John Fernandez (s3653471) 1/12/22
Algorithmic Security Politics
Presentation Paper
Dr. Ragazzi

The Impact of Algorithmic Security Systems on Fundamental Rights

Introduction

Despite the rapid advancement and adoption of algorithmic technologies across nearly all sectors of business and government, concerns regarding their intricacies and impacts have not been adequately addressed. As algorithmic processes become further intertwined into peoples' daily lives it becomes easy to forget about certain concerns in the face of revolutionary changes promoting ease of use, and seamless interaction with services individuals rely on. While the integration of AI technologies into existing processes promises great improvements, it is important to acknowledge the dark specter surrounding the aspects necessary for efficient execution of the technology, which are frequently concealed in part to detract attention from this fact (Pasquale, 2015, p. 6). Any AI process relies upon data, in whatever form it takes, as an input, however when applied to many corporate and particularly government practices, this data can be quite personal, sensitive, and otherwise unattainable. This inherently introduces grave concerns about the treatment of an individual's fundamental rights. Vast frameworks exist to protect the sanctity of fundamental rights outside of an AI environment, yet transgressions still widely occur, frequently forming the basis for lawsuits and other types of legal recourse worldwide. If the protection of fundamental rights cannot be reliably and consistently guaranteed through time-tested transparent processes, with established methods of accountability, the introduction of AI technology in the same sphere with its outright lack of transparency and accountability is a rightfully major consideration for those concerned with fundamental rights. The use of AI does not inherently violate fundamental rights, but it carries a strong potential to do so.

While the use of algorithmic systems can provoke discussions on fundamental rights in any context, their use in the field of law enforcement and criminal justice is a particularly sensitive matter as the practice, noted by Gonzalez Fuster, "touches upon core issues of the relation between the individual and the state" (2020, p. 8). As the fields of criminal justice and law enforcement frequently involve personal information, and can result in the suspension of one's freedom, rights exist to protect the citizen from government overreach, and ensure that such processes are conducted to as fair of a standard as possible. The inclusion of AI or algorithms adds complexity to this already sensitive process. Any use of AI takes decision-making out of the hands of humans to a certain extent. With the lack of transparency or 'black box' nature that many algorithmic systems have, it can frequently be impossible to gauge the level to which fundamental rights are being respected in any application of AI. Additionally, the use of AI in criminal justice or law enforcement applications can be an inherent violation of privacy rights, as such systems can find patterns and connections between data points to a much higher level than humans, and can therefore come to much more detailed, advanced, and accurate conclusions about the nature of specific individuals. The sensitivity of the data being handled, and the fundamental rights concerns that coincide, arise most prevalently in criminal judicial processes. In less personally sensitive matters, including those civil, commercial, and administrative, the use of AI raises significantly fewer rights concerns, and is generally regarded as beneficial for efficiency (Gonzalez Fuster, 2020, p. 45). This goes to show that the premise of AI can theoretically bring massive benefits, however when applied to criminal justice also sparks significant concerns.

Such a sensitive environment surrounding fundamental rights naturally begs the question: What is the impact of algorithmic security systems on fundamental rights? While a broad question, nearly every critical fundamental right can be impacted by the use of algorithmic security systems.

On a more well-known level, the issue of discrimination in AI systems is relatively well known. Algorithms, on a basic level, work to find patterns between existing data. If crime data, for example, happens to show a higher rate of crimes committed by a specific racial group, then an algorithm using that data may discriminate against members of that racial group, and be more likely to brand them as criminals solely based on race. Beyond discrimination however, AI can have much further, less initially obvious impacts on fundamental rights. The EU Charter of Fundamental Rights lists the right to a fair trial, and various other related rights such as the presumption of innocence, and the proportionality of crimes and punishments (European Union, 2012, p. 405). These rights can also be impacted through the use of algorithms, such as in instances of courts using recidivism calculations to determine sentencing. This paper argues that the use of AI could be made in theory to not violate fundamental rights, yet in practice it almost always does.

To more closely examine the impact of such algorithmic systems on fundamental rights, this paper starts by outlining two separate examples of rights affected by AI and concludes with an analysis of their impact as well as the core issue currently driving incompatibility between algorithms and fundamental rights.

