Academic Workshop

Artificial Intelligence and the Rule of Law

30 March 2023

University of Buenos Aires

Within the framework of the AIDP International Colloquium on "AI and Administration of Justice: Predictive Policing and Predictive Justice"

SPEAKERS / COMMENTATORS

Emmanouil Billis (Max Planck Institute for the Study of Crime, Security and Law)

Linus Ensel (Max Planck Institute for the Study of Crime, Security and Law)

Nandor Knust (UiT The Arctic University of Norway)

Katalin Ligeti (University of Luxembourg)

Julian V. Roberts (University of Oxford)

Christian Thönnes (Max Planck Institute for the Study of Crime, Security and Law)

Elizabeth Tiarks (Northumbria University, Newcastle)

Tatiana Tropina (Leiden University)

Date:	Thursday, 30 March 2023
Location:	School of Law, University of Buenos Aires, Argentina
Workshop Organizer:	Otto Hahn Research Group on Alternative and Informal Systems of Crime Control and Criminal Justice (Max Planck Institute for the Study of Crime, Security and Law)
Int. Colloquium Organizer:	International Association of Penal Law (AIDP), Argentine Group
Contact:	Emmanouil Billis (e.billis@csl.mpg.de) Nandor Knust (nandor.knust@uit.no)

PROGRAM

Thursday, 30 March 2023, 09:00-12:00

Chair/Commentator: *Katalin Ligeti*, Professor, University of Luxembourg

Introduction

09:00 **Artificial Intelligence and the Rule of Law: Opportunities and Challenges**
Emmanouil Billis, Research Group Leader, Max Planck Institute

Session 1: Predictive Policing

09:20 **Re-Negotiating the Social Contract: Artificial Intelligence in Predictive Policing and its Impact on the Legitimacy of Social Control through State's Coercive Powers**
Nandor Knust, Associate Professor, UiT The Arctic University of Norway

09:40 **Forecasting Threats or Creating Threats? Artificial Intelligence, Predictive Policing, and the Rule of Law**
Tatiana Tropina, Assistant Professor, Leiden University

10:00 **Guidelines for Human Intervention in Automated Predictive Decision-Making, as Exemplified by the EU Directive on Passenger Name Record Data**
Christian Thönnes, Doctoral Researcher, Max Planck Institute

10:20 **Discussion**

Session 2: Predictive Justice

10:40 **The Role of AI at Sentencing: Enhancing or Impeding Rule-of-Law Requirements**
Julian V. Roberts, Professor, University of Oxford

11:00 **Predictive Justice and the Purposes of Sentencing in England and Wales**
Elizabeth Tiarks, Assistant Professor, Northumbria University

11:20 **The Compatibility of Algorithm-Based Sentencing with the Notion of Culpability and the Right to Be Heard Before a Court in German Doctrine**
Linus Ensel, Doctoral Researcher, Max Planck Institute

11:40 **Discussion and Conclusions**

WORKSHOP DESCRIPTION

In our globalized world, threats to the peaceful coexistence of humans and to public security have become global as well. In response to successive incidents of terrorism and highly sophisticated forms of serious national and transnational crime, a new security architecture is emerging. It is characterized by a global transformation of traditional legal notions and the blurring of boundaries between security and criminal law, prevention and repression, crime and migration control, as well as the exercise of public and private power. Personal freedom- and privacy-preserving principles such as the requirement of a concrete suspicion or threat prior to launching investigations are replaced with generalized mass surveillance, and, at the institutional level, informational separation makes way for the interoperability of national and international databases.

An increasing reliance on modern Artificial Intelligence (AI) tools for the purpose of the automated detection and suppression of security threats and possible crimes forms an integral part of this phenomenon. Proponents of these tools purport AI's ability to enhance law enforcement and criminal justice in many ways: It can, they say, make the exercise of public authority more effective and efficient, legal decisions less biased and noisy, and interferences with fundamental rights less freedom-threatening. However, the use of these new technologies also entails many rule-of-law and human rights concerns. AI has the potential to radically expand the states' powers of surveillance and coercion, as well as to drastically accelerate the aforementioned transformation process. As we introduce automation into our legal notions, they are transformed – potentially beyond recognition. This trend needs to be counterbalanced by bolstering established rule-of-law safeguards and procedural guarantees. In liberal democratic orders employing AI technology, the primary focus must be, hence, on fulfilling the constitutive requirements of the rule of law: the principle of legality, including the consistent and impartial application of foreseeable, clear and transparent norms and institutions; the principles of equality and proportionality; the nonarbitrary use of power and the respect of fundamental rights and procedural guarantees; the separation of state powers and the control of their exercise by independent and impartial judicial organs.

This workshop, organized by the Max Planck Society's Otto Hahn Research Group on Alternative and Informal Systems of Crime Control and Criminal Justice within the framework of the AIDP International Colloquium on "AI and Administration of Justice: Predictive Policing and Predictive Justice," is designed to tackle these complex questions. It focuses on the meaning and significance of rule of law and human rights considerations when designing and employing AI tools for security, crime control, and criminal justice purposes. We will be honing in on both the promises and perils of AI by integrating criminal and security law concerns into our cogitation on preventive justice – just as the emerging security architecture transcends these disciplinary boundaries. Participants are called to explore and discuss issues ranging from the practical opportunities and the rule-of-law challenges created by specific predictive policing instruments to the transformation of traditional notions, such as culpability or punishment, under the paradigm of predictive justice. Particularly sensitive matters connected to broader socio-legal and legal-ethical questions about justice, legitimacy, and democracy, such as the issues of privacy and data protection, transparency, bias, deception, explicability, and accountability, constitute significant subjects to be addressed during the workshop.

DESCRIPCIÓN DEL TALLER

En un mundo globalizado como el nuestro, las amenazas a la pacífica coexistencia de los humanos y a la seguridad pública también se han hecho globales. En respuesta a los continuos incidentes terroristas y a las formas de delincuencia grave nacional y transnacional de gran sofisticación, está surgiendo una nueva arquitectura de seguridad. Esta se caracteriza por una transformación global de los conceptos jurídicos tradicionales y la difuminación de los límites entre seguridad y derecho penal, prevención y represión, control de la delincuencia y de la migración, así como el ejercicio del poder público y privado. Los principios de libertad individual y respeto por la intimidad, como el requisito de existencia de sospechas concretas o amenazas antes de iniciar una investigación, son sustituidos por un vigilancia masiva y, a nivel institucional, la separación de la información deja paso a la interoperabilidad de las bases de datos nacionales e internacionales.

El aumento de la confianza en las herramientas modernas de inteligencia artificial para la detección y supresión automatizadas de amenazas para la seguridad y posibles formas de delitos forma parte de este fenómeno. Los partidarios de estas herramientas defienden la capacidad de la inteligencia artificial de mejorar la aplicación de las leyes y la justicia penal de muchas formas. Afirman que puede realizar la labor de las autoridades públicas con mayor eficiencia y eficacia, tomar decisiones legales con menos sesgos y "ruidos", y conseguir que las interferencias con los derechos fundamentales supongan un peligro menor para las libertades. Sin embargo, el uso de estas nuevas tecnologías también entraña muchos problemas para el Estado de Derecho y los derechos humanos. La inteligencia artificial tiene el potencial de ampliar radicalmente la capacidad de los estados de control y coerción, así como de acelerar considerablemente el proceso de transformación mencionado. A medida que introducimos la automatización en nuestros conceptos jurídicos, estos se transforman (posiblemente, más de lo que seamos capaces de identificar). Esta tendencia debe contrarrestarse reforzando las salvaguardias del Estado de Derecho y las garantías procesales establecidas. Por ello, en los sistemas democráticos liberales en los que se empleen tecnologías de inteligencia artificial, la prioridad ha de ser cumplir los requisitos esenciales del Estado de Derecho: el principio de legalidad, incluida la aplicación constante e imparcial de normas e instituciones previsibles, claras y transparentes; los principios de igualdad y proporcionalidad; el uso no arbitrario del poder y el respeto por los derechos fundamentales y las garantías procesales; la separación de los poderes del estado y el control de estos a través de órganos judiciales imparciales e independientes.

