Andrew Thompson

20 November 2023

FYS 112-H03

Dr. Petrosillo

Word count: 1,617

Data Analysis, Artificial Intelligence, and Ethics in Crime Prevention

Steven Spielberg's *Minority Report* introduces a futuristic world where crimes can be predicted before they happen; unique ethical issues become evident. Modern crime prevention has reached new horizons by implementing a newly developed artificial intelligence algorithm that can predict crime up to a week before it happens. Because of the parallelism between the 2002 science-fiction film *Minority Report* and existing technology, new ethical considerations have risen in the world of criminology. It is important to immediately and effectively consider ethical implications, as crime prediction is no longer fictitious. Constant data collection used to analyze and predict the criminal behavior of U.S. citizens necessitates data ethics, particularly in crime prevention and behavioral analysis. As daily targets of data collection used to analyze and predict criminal behavior, it is essential that ethical data standards are introduced in crime prevention and behavioral analysis.

The initial element of preventing crime with predictive analysis by collecting data has both desirable and undesirable applications. While many are unaware, data is constantly gathered; the way people live, shopping habits, browser searches, and what people buy are some examples of stored information. Companies like Amazon, Microsoft, and even the government have database warehouses full of infrastructure to store data, because mass amounts of data are created daily. Often stored for purposes like marketing, improved services, and estimation

models, data collection is not inherently harmful. In fact, intentions with data collection are generally good. Despite good intentions, data collection can also reduce privacy. One example, The PATRIOT Act, allows law enforcement to collect individual data and even tap phone lines because "data can be collected under the authority of a [government ordered] FISA request," which would require another corporation to hand over information (Squitieri 2019). The general public does not appreciate when their data can be so easily passed around, especially without much more need than a simple request. These less desirable privacy infringements prove that data collection is not always what the public perceives as beneficial. Therefore, it is important to be knowledgeable about the different applications of data collection and to consider the pros and cons of each to define the future ethical framework that data science should follow.

Contrasting old and new statistical analysis applications in crime prevention shows how statistical analysis is used today. According to criminology statistician James Alan Fox, one area for improvement of old data analysis applications in crime was the need for more precision in explaining crime rate trends. He claims that using line segment predictive models makes research less superfluous (9). Two lines would be used to predict trends instead of a curve because curves can bend in any random direction without notice. While line-segment prediction may have been an excellent method in 1975, modern data science is inherently more accurate, considering almost 50 years of improvements in statistical analysis since 1975. In fact, the improvements lead to shocking prediction capabilities. A new artificial intelligence algorithm aids in making highly accurate predictions via data analysis. As a result of modern data analysis and artificial intelligence, this algorithm can predict crimes up to a week in advance (Wood). These improvements, giving the ability to predict future behavior, have many implications, such as the ability to prevent crime. This algorithm was 90% accurate on a receiver operating characteristic

curve; crimes were accurately predicted "within ~1,000 ft" of where they would later occur (Rotaru et al. "Abstract"). A receiver operating characteristic curve compares true-positive and false-positive rates. In other words, when the algorithm predicted a crime, there was a 90% probability that it did occur. This shows that data analysis capabilities via artificial intelligence introduce an advanced age of data science.

The enhanced capabilities of AI in predicting crime create data science scenarios that raise ethical questions. Artificial intelligence's ability to outperform humans in some situations raises significant questions. Should we continue to rely on humans, or should we trust an algorithm to make certain life-impacting decisions? For example, in the courtroom, it must be determined whether an artificial intelligence algorithm—or a less precise human—decides if a defendant is released on bail. 554,689 defendants were sorted by a judge and an artificial intelligence algorithm that determined 400,000 to be offered bail. Once the two lists were compared, there was no close competition. The defendants on the A.I. generated list were 25% less likely to commit a crime while waiting for their trial (Gladwell 39). This brings up a significant issue. How do we determine if using artificial intelligence and data analysis is ethical? In this situation, the A.I. was more correct than the judge, who had years of experience. Given its unique capabilities surpassing a human judge, would it be correct to allocate bail decisions to artificial intelligence? These applications of artificial intelligence have multiple implications. In another scenario brought up by Spielberg's *Minority Report*, if a murder is predicted, and prevented, is the criminal guilty of murder? Are they a criminal at all? How will the proper, ethical route of action be determined? This ethical issue presides because while trusting artificial intelligence makes quantitative sense in one situation, a reasonable level of trust for these algorithms must be defined.

Data is collected and used to make predictions, with or without consent, so promoting data ethics is essential. This is because the public will more likely approve of data science projects with a "clear public benefit" (Drew 5). This means that data analysis should be used for the common good and should be beneficial, bringing some sort of good about. It would be fallible to believe that data science projects are inherently good, when the intentions are not. Projects need public approval, which can only happen when ethics are accounted for.

