Original Research Article

The datafication revolution in criminal justice: An empirical exploration of frames portraying data-driven technologies for crime prevention and control

Anita Lavorgna (iD) **and Pamela Ugwudike** (iD)

Abstract

The proliferation of big data analytics in criminal justice suggests that there are positive frames and imaginaries legitimising them and depicting them as the panacea for efficient crime control. Criminological and criminal justice scholarship has paid insufficient attention to these frames and their accompanying narratives. To address the gap created by the lack of theoretical and empirical insight in this area, this article draws on a study that systematically reviewed and compared multidisciplinary academic abstracts on the data-driven tools now shaping decision-making across several justice systems. Using insights distilled from the study, the article proposes three frames (optimistic, neutral, oppositional) for understanding how the technologies are portrayed. Inherent in the frames are a set of narratives emphasising their ostensible status as vital crime control mechanisms. These narratives obfuscate the harms of data-driven technologies and evince idealistic imaginaries of their capabilities. The narratives are bolstered by unequal structural arrangements, specifically the unevenly distributed digital capital with which some are empowered to participate in technology development for criminal justice application and other forms of penal governance. In unravelling these issues, the article advances current understanding of the dynamics that sustain the depiction of data-driven technologies as prime crime prevention and law enforcement tools.

Keywords

Dataveillance, datafication, big data, crime control, digital capital

Introduction

Datafication is a rapidly proliferating phenomenon in contemporary Western societies, and it refers to 'the ubiquitous quantification of social life' (Baack, 2015: 2). It manifests itself mainly in big data analytics, and it stems from the so-called data revolution (United Nations, 2014) of the past few decades. Datafication is also a by-product of the accumulation of large-scale administrative data and of behavioural and social data from heightened human interaction with smart devices and online platforms including social networking sites (Boyd and Crawford, 2012), and it is increasingly used to inform decision-making within and beyond the public, political and economic sectors.

There is a fast-growing body of critical literature on the proliferation of datafication in the form of the data-driven models now increasingly applied in justice systems across the world for predicting risk (Angwin and Larson, 2016), forecasting crime hotspots (Ensign et al., 2018; Lum and

Isaac, 2016) and implementing the biometric identification of targeted individuals (Bennett Moses and Chan, 2018; Fussey, 2019). The rapid rise of data-driven technologies in justice systems highlights their growing popularity as prime crime control tools despite the growing corpus of studies pointing to their potential harms (Barocas and Selbst, 2016; Ferguson, 2017; Fussey, 2019; Hao and Stray, 2019; Oswald et al., 2018). This calls for close scholarly scrutiny of how the technologies are framed and presented to the state and the general public, both of whom

Department of Sociology, Social Policy and Criminology, University of Southampton, Southampton, UK

Corresponding author:
Anita Lavorgna, Department of Sociology, Social Policy and Criminology, University of Southampton, Murray Building, 58 Salisbury Rd, Southampton SO171BJ, UK.
Email: a.lavorgna@soton.ac.uk

may be considered the core stakeholders. Therefore, this article provides an empirical account of the frames that appear to legitimise the technologies. This is explored through two fundamental concepts: utopic 'sociotechnical imaginaries' (Jasanoff et al., 2007), which emphasise the benefits of datafication whilst obscuring its harms, and 'digital capital' (Van Dijk, 2005), which refers to the resources required for technology development. Motivated actors with access to such capital can dominate technology development and the truth claims made about their capabilities.

To fulfil its objectives, the article draws on the findings of a narrative review of abstracts from academic articles on the technologies and their capabilities. Using insights gained from the narrative review, we propose the following frames for understanding how the capabilities of data-driven technologies are depicted in the extant literature: (a) optimistic frames which endorse the tools and their ostensible status as the panacea for cost-effective and efficient crime control; (b) neutral frames that are ambivalent towards the potential harms and benefits of datafication although some are solutionist in that they acknowledge the potential harms of datafication and proffer remedial artificial intelligence (AI) models and (c) oppositional frames which emphasise several harms of datafication and reject the view that data-driven tools constitute the panacea for crime control.

Datafication and the proliferation of data-driven technologies

Data-driven technologies that have emerged as part of the datafication phenomenon are being deployed for a range of tasks. Some examples are social media mining, sentiment analysis and natural language processing. These technologies now inform decision-making at several levels of the criminal justice process. For example, experimental algorithmic tools can influence police operational decisions to improve a police force's decision-making, such as how to allocate policing resources (Mastrobuoni, 2020; Oswald et al., 2018). In many countries, police services are also using new data-driven technologies for intelligence and surveillance-led policing (Bennett Moses and Chan, 2018). The technologies, which are sometimes produced through collaborations between academics, the state and non-state actors, are also influencing other criminal justice services. In court, they can inform decisions about sentence severity, and in the penal system, they can determine the intensity of interventions, parole decisions and prison security classifications (e.g. MOJ, 2019). To perform their tasks, the technologies automate a combination of administrative data from criminal justice services and/or big data from other sources.