En este taller, organizado por el Grupo de investigación Otto Hahn sobre sistemas alternativos e informales de control de la delincuencia y justicia penal de la Sociedad Max Planck dentro del marco del Coloquio Internacional de la Asociación Internacional de Derecho Penal (AIDP) "Inteligencia Artificial y administración de justicia: la policía predictiva y la justicia predictiva", está diseñado para abordar estas complejas cuestiones. Nos centraremos en el significado y la importancia del Estado de Derecho y en reflexiones sobre los derechos humanos a la hora de diseñar y utilizar herramientas de inteligencia artificial para la seguridad, el control de la delincuencia y la justicia penal. Nuestro principal punto de interés serán tanto las promesas como los peligros de la inteligencia artificial y, además, incluiremos cuestiones de derecho penal y de seguridad en nuestra reflexión sobre la justicia preventiva (simplemente porque la incipiente arquitectura de seguridad trasciende los límites de estas disciplinas). Invitamos a los participantes a que exploren y debatan aspectos que van desde las oportunidades prácticas y los retos para el Estado de Derecho generados por determinados instrumentos de policía predictiva hasta la transformación de nociones tradicionales como la culpabilidad o la pena bajo el paradigma de la justicia predictiva. Los asuntos especialmente delicados relacionados con cuestiones sociojurídicas y ético-legales más amplias sobre justicia, legitimidad y democracia, como la privacidad y la protección de datos, la transparencia, los sesgos, el engaño, la explicabilidad y la responsabilidad, son temas importantes que serán tratados en el taller.

ABSTRACTS

Artificial Intelligence and the Rule of Law: Opportunities and Challenges

Emmanouil Billis

In the "global risk society" crime is becoming more sophisticated, complex, and transnational, while enforcement and judicial systems are becoming ponderous and overloaded. As a result, the practical significance of mechanisms and institutions aimed at enhancing (national and transnational) law enforcement and improving justice administration has grown. A key aspect in this regard is the revolutionary importance of Artificial Intelligence (AI) for many policy sectors. Numerous legal orders are currently resorting to this technology with the goal of strengthening the efficiency and effectiveness of crime control and criminal justice systems and of optimizing the decision-making processes. In an era of multiple novel challenges in the fight against crime, a plethora of AI applications has emerged in parallel with traditional enforcement and judicial practices set to serve a variety of purposes: from predictive policing, crime prevention, and crime detection to risk and recidivism assessment, the processing of evidence, and the determination of criminal punishment.

Fundamental research and legal policy are called to address not only the opportunities but also the considerable risks for a peaceful coexistence of humans that come with such an evolution mainly in two ways. On the one hand, compared to prior (conventional) technological advancements, employing new AI technology to realize ambitious anti-crime plans might result in broader, more direct, and multi-layered threats to established rights and freedoms. In view of this, extensive *a priori* bans of AI uses identified as particularly dangerous for individuals or societies may be deemed necessary. On the other hand, the debate over the "inevitability" of the expansion of AI should not only be about getting the most out of this technology in terms of effective crime fighting. The first priority must be to create the algorithms and to program the machines in accordance with the overriding objectives, i.e., to protect and secure respect for the most basic human and social values.

In terms of the relationship between AI and the rule of law specifically, the challenge is twofold: to proactively program AI tools in a way that excludes any arbitrariness in their decision-making processes; and to optimize the operation and learning processes of AI with the overall purpose of complementing the traditional justice sector in producing more accurate, objective, and fair results. This introductory contribution discusses characteristic questions and problems of contemporary importance for legal theory, policy, and practice associated with the key notions of human dignity, legality, proportionality, privacy, equality, and procedural justice. It focuses on the meaning and significance of rule of law and human rights considerations in designing and employing AI tools for crime control and criminal justice purposes.

Re-Negotiating the Social Contract: Artificial Intelligence in Predictive Policing and its Impact on the Legitimacy of Social Control through State's Coercive Powers

Nandor Knust

Starting point of this presentation is the general idea of the overall function of criminal law as the state's *ultima ratio* instrument for guaranteeing social order and peace. The *raison d'être* of public authority, that is, the safeguarding of freedom and human dignity of all individuals, can be seen in the notions of the social contract and the rule of law, which define and navigate our social existence in the community and society.

The social contract is an agreement to form one entity, a collectivity, which by definition is more than just an aggregation of individual interests and wills. By collectively renouncing the rights and freedoms, which the individual has in the "state of nature", and transferring these rights to the collective body, a new "person" (sovereign/state) is formed. Included in this version of the social contract is the idea of reciprocated duties: the sovereign/state is committed to the good of its individuals, and in turn each individual is committed to the common good. This social construct is backed and controlled by the concept of the rule of law. In rule-of-law orders a main task of public authority is to safeguard the freedom and human dignity of all individuals.

In modern societies social control and peace are maintained by law enforcement and criminal justice systems, where decisions with severe consequences for the freedoms and behavior of individuals are taken. Recently, such decisions have been increasingly influenced by special algorithms aimed at predicting criminal behavior and by the corresponding preventive police action. The increased use of predictive software by the police creates the danger of widespread and uncontrolled use of data and meta-data. The integration of incomprehensible algorithms and autonomous machines into the law enforcement and criminal justice decision-making systems poses a threat to the above-mentioned ideas of the social contract and the rule of law. Unlike crime forecasting approaches by human-machine-interface guided methods, the new algorithms are capable to decide for themselves what and how they learn and choose. This makes them not only non-transparent or non-understandable for people without the necessary technical skills but may even create a "black box" for their developers due to their immense complexity in the decision-making process (especially Deep Learning).

The use of algorithmic forecasting software completely changes police and justice decision-making processes and the relationship of the police with society. The characteristics of this new type of algorithms raise several questions about the role of artificial intelligence in social control but also in changing normative expectations and the predictability of state action. These decision-making processes have enormous implications for the structural organization of institutions within crime prevention, crime control, and criminal justice. Against this backdrop, this presentation discusses whether the changing landscape of crime control by means of predictive policing needs a "re-negotiation" of the social contract between the state as the owner of the power monopoly and its citizens.

Forecasting Threats or Creating Threats? Artificial Intelligence, Predictive Policing, and the Rule of Law

Tatiana Tropina

Digital technology is increasingly transforming not only criminal behavior but also ways to investigate and disrupt it. The last two decades have seen a significant change in approaches to combatting crime. The focus of law enforcement is gradually shifting from investigation of offenses that have already taken place to the idea of predictive, intelligence-driven policing. Initially, this shift was facilitated by growth in volumes of digital data that could potentially be collected and analyzed to forecast future crimes. However, in recent years, this approach has been evolving with the growing incorporation of AI decision-making systems in predictive policing. Algorithmic risk assessment, profiling, biometric identification systems, and emerging tools, such as emotion recognition technologies, have been employed by law enforcement to identify potential criminal activities and forecast criminal behavior.

