Johnson makes a great point here. Let’s suppose we have an AI model trained on something like crime or political data that has a 90% accuracy rating for producing factual unbiased data predictions. This sounds good on the surface, however having a 90% accuracy rating will lead to reinforcing the 10% of incorrect and biased data predictions making it harder to retrain the model without starting from scratch. This creates a feedback loop where the model's errors are perpetuated, potentially amplifying biases and spreading misinformation. As a result, ethical considerations and rigorous testing are essential to mitigate these risks and ensure AI systems promote fairness and accuracy. The level of fairness can also be a hindrance in certain situations as Fu points out. “The proposed fair ML algorithms that require impact parity, while conceptually appealing, can make everyone worse off, including the very class they aim to protect,” (Fu et. al, 2021). For instance if someone

of a protected class or race were to apply for a loan, and an AI model was used to sort who was eligible, it could artificially boost some people over others. Furthermore leading to unfair treatment and possibly a net negative for the business offering the loan. Overall in terms of machine learning and producing AI for public use, it is important to understand the context of data and when to apply AI responsibly, ensuring it aligns with ethical principles and does not inadvertently cause harm. This requires careful consideration of data sources, model design, and potential societal impacts to avoid perpetuating inequities or unintended consequences.

Criminology

To study criminology is to study its prevention, and find patterns to behavior as well as make predictions. Faster technology has demanded that we speed up as well, and has forced us to adapt and use new methods of prediction. “The proliferation of the Internet and mobile devices led to the digitization of society, which encouraged the digitization of crime,” (Leukfeldt & Holt, 2019). The development of cybercrime has been widespread and in many cases devastating due to its unpredictability and challenges with deterrence. This has forced criminologists to consider alternate ways of approaching this topic so as to emerge one step ahead of offenders. For example, in many areas of the country, law enforcement is utilizing AI algorithms to analyze crime data and predict where to send their officers to boost department efficiency. Let’s consider the data being input to these programs. Type of offense, location, demographics of offender, perhaps repeated offenses too. All of this data together can provide a comprehensive look into the crime distribution of a given precinct. While perceived as a roadmap to catching criminals in the act, if not handled responsibly, this data can be a roadmap to assumptions and bias. As previously discussed, biased data has the potential to harm systems. “Data are means of communication, knowledge and entertainment, as well as resources for analytic and criminal

manipulation and community mobilization,” (Smith et. al, 2017). Any field of study relies heavily on data to formulate theories and make new discoveries. However, it is imperative for criminology, when faced with new challenges in the world of crime, to ensure proper and responsible use of data as well as programs formulated off of this data. The potential for harmful outcomes from this usage of AI is not simply encouraging a gambling habit or passing a test you didn’t study for. Criminological use of artificial intelligence influences people’s lives and livelihoods. The data predictions produced are not to be taken lightly or at face value.

Common Ground

Within this interdisciplinary research, several discoveries were made through comparison of the disciplines involved. Although within their own separate realms and containing their own theories, these disciplines were discovered to have common ground amongst themselves. The first and most prominent area of common ground is between all three disciplines. Data science, machine learning, and criminology all agree to some degree that problems with algorithmic bias primarily tie back to the data used (Hooker, 2021; Johnson, 2020; Shin & Shin, 2023; Smith et al. 2017). While the degree in which data is to blame as well as other contributing factors varies, it is widely agreed upon that data is one of the most influential factors. Another area of common ground is the demonstrated need for more diverse data to produce AI machines better equipped to handle today’s challenging distribution of crime (Smith et al. 2017; Fu et al. 2022). These points of agreement from different disciplines eases the process of integrating them and developing a deeper understanding. To find common ground is to make connections and uncover underlying insights on the topic of research.

Disciplinary Conflicts

While many instances of common ground could be identified, there still stood to be a particular insight the involved disciplines could not agree upon. In each discipline there is discussion, mention and definition of the term bias. Criminology defines bias in terms of stereotypes and prejudice that influences criminal justice whereas data science defines it in terms of an error that causes inaccurate results. To mediate this conflict in an interdisciplinary context, it is necessary to compare and contrast definitions in search of a common understanding. By taking specific insights about the term bias and compiling them into a singular comprehensive definition, conflict can be overcome and the different disciplines appeased. In this sense, it may also be pertinent to create an entirely new definition for the purposes of an interdisciplinary approach. However, a new definition is not without input from previous definitions.