Predictive Policing Software

Recent years have seen a trend towards "predictive policing" within communities, as police departments attempt to use the data at their disposal to increase their law enforcement capacity (Gonzalez Fuster, 2020, p. 22). Law enforcement agencies around the world are increasingly moving towards algorithmic systems due to various contributing factors, such as initiatives to limit police resources, an increase in the amount and complexity of data available, and a growing

perception that police ought to operate in a preventative rather than responsive manner (Gonzalez Fuster, 2020, p. 23).

One specific example of predictive policing software that has seen widespread use across departments in the United States is aptly named PredPol (Sankin et al., 2021). One of many privately made software packages currently available, PredPol sells its risk assessment algorithm to police departments advertising the ability to predict where and when crime will most likely occur (Sankin et al., 2021). The platform has seen great popularity, with "more than one in 33 U.S. residents … potentially subject to police patrol decisions directed by crime-prediction software" between 2018 and 2021 (Sankin et al., 2021).

While the exact functioning of the PredPol algorithm is unknown by design, researchers have been able to come to certain conclusions regarding its methods. Like other predictive policing software, the creators of PredPol maintain the 'black box' nature of their algorithm behind arguments of protecting trade secrets, which may be a legitimate interest, but prevents authorities from truly verifying the software does not violate rights (Chohlas-Wood, 2020). Independent researchers have however identified that rather than working to eliminate bias, PredPol's software perpetuates them (Sankin et al., 2021). A report conducted by researchers at Gizmodo found over 5.9 million PredPol crime predictions on an unsecured server, which already poses a major privacy issue, and analyzed them finding that the software consistently targeted Black and Latino areas (Sankin et al., 2021). The report came to various troubling conclusions about how the algorithm directed police towards predominantly minority areas, such as finding that neighborhoods in Michigan where PredPol recommended police patrols had nine times the proportion of black residents as the city average (Sankin et al., 2021).

Two immediately apparent fundamental rights issues appear out of the discussion surrounding PredPol. First, it is clear that racial discrimination is occurring, with the areas identified by PredPol having significant racial ties to minority groups, while areas with predominantly white citizens were frequently left untouched (Sankin et al., 2021). Second, the black box nature of the algorithm, combined with the fact that defense attorneys are not informed when crime prediction software lays the groundwork for an arrest, significantly limits an attorney's ability to provide their client with their right to a fair defense.

Furthermore, the efficacy of using algorithms to predict crime has frequently been called into question, as despite studies showing that targeted police patrols do work to address crime (Mohler et al., 2015), these findings have been criticized as being part of a self-generated feedback loop. Suresh Venkatasubramanian, a member of the board of directors of the American Civil Liberties Union (ACLU) in Utah, spoke of this feedback loop in regard to the PredPol software, stating, "Because this data is collected as a by-product of police activity, predictions made on the basis of patterns learned from this data do not pertain to future instances of crime on the whole … In this sense, predictive policing is aptly named: it is predicting future policing, not future crime" (Hicks, 2021). The existence of this feedback loop has also been observed by other researchers addressing the topic of predictive policing and finding on a broad level that if more police are dispatched to an area, a 'higher' rate of crime naturally follows (Reese, 2022). This is not to say that the true rates of crime actually rise, but that the presence of more police officers leads to a higher *reported* rate of crime.

Recidivism Risk Assessments

Another algorithmically driven tool commonly used in criminal justice is the recidivism risk assessment, frequently used by courts to help determine an individual's likelihood for reoffending, and therefore contributing to the determination of punishment (Chohlas-Wood, 2020). Algorithmic models in this field are highly desired, as they stand the chance of bringing a level of consistency and accuracy to judicial decisions which can frequently hold a subjective nature. In order for this to occur however, they must definitively not come into conflict with fundamental rights, yet they almost always do.

One particular example of recidivism risk assessment algorithms has been a system used in the United States called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) ("Criminal Law - Sentencing," 2016-2017, p. 1531). COMPAS was the focus of a legal battle in the Wisconsin Supreme Court, where an individual challenged his prison sentence on the grounds of the state's usage of the recidivism algorithm ("Criminal Law - Sentencing," 2016-2017, p. 1531). Defense attorneys claimed that the Court's use of COMPAS violated the defendant's right to due process, due to the algorithm's reliance on historical generalized data, which the defense claimed prevented individualized sentencing ("Criminal Law - Sentencing," 2016-2017, p. 1531). Furthermore, the defense argued that the use of gender in the algorithmic calculations constituted a further breach of due process ("Criminal Law - Sentencing," 2016-2017, p. 1532).