Yet the promise of AI technologies is a double-edged sword. Significant advantages of using algorithmic decision-making by law enforcement give rise to even more significant challenges. While employed to

help police foresee future threats, AI technologies in predictive policing have been increasingly considered a growing threat to the rule of law. Various problems associated with the use of these tools – the lack of transparency and scrutiny, racial and gender biases, stigmatization, and oppression of disadvantaged groups – can reinforce existing inequalities, lead to abuse of power, undermine fundamental rights, erode regimes of accountability of state institutions, and, ultimately, cause the loss of societal trust in law enforcement and criminal justice.

This presentation aims to discuss the tension between the rule of law and the rapid evolvement of AI tools designed and employed for the purpose of forecasting criminal behavior. After briefly scoping the topic of AI application in predictive policing, the presentation will illustrate the integration of AI tools into the risk-based prevention practices of law enforcement agencies and discuss associated benefits and problems. It will examine how algorithm-based decision-making in predictive policing can potentially affect the rule of law, especially by creating discriminatory patterns and reinforcing social inequalities combined with the lack of transparency and mechanisms for redress.

The analysis will further consider how these challenges can be addressed by strengthening the rule-of-law requirements for predictive policing in the context of AI. This discussion will focus on the permissibility for law enforcement to use certain AI tools and the conditions and safeguards for such use. It will also examine the possibility of "red lines" imposed by the rule of law, which would justify moratoriums or prohibitions on the development and use of some AI technologies in predictive policing.

The presentation will conclude by looking into how the challenges to the rule of law are being tackled in the current development of regulatory standards for AI. It will briefly assess the relevant discussions around the European Commission's Proposal for a Regulation on Artificial Intelligence, particularly the debates on biometric data identification tools by law enforcement agencies. Lastly, it will consider the Council of Europe's work on the Convention on artificial intelligence, human rights, democracy, and the rule of law.

Guidelines for Human Intervention in Automated Predictive Decision-Making, as Exemplified by the EU Directive on Passenger Name Record Data

Christian Thönnes

In its *Ligue des droits humains* decision, the Court of Justice of the European Union (CJEU) was asked to decide on one of the first EU-wide, large-scale Predictive Policing instruments – the EU Directive on Passenger Name Record Data (PNR Directive). According to the PNR Directive, EU Member States must require air carriers to transmit a set of data for each passenger to national security authorities, so-called Passenger Information Units (PIUs). PIUs then compare all PNR datasets against pre-existing databases (Art. 6 § 3 letter a) and so-called "pre-determined criteria" (Art. 6 § 3 letter b). "Pre-determined criteria" are algorithms containing (allegedly) suspicious flight patterns. Rather than recognizing known suspects and criminals, these algorithms are explicitly aimed at conjuring up new suspicions of future crimes. They do so by targeting previously unknown citizens within the gigantic group of all European flight passengers, based simply on how they choose to travel. These "pre-determined criteria" were therefore widely regarded as a blueprint for the deployment of self-learning algorithms at EU borders.

Crucially, the PNR Directive stipulates that all hits automatically generated through comparisons against databases or pre-determined criteria must be "individually reviewed by non-automated means" (meaning: by humans) before further investigative police measures may be undertaken (Art. 6 § 5). Hence, the PNR Directive, in addition to raising all sorts of human rights and rule-of-law concerns, is an

important opportunity to set standards for individual human intervention in automated decision-making processes within predictive policing frameworks. In its landmark decision on the PNR Directive, the CJEU declined to invalidate the PNR Directive, limited the use of machine-learning technologies within the PNR system and emphasized that "Member States must ensure that the PIU establishes, in a clear and precise manner, objective review criteria" for automated hits (para 206). In my contribution, however, I will argue that the Court failed in fully using its opportunity to set clear standards for predictive policing systems, since it neglected to specify the purpose and content of criteria for human review. Much rather, it delegated their formulation to Member States, merely stating that their main purpose ought to be the prevention of false-positives – the frequent occurrence of which is a mathematical near-certainty within the PNR system since it obliges European security authorities to look for the proverbial needle in a haystack.

In my contribution, I investigate which purpose human interventions fulfill and what the content of criteria for human review in the PNR system could be. In so doing, I introduce a conceptual distinction between epistemic and expressive human interventions. While the former are aimed at generating additional knowledge or correcting factual mistakes regarding an automated hit, the latter mainly serve a communicative function. I argue that the CJEU's approach in *Ligue des droits humains* is flawed because it limits the purpose of human intervention to generating more *human knowledge* when it ought to also aim at generating more *human communication.*

The concept of expressive intervention holds that the socio-legal practice of reason-giving operates in legally defined relationships of mutual recognition. Giving reasons for a legal decision aims at expressively stabilizing the affected person's status as an autonomous moral agent. By explaining our decisions to each other, we mutually recognize our statuses as legal subjects in a free society who are worthy of understanding because we are capable to be motivationally guided by reason. A recognition-based approach to human intervention in automated predictive decision-making is based on the idea that our legal order ascribes the faculties necessary for mutual recognition to humans but not to machines. Machines cannot recognize, because they are not recognized. Hence, the mutual recognition usually operating in legal processes is lacking when the decision to undertake an invasive police measure is made (or meaningfully influenced) by a machine. This lack is especially significant in the PNR system because the prognosis that someone will commit a serious crime in the future involves an unexpected expressive judgment about the content of their character and place in society. Hence, this lack must be compensated through procedural rule-of-law-based guarantees aimed at expressively stabilizing the relationship of mutual recognition between state and citizen.

In my contribution, I sketch out what these procedural guarantees could be – and how, with increasing complexity of automated decision-making technologies, expressive interventions will grow in doctrinal utility vis-à-vis their epistemic counterparts.

The Role of AI at Sentencing: Enhancing or Impeding Rule-of-Law Requirements

Julian V. Roberts

The sentencing process engages many rule-of-law safeguards and procedural requirements, including transparency, impartiality, equality of treatment, and offender allocution. Most jurisdictions allow courts wide discretion at sentencing. As a result, outcomes are often hard to predict, and the system lacks transparency. In addition, judicial and systemic bias may result in disparity of treatment. In all western jurisdictions racial or ethnic minorities are more likely to be sentenced to custody (and for longer periods of imprisonment). As a solution, a number of scholars have argued that AI can effectively

replace human judges at sentencing. Court staff or judicial officers would enter all legally-relevant information about the offense and the offender, and the AI program would then apply the principles of sentencing and generate a sentence. It is argued that this would increase fairness and have a number of other benefits for the sentencing process. Is this likely to enhance rule-of-law values at sentencing?

In this presentation I imagine what sentencing would look like if a program were devised to determine sentence. I will argue that AI enhances certain requirements and undermines others. On the positive side, AI can improve the likelihood of impartial treatment by constraining a potentially biased human decision-maker. A judge could compare his or her proposed sentence to one generated by an algorithm which has processed all the information available to a court. AI will be better able to devise a sentence that conforms to established sentencing principles such as proportionality, restraint, and equity. At the same time, AI will be blind to extra-legal factors which may have influenced a human judge – for example, the defendant's race, ethnicity, immigration status, and employment record. Impartiality can also be improved, since sentencing patterns or sentencing judgments can be scrutinized by Artificial Intelligence to identify sources of bias or specific courts or judges which routinely depart from the established sentence range or tariff for a specific offense. In jurisdictions which operate formal sentencing guidelines (such as England and Wales, South Korea, and many US states), AI can identify elements of these guidelines which trigger inequitable outcomes. For example, AI is much better able (than human researchers|) to identify indirect sources of discrimination at sentencing.