The private sector and academic researchers (particularly data and computer scientists, software developers as well as mathematicians and statisticians) are at the forefront of the datafication phenomenon (e.g. Brantingham et al., 2018; Mohler et al., 2015). Their frames about the technologies can evoke and legitimise sociotechnical imaginaries (Jasanoff et al., 2007; Jasanoff, 2015) of criminal justice technologies as the panacea for crime control. The concept of sociotechnical imaginaries is defined as 'collectively held and performed visions of desirable futures […] animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology' (Jasanoff, 2015: 25). As such, these imaginaries depart from expectations of innovation of science and technology (Jasanoff et al., 2007); however, they do not only embed visions of technological possibilities, but they also express a shared understanding of the common good (Jasanoff, 2015).

Therefore, imaginaries of 'smart' criminal justice can be held not only by the proponents and developers of the technologies but also by state consumers of technological products and even ordinary members of the public (Jasanoff et al., 2007). But, collectively held imaginaries are informed by specific framings and visions of the benefits of technological development that can obfuscate technological harms, as frames organise knowledge and influence thinking (Borah, 2011; Entman, 1993). An example is the capacity for data-driven technologies to reproduce biases embedded in foundational data. In replicating biases, the technologies connect the past and the future and, in the process, sustain a social order marked by historical inequalities. As Jewitt et al. (2015: 90) put it, such imaginaries can influence how people 'see and organise the world, in its histories as well as its futures'. The frames should as such be investigated since they are not simple visualisations of future possibilities: they can inform the design of technologies that either rebalance or maintain an unequal social order and can fuel expansionary applications of the technologies by justice systems.

This trend overlooks the caveats proffered by critical data scientists who emphasise the potential for big data analytics to rely on biased administrative data that prompt algorithmic replications of existing inequalities (e.g. Starr, 2014). Notwithstanding this, the criminal justice system is becoming increasingly datafied. For example, more than 150 data-driven predictive algorithms are now applied in penal systems across the world, including low- and middle-income countries (Fazel et al., 2012). Yet, even if academic attention on this issue is certainly rising (for instance in the legal discipline, see among others Custers, 2013; Hamilton, 2015; Law Society, 2019; Starr, 2014), criminological and criminal justice scholarship has remained relatively silent on this subject matter. The concept of sociotechnical imaginaries is therefore useful since it alerts us to the need for critical analysis of the frames that instigate and sustain imaginaries of 'smart' criminal justice or datafication as the panacea for crime control.

Reflecting this perception, in several countries, there have been official declarations about the merits of data-driven predictive algorithms and their importance in policing and criminal justice decision-making (only a few examples: in Australia see AFP, 2021; in the United Kingdom see MOJ, 2019; in France, see Politico, 2020). Authorities using predictive policing software similarly evoke the sociotechnical imaginaries depicting the tools as capable of achieving crime prevention aims, stressing their technocratic benefits such their ability to provide efficient and cost-effective solutions to the problems of crime and deviance (e.g. Dearden, 2017; Kelion, 2019 MOJ, 2015; National Offences Management Service, 2016). The technologies are also generally depicted in policy documents as scientifically objective and capable of a higher degree of accuracy and impartiality than human actors or when combined with professional judgement (Home Office, 2002). These presumed benefits denote the influence of positive frames and sociotechnical imaginaries regarding the capacity of technologies to improve human decision-making. It appears that equal consideration is not given to technological limitations. However, speculation aside, empirical insight is needed to unravel the frames and their implications.

The critical scholarship on data-driven technologies

Rejecting the positive depictions of datafication inspired by sociotechnical imaginaries, critical research has underlined how big data are not easy to manage, especially when potentially important data need to be identified and prioritised amongst a white noise of basic rumours and potentially misleading information (Middleton et al., 2020). In addition, there are no certainties provided by big data analyses, but only projections of patterns based on computational models. Indeed, a large corpus of research highlighting the social harms of the technologies has emerged. Central to this body of work is the view that as data-driven technologies continue to proliferate, their depiction by their academic and commercial developers as the panacea for crime control seems to obscure their challenges. This fast-growing scholarship is sometimes described collectively as critical data studies. Moving away from the earlier myopic idealism of certain sociotechnical imaginaries that have long surrounded the emergence of AI (Natale and Ballatore, 2020), this scholarship is increasingly recognising the harms of big data when automated by algorithms across the public and private sectors. Insights into these harms have emerged from various disciplines including criminology (Hannah-Moffat, 2018; Ugwudike, 2020), legal studies (Custers, 2012, 2013; Hamilton, 2015; Starr, 2014), media and cultural studies (Dencik et al., 2018; Iliadis and Russo, 2016), and computer science (Hao and Stray, 2019).

What emerges from these fields of study is that technologies designed by academics and others sometimes in collaboration with state authorities (such as the police) and sometimes with commercial motives can reproduce existing structural inequalities for several reasons. One is that data (in)justice (i.e. the social harms and benefits of datafication) can depend on socioeconomic status (Dencik et al., 2018; Taylor, 2017; Završnik, 2019). Importantly, the Bourdieu-inspired concept of digital capital (Bourdieu, 1986) can help us understand the uneven distribution of technological harms and benefits. From a sociological perspective, digital capital can be defined as the ability to not only access digital techs but to also design and capitalise a range of resources linked to digital technology (Van Dijk, 2005). Depending on the technology, digital capital, particularly via design and commercialisation capabilities, tends to be endowed in wealthy corporations and other elites such as academics and researchers. Perhaps relatedly, these groups also seem largely immune from technological harms.