To this point, two glaring ethical considerations for data scientists are the reasoning and transparency behind a data science project. Reasoning is one of the four significant considerations to make regarding data science ethics. If there is no reason to analyze the data, then the project is unnecessary, and data should not be collected. Without a "clear policy or operational need," data collection is unethical (Drew 5). This outlook encourages ethical data collection and discourages other applications. Because reasoning is the backbone for data analysis, reason is always important to consider when deciding if any analysis should be done on collected data. After that, transparency plays a role in ensuring that data science projects are held to ethical standards. In the realm of crime, data is available from multiple sources. Crime data sources include FBI tabulations, police expenditures, the Consumer Price Index, and others (Fox 5-6). It is crucial to address where a project is getting its information; knowing the source contributes to transparency. Understanding where individualized data comes from and who is sharing it with external sources destroys trust. Because it is unlikely that personal actions will limit data collection by organizations like the FBI, there is importance in enforcing data science ethics among these groups.

Security is the most prominent ethical concern among the average citizen, who is well-versed in shared, lost, and stolen data. The lack of security in some situations has made data

collection a source of unease. Governments and corporations put massive efforts into protecting their data, not only because of its monetary value but also because of the backlash experienced by companies that leak data. To overcome potential risks, infrastructure is constantly being developed to securely hold data (Drew 8). Ultimately, those who collect data should be held to a high standard, and the development of infrastructure shows that effort is being put in to preserve this security. Because proprietary data storage techniques are constantly developing, effective techniques must become the data collection and storage standard.

The trustworthiness factor is multifaceted because, whereas artificial intelligence can improve upon current skills in one situation, conversely, it can drag decision-making abilities down. Matt Wood demonstrates that using predictive algorithms to prevent crimes "requires more resources to arrest more people in response to crime in a wealthy area and draws police resources away from lower socioeconomic status areas." Specifically, while the algorithm is not magical, it predicts future crimes. The concession is that policing resources were stretched too thin trying to prevent these crimes. The department consequentially focused on protecting wealthier areas, leaving those with a low socioeconomic status less effective protection. Regarding trustworthiness, citizens must believe that data analysis applications promote their welfare and do not enforce biases. In this case, it is difficult to place trust in a policing system's predictive algorithm that is not being used in a way to prevent all crimes and is being used in a specific manner to target high-profile crimes in wealthier areas. Ethics play a prominent role in deciding if A.I. or mankind makes the most effective law enforcement decisions.

In his *Minority Report* world, John Anderton discovers that errors in crime prediction were covered up to maintain public faith in Pre-Crime. The antagonist attempts to remove Anderton from the equation to protect these secrets. This is the epitome of unethical behavior in

data analysis. It followed none of the previously defined standards. This cover-up behavior broke transparency because it was kept secret, security because even the Pre-Crime division did not know it was happening, trustworthiness because they were lying, and the reasoning for it was to manipulate others into trusting the crime prevention. Because real-world parallelism between the crime prediction abilities already exists, consideration of data collection and analysis ethics is essential. Artificial intelligence in the world can predict crime and behavior, but there are real-world examples of flawed decision-making by those with access to crime predictions algorithms. Parallelism is popping up. At what point is a real-world, dystopian resemblance to *Minority Report* too great?

Works Cited

Dihuni. "Every Day Big Data Statistics – 2.5 Quintillion Bytes of Data Created Daily." *Digital*

    Transformation Simplified, 10 Apr. 2020, www.dihuni.com/2020/04/10/every-day-big-

    data-statistics-2-5-quintillion-bytes-of-data-created-daily.

Drew, Cat. "Data Science Ethics in Government." *Philosophical Transactions: Mathematical,*

    Physical and Engineering Sciences, vol. 374, no. 2083, 28 Dec. 2016, pp. 1–11.

    www.jstor.org/stable/26115839. Accessed 12 Oct. 2023.

Fox, James Alan. *Forecasting Crime Data*. D.C. Heath and Company, 1978.

Gladwell, Malcolm. *Talking to Strangers: What We Should Know about the People We Don't*

    Know. Little, Brown and Company, 2019.

Minority Report. Directed by Steven Spielberg, 20th Century Studios, 2002.

Rotaru, V., Huang, Y., Li, T. *et al.* Event-level prediction of urban crime reveals a signature of

    enforcement bias in US cities. *Nat Hum Behav* 6, 1056–1068 (2022).

    doi.org/10.1038/s41562-022-01372-0. Accessed 24 Oct. 2023

Squitieri, Chad. "CONFRONTING BIG DATA: APPLYING THE CONFRONTATION

    CLAUSE TO GOVERNMENT DATA COLLECTION." *Virginia Law Review*, vol. 101,

    no. 7, 2015, pp. 2011–49. *JSTOR*, www.jstor.org/stable/24643632. Accessed 24 Oct.

    2023.

Wood, Matt. "Algorithm Predicts Crime a Week in Advance, but Reveals Bias in Police

    Response." *Biological Sciences Division | The University of Chicago*, 30 June 2022,

    biologicalsciences.uchicago.edu/news/algorithm-predicts-crime-police-bias.