At the same time, under some machine-learning proposals, AI may also undermine the rule of law. For example, algorithms currently used for risk prediction are notoriously lacking in transparency. In addition, replacing a human judge with an algorithm would mean there is no need for a sentencing hearing. For example, one of the most fundamental requirements is for the litigant to be heard. Without a sentencing *hearing*, the offender loses voice. While submissions on the defendant's behalf can be entered into the program (along with other submissions and information), this is at best an unacceptable substitute for offender allocution. If courts reduce the reduce the opportunity for defendants to be heard at the sentencing hearing, this will ultimately compromise perceptions of sentencing legitimacy. A viva-voce sentencing hearing is therefore an indispensable element of the sentencing process. It is also the opportunity for the parties and the victim to interact and exchange perspectives in a way that is impossible if sentencing is determined by an algorithm.

To summarize, in this presentation I argue that AI should supplement and not supplant the sentencing judge. AI can make a significant contribution to improving rule-of-law values at sentencing – but not through the replacement of human decision-making. Instead, AI should be used to monitor sentencing practices and judgments – a task it can perform better than human researchers. The outcomes of this monitoring can be used by sentencing commissions that issue guidelines, courts of appeal that issue guideline judgments, and trial court sentencers. As a result, sentencing will more closely fulfill the rule-of-law requirements ascribed to it. Expanding the role of AI at sentencing is necessary, but courts need to preserve sentencing as an essentially human enterprise.

Predictive Justice and the Purposes of Sentencing in England and Wales

Elizabeth Tiarks

Transparency in sentencing is important for penal legitimacy and encouraging public confidence in sentencing. This talk will consider the impact of predictive risk assessments on transparency in

sentencing, using a specific part of the sentencing process in the jurisdiction of England and Wales as a case study: the sentencer's choice of which purpose of sentencing to pursue.

There are five statutory purposes of sentencing in England and Wales: the punishment of offenders; the reduction of crime (including by deterrence); the reform and rehabilitation of offenders; the protection of the public; and reparation. These purposes can come into conflict and cannot all be satisfied when an offender is sentenced. There is expressly no hierarchy between these purposes and therefore it is up to the sentencer to decide in any given case which to prefer. This creates some opacity in this part of the process, as it is often unclear why a sentencer chooses a particular purpose of sentencing. Whilst there are measures in place to structure the sentencing process and encourage consistency of approach (in the form of sentencing guidelines), there remains some lack of transparency about the way that the decision concerning which purpose of sentencing to pursue is made.

The talk will explore how the lack of transparency in this part of the sentencing process is impacted by the use of predictive algorithms in sentencing in England and Wales. The Offender Assessment System (OASys) and its algorithmic components, such as the Offender Group Reconviction Scale (OGRS), provide a predictive risk score for risk of harm and risk of reoffending, which is used to inform both parole and sentencing decisions. The provision of this score can influence which purpose of sentencing a sentencer chooses, but it is not clear how, or by how much.

The risk of harm and reoffending is arguably most pertinent to the “protection of the public” purpose, which could encourage sentencers to focus more on this purpose of sentencing. However, it is also possible that sentencers might use the risk assessment to support one of the other four purposes, perhaps deeming a low-risk score indicative of the suitability of a “reform and rehabilitation” approach. The issue is further complicated by the varying levels of confidence in predictive algorithms amongst sentencers and the extent to which they are willing to take into account or dismiss the risk score. Ultimately, it is difficult to know precisely how predictive risk assessments affect sentencers’ decision-making. It will be argued that they increase uncertainty about the process by which different purposes of sentencing are selected, exacerbating the existing problem and reducing transparency.

The Compatibility of Algorithm-Based Sentencing with the Notion of Culpability and the Right to Be Heard Before a Court in German Doctrine

Linus Ensel

In the light of considerable interregional variance in sentencing in Germany and an accompanying deficit of reasoning in sentencing decisions, the call for a rationalization of the sentencing process seems natural. One way to achieve such a rationalization is through the use of modern technology. In particular, two approaches stand out, which I would like to address in my contribution.

A first option would be a system that incorporates machine learning (ML approach). The system would be fed with a large number of cases, thus “learning” the correlation between the facts of the case and the sentence that was given. In application, the judge would provide the system with the relevant facts of the case (input data) and would receive an aggregate average penalty (output data).

Alternatively, one could define the specific sentencing criteria in a statutory and computer-oriented manner and incorporate them into a (non-learning) algorithm (NLA approach). When provided with the facts of the case, this algorithm could then calculate a certain sentence or a sentencing range.

Both approaches are conceivable in a fully automated variant (FA) as well as part of a decision support system (DSS), whereby the latter would still require a final human decision. The implementation of either of these tools (ML/NLA), in either variant (FA/DSS), would face several obstacles – technical, legal, and ethical. In my presentation, I will focus on two particular obstacles that might arise for constitutional reasons.

According to the German Federal Constitutional Court (*Bundesverfassungsgericht*), the principle of culpability is of constitutional rank. It is based upon the rule of law (Art. 20 section 3 of the German constitution, *Grundgesetz, "GG"*) and the guaranty of human dignity (Art. 1 section 1 GG). In the context of criminal proceedings, this guaranty includes the prohibition of degrading a person to a mere object of the trial. I will make the argument that – at least in the sensitive area of criminal convictions – a fully automated approach would violate that guaranty and therefore the principle of culpability. Even with DSS approaches, automation bias could potentially lead to *de facto* untested decisions. To prevent this, the human judge's duties and the final decision-making process – subsequent to the machine decision – must be clearly defined.

Furthermore, according to the principle of culpability, the penalty imposed must be "justly proportionate to the gravity of the offence and the degree of culpability of the offender". The German legislator has stated that the assessment of culpability is not a "strict scientific finding", but a "moral evaluation process within the legal community". Some even speak of an "act of social composition". I will argue that the comparative sentencing approach of a machine learning system would not be consistent with this conception of culpability.

The next yardstick against which I will measure the ML and NLA approaches is the right to be heard before a court, which is guaranteed in Art. 103 section 1 GG. This right also derives from the rule of law (Art. 20 section GG) and the guaranty of human dignity (Art. 1 section 1 GG). Article 103 section 1 GG contains a right of the accused to express themselves to the court and a right to have their word taken into account. However, these two manifestations would be of little value if the accused were unable to inform themselves about all facts relevant to the decision prior. The right to be heard before a court therefore also implies a requirement of transparency. I will point out issues regarding transparency within ML and NLA approaches and conclude that a ML approach would not meet the requirements regarding the explainability of the system – at least not with the means available today. The implementation of a NLA approach, on the other hand, would be compatible with the right to be heard before a court if the algorithm's calculation path is disclosed and all relevant aspects of the case can be considered when the final human decision is made.