In contrast, the socioeconomically marginal seems more vulnerable to harms such as 'dataveillance', that is, 'a form of continuous surveillance' (van Dijck, 2014: 198) using digital methods as has been noted by studies of the surveillance of defendants and convicted people (Willoughby and Nellis, 2016). Groups such as ethnic minorities and the disabled are also more vulnerable than others to misidentification given their noted underrepresentation in the data on which AI surveillance and identification systems rely (Law Society, 2019). Furthermore, the socioeconomically marginal and particularly those amongst them who are minorities can experience higher rates of policing activity in their residential area. This can be triggered by predictive policing algorithms that are prompted by racially biased arrest data to designate their geographical region as 'high crime' and in need of enhanced policing activity (Ferguson, 2017; Harcourt, 2015; Lum and Isaac, 2016). Minorities are also more vulnerable than others to overprediction of recidivism risks by algorithms relying on similar data (Angwin and Larson, 2016; Eaglin, 2017; Hannah-Moffat, 2018; Hao and Stray, 2019; Ugwudike, 2020).

Given these trends, it has been argued that data harms are proliferating at the same rate as the technologies exploiting the increased availability of big data, but knowledge of the harms and how to attenuate them is not (Taylor, 2017). An upshot of this is that the general public and the services relying on the data-driven algorithms can overestimate the veracity of the technologies (Hollywood et al., 2018). Positive frames depicting the tools as the panacea for crime control can exacerbate this problem. But there is a dearth of research on how the technologies are framed. To address the gap in knowledge, this article unravels the dominant frames and their implications, shedding light on how they are conjoined to issues of digital capital and sociotechnical imaginaries.

Methodology

Our narrative review of abstracts sought to advance understandings of the frames legitimising the datafication of criminal justice by answering the following questions: (a) how are the capabilities of datafication and data-driven technologies for criminal justice framed and presented to potential stakeholders and the general public alike? (b) what explains the different frames identified and (c) what are the implications?

Data search criteria

We focused on academic literature retrieved through the citation indexing service Web of Science (WoS), which indexes the text and metadata of most peer-reviewed academic journals, books, abstracts, conference papers, reports and other relevant literature. We looked for 'topics', a category which encompasses titles, abstracts, authors and keywords. Search terms used were as follows: TOPIC: ('big data' AND 'criminal justice') OR ('big data' AND policing) OR (mining AND 'criminal justice') OR (mining AND policing) OR (artificial AND 'criminal justice') OR (artificial AND policing) OR (machine AND 'criminal justice') OR (machine AND policing) OR (algorithm* AND 'criminal justice') OR (algorithm* AND policing). The keyword search returned 1773 results, which became 1330 once we refined the search using the exclusion criteria set out further below. We exported all titles, abstracts and associated metadata in our search results. The abstracts were then manually screened again by two researchers to enhance interrater reliability. The aim was to assess relevance since keyword searches notoriously produce false positives. Following this manual screening, a total of 493 abstracts were selected for in-depth analysis.

Data exclusion criteria and research limitations

We focused on material published in the last 11 years (January 2009 to December 2019) to reflect the period that has witnessed the rapid expansion of the big data phenomenon. Also, to make qualitative analysis manageable, our searches focused on terms commonly used in the academic literature on datafication and data-driven technologies as they pertain to criminal justice.

We recognise that the dataset used has a number of limitations. First, it is worth noting that a limitation of our data gathering approach is that WoS indexes the text and metadata of most relevant material, but some journals, books and conferences, for instance, might be overlooked. In this regard, we consider our study to be exploratory in nature and useful to start a discussion on sociotechnical imaginaries and digital capital in the context of data-driven technologies for crime prevention and control, not to be an end point. Whilst it is important to note that bibliometric evidence suggests that the coverage of WoS databases is competitive in the social sciences (Martín-Martín et al., 2018), additional research should consider other databases, to ensure a more comprehensive coverage. Also, we recognise that looking at abstracts does not tell us the whole story (especially considering that abstracts in different disciplines are written in quite different ways), and a proper meta-analysis on the topic here addressed would be beneficial. Yet, as recognised in the literature, abstracts are of fundamental importance for screening academic publications, to the point that the necessity and urgency to improve abstract reporting have been broadly discussed (see e.g. Guo and Iribarren, 2014; Saint et al., 2000) – as such, we believe they offer crucial information for our analysis.

Another limitation of this study that needs to be acknowledged is that we considered only studies and reports published in English. Even if English has become the *lingua franca* for researchers, we acknowledge that our language restriction (which was imposed by resource limitations) might lead to some biases in the results, for instance we might have overlooked important research of more local or regional interest.