Ch. 12: “Constructing a More Comprehensive Understanding or Theory”

It is necessary, when utilizing an interdisciplinary approach to research, that a more comprehensive understanding of the information received is developed. There are several applicable strategies to this goal, but perhaps the most effective is that of sequential integration. Sequential integration for this specific research question involves a cause and effect reaction and order to an integration of disciplines. When we view data science, machine learning, and criminology in this way, we see a flow chart start to evolve and a visualization of the flow of data from one stage to the next. First processing in data science, then training and analysis in machine learning, and finally data is applied to real world threats in criminology. “The theories are complementary, but each captures only a piece of the broader puzzle,” (Repko & Szostak, 2021, p.336). To construct a more comprehensive understanding of the topic of algorithmic bias, we follow the lifecycle and transformation of data through the disciplines. In this way, we can

identify certain hazards that are more likely to contribute to the development of algorithmic bias. For example, incomplete processing and cleaning in data science, human error in the coding process in machine learning, or improper usage in practice in criminology. Visualizing the flow and changes this data goes through is the most effective way to prevent an undesired outcome. Through this visualization, we are able to develop a more comprehensive understanding of this specific life cycle of data, and be more sensitive to the threat of algorithmic bias.

Ch. 13: “Reflecting On, Testing, and Communicating the Understanding or Theory”

There are many ways to reflect on one’s understanding and many aspects of the understanding to reflect on. Throughout the process of researching this topic, there was intellectual growth and broadening of insights into new and previously untouched disciplines. While knowledge and insight was gained in this process, it’s only logical to assume that some bias was gained as well. As hard as we try to scour it from academic investigations, some bias is bound to surface. I can admit that throughout this report of my findings, I have depicted artificial intelligence as scary, dangerous, and a tool that trends toward negative outcomes. However, AI has many benefits and has contributed to many successes in our time. While this tool does have the potential to cause harm, it also has great potential to promote growth and advancement in our society. The best testing of an understanding happens when you circle back to where the work began, at the question. Afterall, isn’t that why we need understanding in the first place? To construct an answer? To be able to answer the research question is to have a solid and successful understanding, capable of being applied to real world problems. “Testing the more comprehensive understanding or theory is quite different from reflecting on it, though arguably testing may involve reflection” (Repko & Szostak, 2021, p.367). In future research, next steps may be considered by including a legal discipline to examine how the use of artificial

intelligence holds up in court against accusations of algorithmic bias. This would further expand the understanding of the topic and how the world perceives it. An overall narrative communication of the aforementioned understanding was effective, detailing specific research and insights into the research methods used.

Conclusion

Throughout this interdisciplinary approach to research, many insightful and deeper connections were made between disciplines involved. In a world where cybercriminals are evolving to use new techniques and methods of offending, it's necessary for law enforcement and criminologists to utilize tools such as artificial intelligence to combat them. While these tools offer many benefits, there is a potential for harmful algorithmic bias. Algorithmic biases are errors in a machine learning algorithm that produces unfair results. As discovered through interdisciplinary research, this bias impacts cybercrime detection by creating feedback loops, predicting high-crime areas incorrectly, and furthering the proliferation of bias. To avoid these negative outcomes, certain steps have to be taken in the data lifecycle. Starting with proper and thorough processing of data and continuing to responsible use for training with artificial intelligence. Finally, the end-product use by criminologists must be monitored for unintended purposes. This investigation has produced a deeper consideration of human responsibility in the development and usage of machine learning algorithms. Understanding of the consequences as well as the necessary protocols to prevent negative outcomes is imperative to retain. Further communication of these findings has the potential to create change and encourage closer monitoring in the use of artificial intelligence in cybercrime detection. Although we must rise to meet the demand of emerging technology, we cannot do so irresponsibly.

References

- Fu, R., Aseri, M., Singh, P. V., & Srinivasan, K. (2022). “Un”fair machine learning algorithms. *Management Science*, 68(6), 4173–4195. <https://doi.org/10.1287/mnsc.2021.4065>
- Hooker, S. (2021). Moving beyond “algorithmic bias is a data problem.” *Patterns*, 2(4), 100241. <https://doi.org/10.1016/j.patter.2021.100241>
- Johnson, G. M. (2021). Algorithmic bias: On the implicit biases of social technology. *Synthese*, 198(10), 9941–9961. <https://www.jstor.org/stable/48692581>
- Leukfeldt, R., & Holt, T. J. (Eds.). (2019). *The Human Factor of Cybercrime* (1st ed.). Routledge. <https://doi.org/10.4324/9780429460593>
- Repko, A. F. & Szostak, R. (2021). *Interdisciplinary research: Process and theory* (4th ed.). SAGE Publications, Inc.
- Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial intelligence: Definition and background. In *Mission AI* (pp. 15–41). Springer, Cham. https://doi.org/10.1007/978-3-031-21448-6_2
- Shin, D., & Shin, E. Y. (2023). Data’s impact on algorithmic bias. *Computer*, 56(6), 90–94. <https://doi.org/10.1109/MC.2023.3262909>
- Smith, G. J. D., Moses, L. B., & Chan, J. (2017). The challenges of doing big data and criminology: The datafication of criminal justice. *British Journal of Criminology*, 57(2), 259–274. <https://doi.org/10.1093/bjc/azw096>