RESÚMENES

Inteligencia artificial y Estado de Derecho: oportunidades y retos

Emmanouil Billis

En la "sociedad del riesgo global", los delitos son cada vez más sofisticados, complejos y transnacionales, mientras que las fuerzas del orden y los sistemas judiciales se hacen lentos y están saturados. Como resultado, ha crecido la importancia práctica de los mecanismos e instituciones orientados a mejorar la aplicación de la ley (nacional y transnacional) y la administración de la justicia. Un aspecto clave en este sentido es la revolucionaria relevancia de la inteligencia artificial en muchos ámbitos normativos. Un gran número de ordenamientos jurídicos recurren en la actualidad a esta tecnología con el objetivo de aumentar la eficiencia y eficacia de los sistemas de control de la delincuencia y de justicia penal y de optimizar los procesos de toma de decisiones. En una era de numerosos retos nuevos en la lucha contra la delincuencia, ha surgido un sinfín de aplicaciones de la inteligencia artificial que actúan en paralelo con las prácticas tradicionales de las fuerzas del orden y los sistemas judiciales y están destinadas a cumplir diversos objetivos: desde la policía predictiva, la prevención de la delincuencia y la detección de delitos hasta la evaluación de los riesgos y la reincidencia, el procesamiento de las pruebas y la determinación de la sanción penal.

La investigación básica y la política legal son llamadas a abordar, además de las oportunidades, los notables riesgos que dicha evolución plantea para una coexistencia pacífica de los humanos y que, principalmente, se perfilan en dos formas. Por un lado, en comparación con los avances tecnológicos anteriores (convencionales), el uso de nuevas tecnologías de inteligencia artificial con el fin de elaborar planes ambiciosos para la lucha contra la delincuencia puede desembocar en amenazas de mayor alcance, más directas y a distintos niveles que ponen en riesgo los derechos y libertades establecidos. En vista de esto puede considerarse necesario aplicar *a priori* importantes prohibiciones a los usos de la inteligencia artificial identificados como especialmente peligrosos para los individuos o las sociedades. Por otro lado, el debate sobre la "inevitabilidad" de la expansión de la inteligencia artificial no debería estar centrado exclusivamente en sacar el máximo partido de esta tecnología en lo que a una lucha eficaz contra la delincuencia respecta. La prioridad debe ser crear los algoritmos y programar las máquinas en función de los objetivos más importantes, por ejemplo, proteger y garantizar el respeto por los valores humanos y sociales más básicos.

En lo que se refiere a la relación entre la inteligencia artificial y el Estado de Derecho en concreto, el reto tiene una doble vertiente: programar de forma proactiva las herramientas de inteligencia artificial de manera que se excluya cualquier arbitrariedad en su proceso de toma de decisiones y optimizar los procesos de funcionamiento y aprendizaje de la inteligencia artificial con el objetivo general de complementar a la justicia tradicional a la hora de obtener unos resultados más precisos, imparciales y equitativos. En esta contribución introductoria, analizaremos las cuestiones y los problemas típicos que son de relevancia actualmente para la teoría, las políticas y las prácticas legales asociados a las nociones clave de dignidad humana, legalidad, proporcionalidad, privacidad, igualdad y justicia procesal. Nos centraremos en el significado y la importancia del Estado de Derecho y en reflexiones sobre los derechos humanos a la hora de diseñar y utilizar herramientas de inteligencia artificial para el control de la delincuencia y la justicia penal.

Renegociando el contrato social: el uso de la inteligencia artificial en la prevención del delito y su impacto en la legitimidad del control social a través del poder de coerción estatal

Nandor Knust

Este estudio toma como punto de partida la concepción del derecho penal como instrumento de última ratio para garantizar el orden social y la paz. La *raison d'être* del poder público, el otorgarle la función de protección de la libertad y de la dignidad humana de todos los individuos, se explica a través del concepto del contrato social y del Estado de Derecho, dos nociones básicas sobre las que gira nuestra convivencia y existencia social en la comunidad y en la sociedad.

Como es sabido, en la teoría política el contrato social constituye un acuerdo para formar una entidad o una colectividad, que por definición es algo más que una simple agregación de intereses y voluntades individuales. A través de la renuncia colectiva a los derechos y libertades que el individuo goza en el "estado de naturaleza", y la transmisión de esos derechos al ente colectivo, se genera una nueva forma de "persona" (soberano/estado). A esta concepción del contrato social subyace la idea de unos deberes recíprocos: el soberano/estado está comprometido en proveer el bien de los individuos y, a su vez, cada individuo está comprometido con el bien común.

Esta construcción social se sustenta y, a su vez se ve conformada por el concepto de Estado de Derecho. En los sistemas regidos por los principios del Estado de Derecho, una de las principales funciones del poder público es precisamente salvaguardar la libertad y la dignidad humana de todas las personas, nociones inherentes al estado de derecho.

En las sociedades modernas, el control social y la paz se mantienen a través del cumplimiento de la ley, recurriendo entre otros mecanismos, al control policial y la justicia penal, en cuyo ámbito pueden llegar a tomarse decisiones que afectan gravemente a la libertad y el comportamiento de los individuos.

Actualmente, estas decisiones cada día se ven más influidas por algoritmos diseñados para predecir conductas delictivas, en función de los cuales se llevan a cabo las actuaciones policiales preventivas. El creciente recurso a estos software predictivos por parte de la policía genera el riesgo de que se extienda un uso generalizado e incontrolado de datos y metadatos. El hecho de que los procesos de toma de decisiones estén determinados cada vez más por algoritmos incomprensibles, que además actúan con un creciente grado de autonomía, representa una clara amenaza tanto para las bases del contrato social como para los principios del Estado de Derecho.

A diferencia de los sistemas de prevención del crimen que operan sobre la base de sistemas interfaz hombre-máquina, los nuevos algoritmos son capaces de decidir por sí mismos lo que van a aprender y cómo van a aprender. Esto los convierte no solo en poco transparentes e incomprensibles para toda persona que carezca de los conocimientos técnicos específicos, sino que incluso pueden llegar a crear una "caja negra" (*black box*) que haga que sean indescifrables para los propios programadores, debido a la enorme complejidad del proceso de toma de decisiones (especialmente en el *Deep learning*).

La utilización de software de pronóstico algorítmico altera por completo los procesos de toma de decisiones policiales y judiciales y la relación entre la policía y la sociedad.

Las características de este nuevo tipo de algoritmos plantean diversas cuestiones acerca del papel que desempeña o debe desempeñar la inteligencia artificial en el control social, así como en relación con las expectativas normativas y la previsibilidad de la acción estatal. Estos procesos de toma de decisiones tienen enormes implicaciones para la organización estructural de las instituciones que operan en la prevención y control del delito, así como en el ámbito de la justicia penal.

En este contexto, nuestro estudio analiza la cuestión relativa acerca de si este nuevo panorama en los sistemas de control del crimen a través de la denominada *predictive policing* obliga a "renegociar" el contrato social entre los ciudadanos y el estado que ostenta el monopolio del poder.

¿Predecimos amenazas o las creamos? Inteligencia artificial, policía predictiva y Estado de Derecho
Tatiana Tropina

Las tecnologías digitales están transformando cada vez más no solo las conductas delictivas, sino también las formas de investigarlas y desarticularlas. A lo largo de las dos últimas décadas se ha producido un cambio significativo en los métodos para combatir la delincuencia. El enfoque de las fuerzas del orden está pasando gradualmente de la investigación de delitos que ya se han cometido al concepto de policía predictiva basado en la inteligencia. Inicialmente, este cambio fue propiciado por el crecimiento del volumen de datos digitales que podían recopilarse y analizarse con el objetivo de predecir futuros delitos. Sin embargo, en los últimos años este planteamiento ha ido evolucionando con el aumento del número de sistemas de toma de decisiones basados en la inteligencia artificial que se han incorporado a la policía predictiva. Las fuerzas del orden se han servido de la evaluación de riesgos mediante algoritmos, la elaboración de perfiles, los sistemas de identificación biométrica y las herramientas emergentes como las tecnologías de reconocimiento de emociones con el fin de identificar posibles actividades delictivas y predecir las conductas de este tipo.