Data analysis

We used a thematic synthesis approach to analyse the data and present the findings. First, the data were uploaded to NVivo 11 (a data analysis software package to manage and arrange unstructured information) for coding. We used NVivo to systematically organise and analyse the data manually sampled from WoS, and we applied codebook thematic analysis which is a structured approach to coding that incorporates reflexive elements (Braun et al., 2019). In line with this approach, the main themes (conceptualised as domain summaries, or main 'nodes' in NVivo) and most sub-level nodes (child nodes) were determined in advance of full analysis, but they were then compared and revisited within the context of the abstracts as analysis progressed (see Table 1 for a summary of the nodes). Of course, a certain degree of overlap exists between some of the sub-codes identified (consider, for instance, the division by discipline); in creating categories, we aimed for a good balance between being sufficiently fine-grained for analytical purposes whilst maintaining sufficient accuracy considering the information available in the WoS database. Besides assisting us with qualitative analysis, NVivo allowed us to count the number of times each code was used across the sampled abstracts. An audit trail was maintained throughout, and the two authors worked collaboratively on data coding and interpretation.

Findings

First, we looked at the year of publication (as reported in WoS) to understand the evolution of relevant scholarship

Table 1. Coding framework.

Nodes (and child nodes)

Year (of publication)
Location
Attitude (optimistic, neutral, oppositional)
Discipline (Computer Science or Engineering, Social Sciences (including Criminology), Law, Geography, Economics (including Business and Management), Politics, Health (including Legal Medicine), other)
Technological development (existing, under development or promissory, unclear)
Criminal Justice link (direct link, indirect link)
Criminal Justice sector (Criminal Justice in general, policing, courts and sentencing, prison, probation, youth justice)
Other (discussion of racism, discussion of class-related issues, discussion of gender-related issues, discussion of civil liberties and privacy rights, problems of compliance with copyright law, problems of accountability, transparency is needed for trust, other risks associated with datafication (in general), can create an 'unbalance' towards easy cases, idea that human interpretation is something negative, police online engagement, role of science fiction and media social construction, poor understanding of criminological/criminal justice scholarship emerges from the language)

during the timeframe of interest As shown in Figure 1, research on datafication and the accompanying rise in data-driven technologies is a fast-growing field, and this article is timely given its evaluation of key frames emerging from the scholarship.

Second, we looked for any geographical reference present in the abstract detailing, for example, where the data-driven tool was developed or tested, or the origin of the databases used. A specific location was referred to only in 147 publications. This is problematic since the elision of locational information implies broad generalisability of results and a tendency to overestimate the scope of data-driven technologies. Worryingly, studies now show that such technologies are context specific, and both accuracy and fairness can be undermined if applied in other jurisdictions and contexts (Olver et al., 2014). For tools to have broad applicability, they have to be tested and validated for use beyond the location in which they were created.

Most of the abstracts focused on technologies that have a direct link to criminal justice (355) meaning that the data-driven technology described has been developed or discussed with specific reference to the criminal justice system. The remaining 138 abstracts focused on other uses and sectors but mentioned criminal justice as a potential application. Considering both the direct and the indirect links, most abstracts were directly relevant to policing (including forensics and urban safety), as shown in Figure 2 below (please note that the sum is slightly more than 493 as some publications mentioned/referred to more than one criminal justice sector).

Only a minority of abstracts, however, gave emphasis to a number of key datafication-related problems, as presented in the review of the literature above. More in detail, only a limited number of abstracts referred to potential issues of racism (12), of class-related issues (2) or gender-related issues (2). Similarly, only a few abstracts mentioned explicitly issues of civil liberties and privacy rights (14), problems of compliance with copyright law (2), problems of accountability (5) or generally mentioned the existence of risks associated with datafication (7). Only one abstract stressed that, when applied to criminal justice, datafication could create an 'unbalance' towards easier cases,
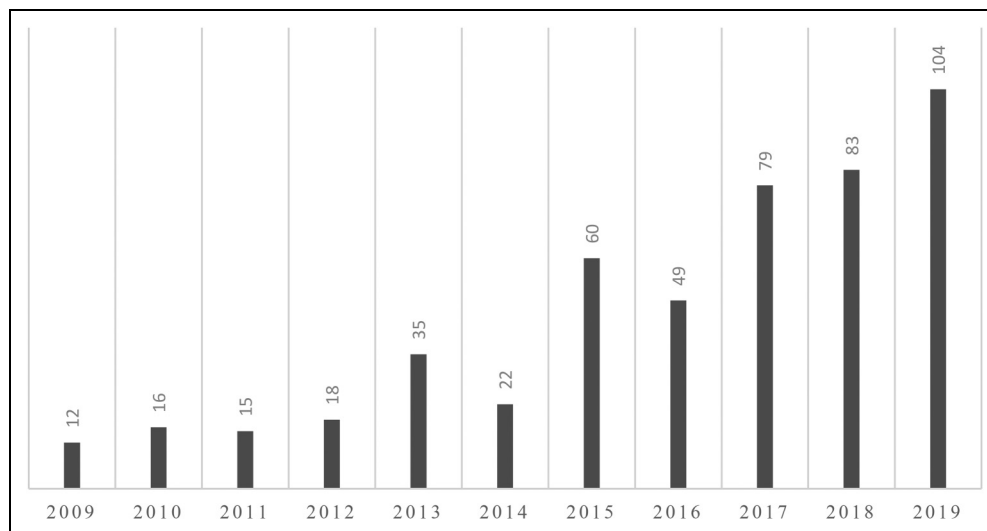
Figure 1. Number of publications per year.