In a surprising decision that has since formed the basis for law school case studies ("State v. Loomis," n.d.) and a piece in the *Harvard Law Review* ("Criminal Law - Sentencing," 2016-2017), the Court upheld the sentence based on the algorithm, yet added certain stipulations confirming the dubious nature of algorithmic systems with regards to fundamental rights ("Criminal Law - Sentencing," 2016-2017, p. 1532). Writing for the Court, Justice Ann Walsh

Bradley admitted that COMPAS can only take into account data for groups similar to the offender, and stressed the importance of individualized sentencing ("Criminal Law - Sentencing," 2016-2017, p. 1532). The Court upheld the sentence on the grounds that the COMPAS report was not the sole basis for the sentencing decision ("Criminal Law - Sentencing," 2016-2017, p. 1532). Justice Bradley noted that while risk scores can play a contributing role in sentencing, they are not allowed to be used "to determine whether an offender is incarcerated", and that judges should exercise caution in using the scores in the first place ("Criminal Law - Sentencing," 2016-2017, p. 1532).

Despite the court in Wisconsin upholding the defendant's sentence, it is unclear to what extent recidivism algorithms can be permissible through the lens of fundamental rights. The *Harvard Law Review* notes that simply "encouraging judicial skepticism of the value of risk assessments alone does little to tell judges how much to discount these assessments" ("Criminal Law - Sentencing," 2016-2017, p. 1534). While this example is very local in nature, it shows the conflicts that can arise when algorithmic processes are applied to judicial proceedings. Given the intensely protected nature of the judicial process, the opportunities for algorithms to breach an individual's rights are nearly endless.

Further Examples

Going beyond the concrete examples listed, algorithms can conceivably have a relatively large impact on many more fundamental rights, beyond those immediately apparent. The rights of fair trial, legal defense, and protection against discrimination were covered, however there still remains others, such as the freedom of expression. If algorithmic systems are deployed on any platform, whether public, private, digital, or physical, with the capacity to monitor interpersonal

exchanges, individuals may be more reserved in their opinions or expressions. Similarly, the EU's right to access of personal information (European Union, 2012, p. 397) would obviously be violated by many algorithmic systems operating within a 'black box', with minimal transparency. Furthermore, rights to data protection, a private life, equal treatment, and more can all be conceivably violated by algorithmic security systems.

Discussion

When analyzing the compatibility of algorithmic security systems broadly with fundamental rights, the core issue that can be identified is the lack of a thorough, uniform, effective code of AI ethics for managing possible interactions with fundamental rights. This is in part due to the recently developed nature of AI technology, but largely due to disagreement over how ethical AI is to be created (Jobin et al., 2019, p. 3-6). Researchers studying diverse sets of AI ethical frameworks found that while there was agreement on the broad principles that AI should possess, such as transparency, fairness, privacy, and more, there were many differences in interpretations of the actual values (Jobin et al., 2019, p. 16). Some of the most identifiable differences are in the steps prescribed by different frameworks for achieving more ethical AI, such as the desire for larger, more all-encompassing datasets on one hand, to reduce bias in algorithmic models, contrasted with the individual desire to have a level of privacy and control over one's data (Jobin et al., 2019, p. 16). Similar to the lack of a consistent code of ethics, Europe has a large number of varying data protection frameworks which may separately apply to the use of algorithmic processes in different situations (Gonzalez Fuster, 2020, p. 14-15). While legal frameworks are separate from ethical guidelines, they are closely related, with legal frameworks usually striving to establish a system abiding by such guidelines. Therefore, without a uniform idea of what

constitutes AI ethics, legal frameworks will expectedly vary significantly. Similarly, without this consistent code it becomes impossible to objectively label any algorithmic initiative as unethical. A baseline, or control level of ethics is necessary, in order to judge deviations from it as positive or negative.

Despite the widely varying nature of ethics guidelines and legal frameworks in AI, even consistency is not universally desired (Roberts et al., 2021, p. 3). Certain researchers have made claims supporting the idea that AI ethics is fluid, and dependent on differing sociopolitical factors (Roberts et al., 2021, p. 3). An article in the journal of *Science and Engineering Ethics* proposed an analogy between AI ethics and foreign electrical outlets, in that different regions of the world have different shapes and voltages of outlets, which achieve the same purpose but through different means (Roberts et al., 2021, p. 3). This approach however seems to perpetuate the current environment of disagreement and inconsistency across ethical interpretations, which leads to the inability to preserve fundamental rights. Unless specific algorithmic actions can be clearly, uniformly labeled as unethical, every action can simply be justified under its own, specially adapted ethical reasoning.