No obstante, la promesa de las tecnologías de inteligencia artificial es un arma de doble filo: los importantes avances en la toma de decisiones mediante algoritmos por parte de las fuerzas del orden plantean retos aún mayores. A pesar de que se emplean para ayudar a la policía a predecir futuras amenazas, se considera cada vez más que el uso de tecnologías de inteligencia artificial para la policía predictiva supone una creciente amenaza para el Estado de Derecho. Diversos problemas asociados a la utilización de estas herramientas (la falta de transparencia y escrutinio, los prejuicios raciales y de género, los estigmas y la opresión de grupos desfavorecidos) pueden reforzar las desigualdades existentes, desembocar en un abuso de poder, vulnerar derechos fundamentales, destruir los regímenes de responsabilidades de las instituciones estatales y, en última instancia, provocar la pérdida de la confianza de la sociedad en las fuerzas del orden y en el sistema de justicia penal.

La finalidad de esta presentación es analizar la tensión creada entre el Estado de Derecho y la rápida evolución de las herramientas de inteligencia artificial diseñadas y empleadas para predecir conductas delictivas. Tras tratar brevemente el tema de la aplicación de la inteligencia artificial para la policía predictiva, examinaremos la integración de estas herramientas en las prácticas de prevención de riesgos de las fuerzas del orden y expondremos las ventajas y los inconvenientes asociados. En concreto, analizaremos cómo podrían afectar al Estado de Derecho las decisiones basadas en algoritmos en el ámbito de la policía predictiva, principalmente creando patrones discriminatorios y aumentando las desigualdades sociales, todo ello combinado con la falta de transparencia y con mecanismos de rectificación.

En el marco de este análisis también consideraremos cómo afrontar estos retos afianzando los requisitos que el Estado de Derecho impone a la policía predictiva en el contexto de la inteligencia artificial. Asimismo, nos centraremos en la permisibilidad de las fuerzas del orden a la hora de usar ciertas herramientas de inteligencia artificial, así como en las condiciones y garantías de dicho uso. Además, analizaremos la posibilidad de que el Estado de Derecho trace ciertas "líneas rojas" que justificarían la suspensión o prohibición del desarrollo y uso de algunas tecnologías de inteligencia artificial para la policía predictiva.

Para finalizar, repasaremos cómo se están abordando los retos a los que se enfrenta el Estado de Derecho en la actual elaboración de leyes para la regulación de la inteligencia artificial. También evaluaremos brevemente los debates pertinentes relativos a la propuesta de la Comisión Europea de regulación de la inteligencia artificial, en particular los debates en torno al uso de herramientas de identificación de datos biométricos por parte de las fuerzas del orden. Por último, reflexionaremos sobre el trabajo del Consejo de Europa en lo que al Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho respecta.

Directrices para la intervención humana en los procesos automatizados de toma de decisiones predictivas automatizadas, según el ejemplo de la Directiva "PNR"

Christian Thönnes

En su sentencia en *Ligue des droits humains*, se pidió al Tribunal de Justicia de la Unión Europea (TJUE) que se pronunciara sobre uno de los primeros instrumentos de policía predictiva (*predictive policing*) a gran escala de la UE: la Directiva de la recolección y transmisión obligatoria de datos de los pasajeros de vuelos (Directiva PNR, *passenger name record*). Según la Directiva PNR, los Estados miembros de la UE deben exigir a las compañías aéreas que transmitan un conjunto de datos de cada pasajero a las autoridades nacionales de seguridad, las denominadas Unidades de Información sobre los Pasajeros (UIP), las cuales comparan todos los conjuntos de datos PNR con bases de datos preexistentes (art. 6 § 3 letra a) y con los denominados "criterios predeterminados" (art. 6 § 3 letra b). Los "criterios predeterminados" son algoritmos que contienen patrones de vuelo (supuestamente) sospechosos. En lugar de reconocer a sospechosos y delincuentes conocidos, estos algoritmos están explícitamente dirigidos a detectar futuros delitos. Lo hacen seleccionando a ciudadanos previamente desconocidos dentro del gigantesco grupo de todos los pasajeros de vuelos europeos, basándose simplemente en cómo eligen viajar. Por ello, estos "criterios predeterminados" se consideran en general como un modelo para el despliegue de tecnologías de inteligencia artificial en el marco de sistemas de autoaprendizaje (*machine learning*) en las fronteras de la UE.

La Directiva PNR estipula que todas las respuestas positivas generadas automáticamente mediante comparaciones con bases de datos o criterios predeterminados deben ser revisadas "individualmente por medios no automatizados" (es decir: por seres humanos) antes de que puedan adoptarse nuevas medidas policiales de investigación (art. 6 § 5). Por lo tanto, la Directiva PNR, además de plantear todo tipo de problemas relacionados con los derechos humanos y el Estado de Derecho, representa una oportunidad para establecer normas para la intervención humana en los procesos automatizados de toma de decisiones dentro de los marcos de *predictive policing*. En su histórica decisión sobre la Directiva PNR, el TJUE se negó a invalidarla, limitó el uso de sistemas de autoaprendizaje (*machine learning*) dentro del sistema PNR y subrayó que los Estados miembros deben garantizar que "la UIP establezca, de forma clara y precisa, criterios de revisión objetivos" para las respuestas positivas automatizadas (apartado 206). En mi contribución, argumentaré que el Tribunal no aprovechó plenamente su oportunidad para establecer normas claras para los sistemas policiales predictivos, ya que omitió especificar el objetivo y el contenido de los criterios para la revisión humana. Más bien, delegó su formulación en los Estados miembros, limitándose a afirmar que su principal objetivo debería ser la prevención de los falsos positivos (*false positives*), cuya frecuente aparición es una casi certeza matemática en el sistema PNR, ya que obliga a las autoridades europeas de seguridad a buscar la aguja en un pajar.

En mi contribución, investigo qué objetivo persiguen las intervenciones humanas y cuál podría ser el contenido de los criterios para la revisión humana en el sistema PNR. Para ello, introduzco una distinción conceptual entre intervenciones humanas epistémicas (*epistemic interventions*) y expresivas (*expressive interventions*). Mientras que las primeras tienen por objeto generar conocimientos adicionales o corregir errores de hecho en relación con una respuesta positiva automatizada, las segundas cumplen principalmente una función comunicativa. Argumento que el enfoque del TJUE en *Ligue des droits humains* es erróneo porque limita el objetivo de la intervención humana a generar más *conocimiento humano*, cuando también debería tener como objetivo generar más *comunicación humana intersubjetiva*.