Figure 2. Criminal justice sectors.

and only one abstract focused on the role of science fiction and social media construction in framing public understanding of the topic.

Increased transparency in big data and their use was suggested as potential solution to increase public trust in datafication in a few studies (4). Quite worryingly, it was suggested that human interpretation is somehow something negative, a problem to be 'fixed' with data-driven technologies (6) – ignoring that data 'do not speak for themselves' (see González-Bailón, 2013) – and the language used in a few abstracts (11) suggested a poor understanding of relevant criminal justice/critical scholarship. This latter fact is probably not surprising, considering that most of the abstract considered emanated from computer science and engineering, with very little multi- or interdisciplinary effort (see Figure 3, which represent the macro-categories of the prevalent disciplines, depending on the specific Journal, Proceedings or else where the study was published). The amount of publications in computer sciences and engineering could also be partially explained by a different publication culture in those disciplines, where it is more common to have papers published by groups of authors.

We also considered the level of technological development, categorising abstracts depending on whether the data-driven technology described was in use (i.e. the publication was discussing/presenting a tool already in use, or the tool has been developed to an extent that it has been tested/trialled in real life) or not in the criminal justice system at the time of writing. As emerges from Figure 4, most of the technologies discussed were under development, and in a number of publications, the phase of development was not clear (e.g. when data-driven technology was discussed in general, without focussing on a specific tool).

Categorising and analysing the frames

How are the capabilities of data-driven technologies framed?

A final code considered was the 'attitude' one – that is, how data-driven technologies were prevalently portrayed in the abstract. Our findings show that the way the technologies and their capabilities are framed can be dimensionalised along three categories: *optimistic, neutral, and oppositional*. Further analysis of these frames revealed overwhelmingly optimistic framings: 379 of the abstracts depicted the tools positively and were overall optimistic about their capabilities and futures, 73 were more nuanced abstracts written with a neutral tone and only 41 used oppositional frames. Some overlap exists between some of the 'attitudes' identified in the same abstract, and between the various frames we will present throughout this section: we should stress that our proposed categorisations do not aim to pigeonhole abstracts into a certain category whilst precluding them from fitting elsewhere, nor to compartment attitudes into single boxes, but simply to provide a framework to discuss the most important trends and patterns identified.

Please note that the following sections provide, in quotation marks, references to the abstracts analyzed. As it would have been impossible to discuss them all within the word limit allowed, we decided not to single out any of them: in our work, we want to highlight some trends and patterns that we think are of great interest in the current debates on data-driven technologies for crime prevention and control, not to debate the merits of specific articles. We also focus on frames relating to the use of data-driven technologies by criminal justice services. The frames may also be relevant in other contexts and to other actors deploying such
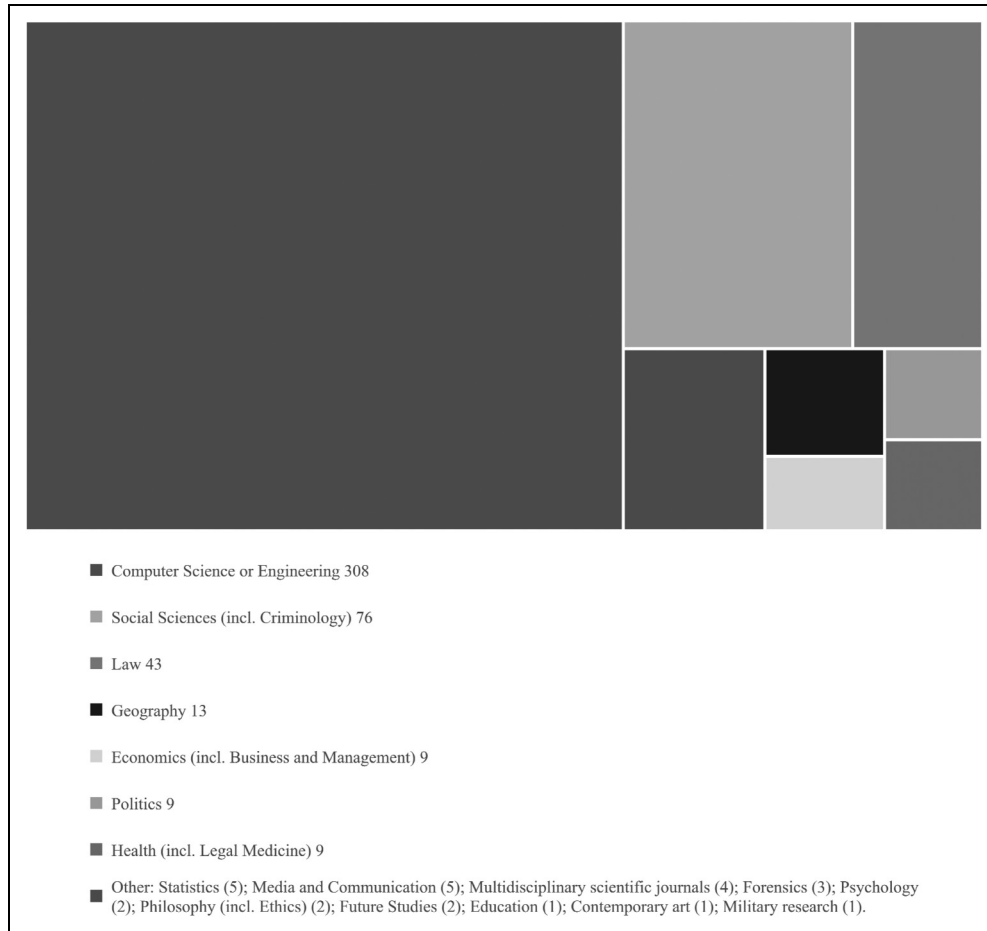
- ■ Computer Science or Engineering 308
- ■ Social Sciences (incl. Criminology) 76
- ■ Law 43
- ■ Geography 13
- ■ Economics (incl. Business and Management) 9
- ■ Politics 9
- ■ Health (incl. Legal Medicine) 9
- ■ Other: Statistics (5); Media and Communication (5); Multidisciplinary scientific journals (4); Forensics (3); Psychology (2); Philosophy (incl. Ethics) (2); Future Studies (2); Education (1); Contemporary art (1); Military research (1).