Relating to the previously discussed examples, a uniform AI ethics framework would ensure that both could be properly labeled as negative. In the case of PredPol, its supposed racial bias and discriminatory tendencies could be analyzed according to a specific framework and assessed objectively on its level of ethical compliance. Similarly, in the case of COMPAS, a clear ethical framework would eliminate ambiguities in how algorithms can be applied, removing the need for tedious and subjective practices such as the encouragement of judicial skepticism in AI results. In order to move towards an environment of more ethical AI, that respects fundamental rights, there must be a clear understanding of what that exactly entails.

Conclusion

The current environment of algorithmic security systems is fraught with fundamental rights concerns, and far from perfect. While the idea of algorithmic assistance in many of these governmental fields is conceptually possible to respect fundamental rights, the infinite number of avenues by which they can still be violated has shown that achieving it in practice is quite difficult. The introduction of a uniform ethical framework would begin to address this issue, bringing algorithmic processes into a realm of consistency that is deemed to be acceptable, however would still be far from an all-encompassing solution. While it is possible to create a specific ethical framework, conceptualizations of ethics can vary across individuals, and there will always be those who believe the established framework of positive ethics is in some way negative.

References

Chohlas-Wood, A. (2020, June 19). *Understanding risk assessment instruments in criminal justice*. The Brookings Institution. Retrieved November 30, 2022, from https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

Criminal law - Sentencing guidelines - Wisconsin supreme court requires warning before use of algorithmic risk assessments in sentencing - State v. Loomis 881 N.W.2d 749 (Wis. 2016). (2016-2017). *Harvard Law Review*, *130*(5), 1530-1537. https://heinonline-org.ezproxy.leidenuniv.nl/HOL/P?h=hein.journals/hlr130&i=1552

European Parliament. (2020, July). *Artificial intelligence and law enforcement: Impact on fundamental rights* (G. Gonzalez Fuster, Author; Research Report No. PE 656.295). https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf

European Union. (2012, October 26). *Charter of fundamental rights of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

Hicks, J. (2021, December 6). *Go read this data analysis that uncovers predictive policing's flawed algorithm*. The Verge. Retrieved November 30, 2022, from https://www.theverge.com/2021/12/6/22814409/go-read-this-gizmodo-analysis-predpol-software-disproportionate-algorithm

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial intelligence: The global landscape of ethics guidelines. *Nature Machine Intelligence*, *1*(9). https://doi.org/10.48550/arXiv.1906.11668

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing.

Journal of the American Statistical Association, *110*(512), 1399-1411. https://doi.org/10.1080/01621459.2015.1077710

Pasquale, F. (2015). Introduction- The need to know. In *The black box society* (pp. 1-18). Harvard University Press.

Reese, H. (2022, February 23). *What happens when police use AI to predict and prevent crime?* JSTOR Daily. Retrieved November 22, 2022, from https://daily.jstor.org/what-happens-when-police-use-ai-to-predict-and-prevent-crime/

Roberts, H., Cowls, J., Hine, E., Mazzi, F., Tsamados, A., Taddeo, M., & Floridi, L. (2021). Achieving a 'Good AI society': Comparing the aims and progress of the EU and the US. *Science and Engineering Ethics*, *27*(6), 1-25. https://doi.org/10.1007/s11948-021-00340-7

Sankin, A., Mehrotra, D., Mattu, S., Cameron, D., Gilbertson, A., Lempres, D., & Lash, J. (2021, December 2). *Crime prediction software promised to be free of biases. New data shows it perpetuates them*. Gizmodo. Retrieved November 30, 2022, from https://gizmodo.com/crime-prediction-software-promised-to-be-free-of-biases-1848138977

State v. Loomis - 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749 [Case Brief]. (n.d.). *LexisNexis Academic*. Retrieved November 30, 2022, from https://www-lexisnexis-com.ezproxy.leidenuniv.nl/community/casebrief/p/casebrief-state-v-loomis?source=UBLbookmarklet