El concepto de intervención expresiva sostiene que la práctica sociojurídica de dar razones opera en relaciones jurídicamente definidas de reconocimiento recíproco (*mutual recognition*). Dar razones de una decisión jurídica pretende estabilizar expresivamente el estatus de la persona afectada como agente moral autónomo. Al explicar nuestras decisiones unos a otros, reconocemos mutuamente nuestra condición de sujetos de derecho en una sociedad libre y dignos de comprensión porque somos capaces de guiarnos motivacionalmente por la razón. Un enfoque basado en el reconocimiento de la intervención humana en la toma de decisiones predictivas automatizadas se basa en la idea de que nuestro ordenamiento jurídico atribuye las facultades necesarias para el reconocimiento recíproco a los seres humanos, pero no a las máquinas. Las máquinas no pueden reconocer porque no son reconocidas. Por lo tanto, el reconocimiento recíproco que suele operar en los procesos jurídicos no está presente cuando la decisión de llevar a cabo una medida policial invasiva es tomada (o influida significativamente) por una máquina. Esta falta de reconocimiento recíproco es especialmente significativa en el sistema PNR ya que el pronóstico de que alguien cometerá un delito grave en el futuro implica un juicio expresivo inesperado sobre el contenido de su carácter y su posición en la sociedad humana. Por lo tanto, esta omisión debe compensarse mediante garantías procesales basadas en el Estado de Derecho dirigidas a estabilizar expresivamente la relación de reconocimiento recíproco entre el Estado y el ciudadano.

En mi contribución, esbozo cuáles podrían ser estas garantías procesales – y cómo, con la creciente complejidad de las tecnologías de toma de decisiones automatizadas, las intervenciones expresivas crecerán en utilidad doctrinal frente a sus contrapartes epistémicas.

El papel de la inteligencia artificial en las sentencias condenatorias: mejora u obstaculización de los requisitos del Estado de Derecho

Julian V. Roberts

El proceso de sentencia abarca numerosas garantías del Estado de Derecho y requisitos procedimentales como la transparencia, la imparcialidad, la igualdad de trato y la declaración final del imputado antes de la sentencia condenatoria. En la mayoría de las jurisdicciones, los jueces gozan de un amplio poder discrecional a la hora de dictar sentencia. En consecuencia, los resultados suelen ser difíciles de predecir y el sistema carece de transparencia. Además, la parcialidad de los tribunales y del sistema puede traducirse en diferencias de trato. En todas las jurisdicciones occidentales, las minorías raciales o étnicas tienen mayor probabilidad de ser condenadas con penas privativas de la libertad (y durante periodos mayores de cárcel). Como solución, varios expertos han señalado que la inteligencia artificial puede sustituir de manera eficaz a los jueces humanos a la hora de dictar las sentencias. El personal del juzgado o los funcionarios judiciales introducirían toda la información legal pertinente sobre el delito y el infractor, y el programa de inteligencia artificial aplicaría los principios de las sentencias condenatorias y generaría la condena. Se argumenta que, de esta forma, se aumentaría la

imparcialidad y se obtendrían otras ventajas en lo que al proceso de sentencia se refiere. ¿Podrá este sistema mejorar los valores del Estado de Derecho a la hora de emitir las sentencias?

En esta presentación imaginaremos cómo serían las penas si se diseñara un programa capaz de dictar sentencias. En mi opinión, la inteligencia artificial favorece ciertos requisitos y menoscaba otros. Siendo positivos, la inteligencia artificial puede aumentar la probabilidad de trato imparcial limitando las decisiones humanas potencialmente sesgadas de los jueces. Así, un juez podría comparar su propuesta de sentencia con la generada por un algoritmo que ha procesado toda la información de la que dispone el tribunal. La inteligencia artificial tendrá una mayor capacidad de emitir una pena conforme a los principios de sentencia establecidos como la proporcionalidad, la mesura y la equidad. Al mismo tiempo, la inteligencia artificial será ajena a los factores extrajudiciales que pueden influir en un juez humano como, por ejemplo, la raza, el origen étnico, la condición de inmigrante o la trayectoria laboral del acusado. Además, puede mejorarse la imparcialidad, ya que la inteligencia artificial es capaz de analizar los patrones de las sentencias condenatorias o los fallos emitidos e identificar fuentes de sesgos o determinados tribunales o jueces que se desvían habitualmente del rango de penas establecido o de las multas impuestas a los distintos delitos. En las jurisdicciones en las que se utilizan directrices formales sobre las sentencias condenatorias (como es el caso de Inglaterra y Gales, Corea del Sur y muchos estados de Estados Unidos), la inteligencia artificial puede identificar elementos en estas directrices que dan lugar a resultados injustos. Por ejemplo, la inteligencia artificial tiene una mayor capacidad (que los investigadores humanos) de identificar fuentes indirectas de discriminación en las sentencias condenatorias.

Asimismo, en algunas propuestas de aprendizaje automático, la inteligencia artificial también puede socavar el Estado de Derecho. Por ejemplo, los algoritmos que se emplean en la actualidad para predecir los riesgos carecen claramente de transparencia. Además, si se sustituyera a los jueces humanos por algoritmos, las audiencias de determinación de la pena dejarían de ser necesarias. Por ejemplo, uno de los requisitos más importantes es escuchar a las partes. Sin la *audiencia* de determinación de la pena, el acusado no puede declarar. Aunque en el programa pueden introducirse alegaciones en nombre del acusado (además de otras alegaciones e información), en el mejor de los casos se trata de un sustituto inaceptable de la declaración final del imputado antes de la sentencia condenatoria. Si el tribunal reduce las oportunidades de los acusados de ser oídos en la audiencia de determinación de la pena, ello acabará afectando a la percepción de la legitimidad de las condenas. Una audiencia oral de determinación de la pena es, por tanto, un elemento indispensable del proceso de sentencia. También brinda a las partes y a la víctima la oportunidad de interactuar e intercambiar perspectivas de una forma que resulta imposible si la sentencia condenatoria la determina un algoritmo.

En resumen, en esta presentación defenderé que la inteligencia artificial debería complementar y no suplantar a los jueces a la hora de dictar las sentencias. La inteligencia artificial puede contribuir de manera significativa a mejorar los valores del Estado de Derecho en las sentencias condenatorias (aunque no mediante la sustitución de las decisiones humanas). En su lugar, la inteligencia artificial debería aplicarse para controlar las prácticas a la hora de dictar sentencias y emitir los fallos, tarea que puede desempeñar mejor que los investigadores humanos. Los resultados de estos controles pueden ser utilizados por las comisiones de sentencia que elaboran directrices, los tribunales de apelación que recopilan fallos a modo de guía y los responsables de dictar sentencias en los tribunales de primera instancia. Como resultado, las penas cumplirán de manera más estricta los requisitos del Estado de Derecho que correspondan. Es necesario ampliar el papel de la inteligencia artificial dentro de la emisión de sentencias condenatorias, aunque los jueces han de mantener esta función como una actividad esencialmente humana.

La justicia predictiva y los objetivos de las sentencias condenatorias en Inglaterra y Gales

Elizabeth Tiarks

La transparencia de las sentencias condenatorias es importante para preservar la legitimidad penal y mejorar la confianza de los ciudadanos en las penas. En esta ponencia consideraremos el impacto que tienen las evaluaciones de riesgos predictivas en la transparencia de las penas tomando como ejemplo una parte específica del proceso de sentencia de la jurisdicción de Inglaterra y Gales: la elección del juez de qué objetivo de la pena perseguir.

Legalmente, en Inglaterra y Gales las penas pueden tener cinco objetivos: sancionar al culpable, reducir los delitos (incluso mediante disuasión), reformar y rehabilitar al culpable, proteger a la ciudadanía y reparar los daños y perjuicios. Estos objetivos pueden entrar en conflicto y no ser satisfechos en su totalidad cuando se condena a un infractor. No existe una jerarquía expresa entre ellos, por lo que es decisión del juez optar por uno u otro en cada caso. Esto genera cierta opacidad en esta parte del proceso, ya que a menudo no suelen quedar claros los motivos por los que un juez se decanta por un determinado objetivo a la hora de dictar sentencia. A pesar de que se han adoptado medidas dirigidas a la estructuración del proceso de sentencia que fomentan un enfoque coherente (en forma de directrices de sentencia), sigue faltando transparencia con respecto a la forma en la que se toma la decisión en lo que al objetivo de la pena se refiere.