Figure 3. Disciplines.

technologies in other sectors, and a useful contribution of this article is that it provides frames that can be explored and expanded on in other contexts by future studies.
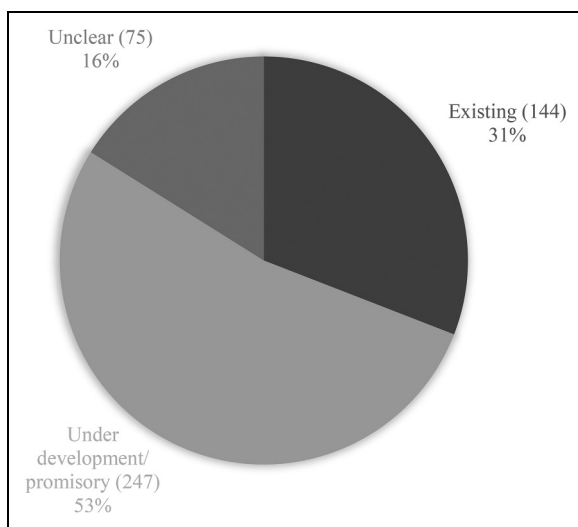

Figure 4. Level of technological development.

Optimistic frames

These were, by far, the most prominent and frequent frames. They depicted the rise of datafication in the criminal justice systems positively and encouraged the automation of 'rich' sources of data 'now readily available', indicating that this would enhance the efficiency and effectiveness of criminal justice. In general, optimistic frames evinced utopic sociotechnical imaginaries, with technoscientific innovations 'encoding visions of the good society' (Jasanoff et al., 2007: 6). They presented AI models as the future or the 'next generation criminal justice' that will provide 'rich' services to the courts and law enforcement and 'power' these services in the future. The key optimistic frames identified were, as such, primarily utopic: *scientific objectivity, technological superiority and vital utility*.

Scientific objectivity. In terms of their capabilities specifically, optimistic frames portrayed the technologies as scientific and capable of attaining levels of objectivity that were beyond human capability, thus indicating what is socially desirable through the use of technology – that is, the minimisation of human intervention, seen as

ineffective or biased. Encapsulating this trend, for instance, one abstract presented a set of metrics to refute the claim that clinical risk predictions were just as effective as algorithmic versions. The tools were described as 'novel' models with the capacity to 'outperform' the limits of human competency. Allusions were also made to the 'scientific value' of 'innovative' big data-driven AI models. This sheds some light on why state authorities that adopt the tools echo similar sentiments (e.g. MOJ, 2019). Thus, such frames seem to provide the basis on which criminal justice services justify the deployment of techs such as predictive policing and risk prediction algorithms. Critics of the frames would argue that they also reflect what Broussard (2018) conceptualses as a 'technochauvinist' belief in the efficiency and scientific neutrality of techs.

Technological superiority. Aligned with the claims made about their scientific objectivity, optimistic frames depicted data-driven technologies as superior to human competencies. Adjectives depicting the tools as high-capability devices were also used in profusion. For example, they were described as 'powerful toolkits' for administering criminal justice and 'optimise procedures', 'smart systems', 'state-of-the-art' algorithms. References were made to the 'high recognition results' of facial recognition software. Models were described as capable of 'precisely forecasting' hot spots. Frames such as 'computationally efficient', 'automatically detect' and 'reliably' were used to describe the capacity of predictive policing models. Much attention was given to the 'advantages' of the models, and sometimes multiple advantages were mentioned including improved prediction, detection and apprehension, hence configuring a specific idea of an advisable social future and setting expectations associated with it. It is not surprising then that police spokespeople can describe data-driven technologies such as predictive policing software as capable of providing 'greater efficiency and effectiveness' (Dearden, 2017).

Vital utility. Frames depicting the models as vital criminal justice tools were prominent in this category. Thus, the tools were described as 'critical for safety and public security', 'crucial' for 'choosing intervention strategies' and for 'evidence-based decision-making' across sectors, and 'rich and intuitive' models capable of providing 'accurate information' with 'great practical value' for crime prevention. Deep-learning models were described as 'fine-tuned' and 'gold standard' tools for expediting crime prevention tasks and 'improving public security'. There were also frames highlighting technocratic benefits for hegemonic social actors, such as 'improved payoffs for the police' and 'prioritising' crime prevention resources. Frames such as 'technology innovation' were used to describe models capable of 'transforming policing', 'upgrading policing' or 'transforming government' by improving efficiency and effectiveness. Alongside the technocratic benefits, these frames also draw attention to

additional managerial benefits of data-driven technologies, particularly their role in enhancing police productivity and public opinion about police legitimacy.

Neutral frames

Neutral frames were non-committal in their analysis of the harms or benefits of datafication. These frames evoked a technorealism that vacillated between utopic and dystopic imaginaries and fell into three categories: *ambivalent*, *informational* and *solutionist*

Ambivalent frames. Most of the neutral frames were nuanced and evoked considerable ambivalence and fluctuation between optimism and pessimism in envisioning social and technoscientific orders, for instance recognising how certain approaches could also generate 'negative externalities'. One abstract, for example, questioned whether decision-making technologies are capable of transforming criminal justice positively and emphasised the need for 'cautious optimism'; another one stressed the need for more 'cautious communication' to the public on the role of certain technologies in possible reforms.

Informational frames. Neutral frames highlighted the potential and limits of data-driven technologies, including the capacity of the technologies to produce either positive outputs (e.g. race-neutral risk scores) or negative outputs (e.g. biased risk scores and invasions of privacy). Reflecting their emphasis on the need for information about the limits and potential of datafication, neutral frames decried the lack of independent research on big data analytics in justice systems. They called for 'independent tests' by researchers other than the developers to establish an 'evidence base' that would set out both the limits and potential of criminal technologies such as predictive policing algorithms. They as such acknowledged the 'allegiance effect' undermining existing research – that is, much of the evidence currently validating the efficiency of data-driven tools have been conducted by the developers themselves or those affiliated to them (Blair et al., 2008). Again, reflecting their informational focus, neutral frames stressed the need for research on technology adoption in real-world practice, emphasising that, if imagined futures justify policy changes and investments, more/better empirical investigation is needed. Such studies are also necessary given the emerging evidence of algorithmic aversion (Burton et al., 2019) and other forms of resistance by front-line practitioners (Sanders et al., 2015).

Solutionist frames. Some neutral frames did not focus solely on developing insights into the limits and potential of datafication in criminal justice. Instead, they evoked a realism that consisted of acknowledging limitations and proffering practical solutions. These solutionist frames were ambivalent about the possibility of creating fair and accurate algorithmic predictions but suggested that robust research could repair technical problems and enhance the

quality of their design and outputs, in this way evoking the aforementioned technochauvinism and a vision of future benefits stemming from technological innovation. Thus, the frames recognised potential barriers to efficient algorithmic output but optimistically proposed technical solutions (e.g. the optimisations of datasets and codes). Models that 'significantly improve' predictive accuracy were presented. It should be noted, however, that whilst mentions of high(er) percentages of accuracy rates were used to justify solutionist frames, from our dataset, it cannot be concluded that these hold up under further scrutiny. Solutions were also offered to maximises the crime control potential of the technologies (e.g. a framework to maximise the payoff of systems useful for facial recognition models for intelligence purposes). Additionally, some proffered fairness measures that could be developed to assess the broader social outcomes of criminal justice algorithms beyond fairness metrics that focus on technical design (e.g. an improved metric 'for considering racial equity concerns in algorithmic design').

Oppositional frames

Abstracts containing oppositional frames were fewer than those comprising the other two frames, but because the oppositional abstracts presented theoretical arguments unlike the largely empirical content of the other three, they produced richer insights. This explains why they had more dimensions (listed further below) than the other frames. Together, the oppositional frames generally rejected the sociotechnical imaginaries that emphasised technological potentialities, even when such rejection meant overlooking implications such as the operational need to consider new uses of data, or the possibility that uninformed or discretionary human decision-making can also have serious adverse consequences. As will be discussed more in detail in the following section, behind oppositional frames, we generally find more subject knowledge of criminal justice systems and sociolegal theorisations (intellectual and critical capital) but less evidence of digital capital.

Meanwhile, oppositional frames depicted the sociotechnical imaginaries as rooted in myths propagated by invested proponents and 'technological evangelists' as well as media exaggerations. In rejecting the imaginaries, the oppositional frames drew attention to the social impacts of datafication including problems stemming from data injustices and the need for robust governance. Thus, they revealed that critical analysis of sociotechnical imaginaries is useful for unravelling 'the implications of technology's social embeddedness for responsible global governance of both knowledge and technology' (Jasanoff et al., 2007). The key oppositional frames were as follows: privacy violations, racial bias, algorithmic opacity, fairness concerns, inadequate legislation, the commercialisation of datafied criminal justice and systemic bias.