En esta ponencia analizaremos cómo la falta de transparencia en esta parte del proceso de sentencia se ve afectada por el uso de algoritmos predictivos en las sentencias condenatorias en Inglaterra y Gales. El sistema de evaluación de infractores ("*Offender Assessment System*" – OASys) y sus componentes algorítmicos, como la escala de reincidencia de grupos de delincuentes ("Offender Group Reconviction Scale" – OGRS), ofrecen una puntuación predictiva de los riesgos de daños y reincidencia, que se utiliza para tomar decisiones sobre la libertad condicional y las condenas. Esta puntuación puede influir en el objetivo de la pena elegido por un juez, aunque no queda claro de qué forma ni en qué medida.

Los riesgos de daños y reincidencia son presumiblemente más adecuados para el objetivo de "proteger a la ciudadanía", por lo que podría fomentarse que los jueces se centraran más en este objetivo de las penas. Sin embargo, también es posible que los jueces hagan uso de las evaluaciones de riesgos para justificar alguno de los otros cuatro objetivos porque quizá consideran que una puntuación baja del riesgo es indicativa de la idoneidad del objetivo de "reformar y rehabilitar". Este asunto se complica aún más cuando existen diferencias entre los jueces en cuanto al nivel de confianza en los algoritmos predictivos y a la disposición que estos presentan a la hora de tener en cuenta o desechar la puntuación de los riesgos. En definitiva, es difícil conocer con precisión de qué forma afectan las evaluaciones de riesgos predictivas las decisiones de los jueces. Defenderemos que aumentan la incertidumbre sobre el proceso por el cual se eligen distintos objetivos en las penas, lo que agrava el problema existente y disminuye la transparencia.

¿Qué tan compatibles son las sentencias algorítmicas con la noción de culpabilidad y con el derecho a ser oído?: Alcances desde la doctrina alemana

Linus Ensel

Al analizar la imposición de penas por parte de los tribunales judiciales de las diferentes regiones de Alemania, se identifican variaciones y fluctuaciones importantes entre las mismas. Asimismo, se aprecia un déficit en el razonamiento de las decisiones judiciales que justifican aquellas sanciones. Dada esta realidad, resulta coherente abogar por la importancia de una racionalización del proceso de

determinación judicial de la pena. Una forma de contribuir con dicha racionalización puede ocurrir a través del empleo de las tecnologías de la información. En la presente ponencia, comentaré dos enfoques al respecto que me parecen pertinentes para este proceso de racionalización.

Una primera manera de racionalizar el proceso de determinación judicial de la pena sería mediante un sistema que incorpore *"machine learning"* (enfoque ML). Este enfoque implicaría que se ingrese la información de un gran número de expedientes judiciales a un sistema digital, de tal forma que el algoritmo "aprenda" la correlación entre los hechos de los casos y las decisiones judiciales emitidas al respecto. En la aplicación de este enfoque, el juez ingresaría al sistema digital los hechos relevantes del caso sobre el cual le corresponde pronunciarse (*input data*) y recibiría como respuesta del sistema una pena media agregada (*output data*).

Alternativamente, el segundo enfoque propone la determinación de las penas de manera reglamentaria a través de las tecnologías de la información, y su incorporación a un algoritmo de "no aprendizaje" – *"non-learning algorithm"* (enfoque NLA). Esto significaría que, cuando el juzgado ingresa al sistema digital los hechos de un caso específico, este algoritmo *calcularía* una condena determinada o un rango de penas.

Ambos enfoques descritos son concebibles tanto en una variante totalmente automatizada (*"fully automated"* – FA), como en el marco de un sistema de apoyo a la toma de decisiones (*"decision support system"* – DSS). La primera variante implica que las tecnologías de la información determinan la pena a ser impuesta, mientras que la segunda significa que estas fungen como apoyo al juez, pero se requiere de una decisión final humana.

Cabe señalar que la aplicación de cualquiera de estos enfoques de uso de las tecnologías de la información (ML/NLA), en cualquiera de sus variantes (FA/DSS), se enfrentaría a diversos obstáculos jurídicos, éticos y técnicos. A continuación, presentaré y analizaré dos de estos obstáculos jurídicos que plantean retos a nivel constitucional.

De conformidad con el Tribunal Constitucional Federal de Alemania (*Bundesverfassungsgericht*), y como es reconocido en diversas Constituciones a nivel internacional, el principio de culpabilidad posee rango constitucional. Esto se fundamenta en virtud del Estado Constitucional de Derecho (art. 20 apdo. 3 de la Constitución alemana, *Grundgesetz*, "GG") y de la garantía de la dignidad humana (art. 1 apdo. 1 GG). En el contexto de los procesos penales, esta garantía incluye la prohibición de degradar a una persona a un mero objeto del proceso judicial. En ese sentido, al menos en el delicado ámbito de las sanciones penales, una variante totalmente automatizada violaría esta garantía y, por tanto, el principio de culpabilidad. Incluso con la variante DSS, el sesgo humano de aceptar decisiones automatizadas (*"automation bias"*) podría conducir potencialmente a decisiones no probadas de facto. Para evitar ello, deben definirse claramente las funciones del juez humano y del proceso decisorio final, posterior a la decisión emitida por parte de las tecnologías de la información.

Además, en base al principio de culpabilidad, la pena impuesta debe ser "específicamente proporcional a la gravedad del delito y al grado de culpabilidad de la persona perpetradora". Al respecto, la legislatura alemana ha señalado que la evaluación de la culpabilidad no es una "constatación científica estricta", sino un "proceso de evaluación moral en el seno de la comunidad jurídica". Algunos se refieren a la misma incluso como un "acto de estructuración social". En esa línea, argumentaré que el enfoque comparativo de imposición de penas de un sistema de *machine learning* no sería coherente con esta concepción de la culpabilidad.

El segundo obstáculo que enfrentan los enfoques ML y NLA es respecto del derecho constitucional a ser oído, garantizado en el Art. 103 apdo. 1 GG. Este derecho también se deriva del Estado Constitucional de Derecho (art. 20 apdo. 3 GG) y de la garantía de la dignidad humana (art. 1 apdo. 1 GG). El artículo 103 apdo. 1 GG contiene el derecho de la persona acusada a expresarse ante el tribunal y el derecho a que su dicho sea tenido en cuenta. Sin embargo, estas dos manifestaciones tendrían poco valor si la persona acusada no pudiera informarse previamente de todos los hechos relevantes para la decisión judicial. Por lo tanto, el derecho a ser oído implica también la existencia de transparencia. Al respecto, en la ponencia examinaré las cuestiones relativas a la transparencia en los enfoques de ML y de NLA, y concluiré que un enfoque de ML no cumpliría los requisitos relativos a la explicabilidad del sistema, al menos no con los medios tecnológicos disponibles en la actualidad. La aplicación de un enfoque NLA, por otra parte, sería compatible con el derecho a ser oído si la ruta de cálculo del algoritmo puede ser puesta a disposición y se pueden considerar todos los aspectos relevantes del caso cuando se toma la decisión humana final.