Privacy violations. These frames focused on privacy issues arising from the datafication of criminal justice such as accessing, storing and using personal data without requisite consent, in a form of 'ubiquitous […] social control'. They also critiqued new trends in data linkage whereby public authorities develop unified databases and data-sharing protocols without the consent of data owners and with implications for data security and privacy – which can expose affected individuals to unwarranted state intervention (Dencik et al., 2018).

Racial bias. Oppositional frames refuted the technological superiority mooted by optimistic frames and pointed out errors produced by big data models when applied in criminal justice settings, leading to 'discriminatory' results. In particular, the presumption of neutrality was said to be harmful since it masked the capacity of the algorithms to produce, for instance, the replication of biases in 'racially biased' administrative data. Concepts such as 'algorithmic injustice' (Birhane, 2021) and 'coded bias' (AJL, 2021) have been used to describe this phenomenon. Social outcomes that can emerge from this include the stereotyping and labelling of affected groups as riskier or more threatening in other ways. Some oppositional frames also referred to the problem of algorithmic feedback loops whereby algorithms learn that linking proxies for race and other protected categories with criminality is accurate and proceeds to do so more frequently in future predictions – with the consequence that 'inequality' becomes 'inherent in all predictions'. This problem has been identified as a feature of some risk prediction technologies and predictive policing algorithms (Cossins, 2018; Ensign et al., 2018). These findings highlight how, as noted earlier, sociotechnical imaginaries of 'smart' criminal justice can mask technology harms including their role in reproducing inequalities, thereby sustaining an unequal social order.

Algorithmic opacity. Several oppositional frames alluded to the opaque operations of black-boxed algorithms. As several scholars suggest, this can undermine, for instance, the transparency of court proceedings and procedural due process (Citron, 2007; Citron and Pasquale, 2014; Steinbock, 2005; Zarsky, 2013) and accountability (Bennett Moses and Chan, 2018; Burrell, 2016; Fontaine, 2016). Indeed, there were several references to the potential for the algorithms to violate human rights, due process safeguards and civil liberties (e.g. algorithms allowing law enforcement officials to carry out procedures that would violate due process rights if algorithms were not applied). Storage of personal data without due permission and illegal searches were two of several examples. Relatedly, some frames emphasised that the hidden nature of black boxed decision-making meant that criminal justice officials

could presume that their outputs were neutral although they could in fact be 'infected with' biased data capable of provoking biased outputs – as mentioned above in our discussion about issues of racial bias.

Oppositional frames also underlined the need for laws such as freedom of information legislation requiring developers providing commercial algorithms for public sector application to release their algorithms for scrutiny. Nevertheless, proprietary laws impede access to algorithmic black boxes, and in doing so, they undermine transparency and accountability (Pasquale, 2016). In the United States for example, the Supreme Court has held in response to requests by a defendant to scrutinise one such algorithm that data-driven predictive technologies applied in penal systems are protected by proprietary rights (Harvard Law Review, 2017; HOC, 2018).

Fairness concerns. Racial and other biases embedded in administrative data were considered inimical to technical efforts by computer scientists and statisticians to achieve a fully effective balance between fairness and accuracy. Accompanying this was a rejection of the tendency to focus on technical solutions to fairness issues whilst overlooking the socioeconomic and sociopolitical contexts that shape technological design, application and outcomes. This echoes recent calls for those developing technical solutions to algorithmic bias and other harms to expand their focus to a broader consideration of structural solutions (Selbst et al., 2018).

Aligned with this, some oppositional frames implied that fairness imperatives cannot be achieved using technical means since the effects of biased administrative data on outputs (risk scores) and wider social outcomes (systemic discrimination) were not measurable. Besides, efforts to achieve fairness can jeopardise predictive accuracy. Ultimately, decision-makers have to subjectively determine the appropriate balance between fairness and accuracy, or even how fairness should be constructed and measured, a reality made clear by researchers such as Hao and Stray (2019).

Inadequate regulation. Some oppositional frames focused more on the need for laws requiring the ethical and fair design of AI models. Data provenance, quality, storage and usage were areas described as in need of robust regulation. Laws regulating relationships or interconnections between developers, users, regulators and the subjects of datafication were also proposed to ensure transparency and accountability. There was a sense that technology design, particularly the aspects relating to data automation, was fast outpacing the introduction of laws and regulations. More robust legal controls beyond ethics-related regulation were as such recommended.

Commercialisation of datafied criminal justice. Oppositional frames also suggested that optimistic frames ignored technological harms and bolstered the lucrative commercialisation of tools such as predictive policing algorithms. From this perspective, the optimistic frames overlooked the fact that developers with digital capital were able to dominate design choices and could as such influence algorithmic outputs. This draws attention to the harms of data injustice (Dencik et al., 2018), whereby developers with digital capital dominate algorithmic design and are less vulnerable than minorities to algorithmic harms (Taylor, 2017; Willoughby and Nellis, 2016).

Aligned with this was the view that the developers of commercial algorithms were being empowered to exercise the 'sovereign power' previously reserved for the state. As others have noted, this provides further impetus for non-state actors to participate in criminal justice policy and governance (Hannah-Moffat, 2018; Ugwudike, 2020). It calls attention to the politics of science and technology innovation since, as noted earlier, the developers exert power through their ability to inform technology design and criminal justice expenditures (Jasanoff, 2015).

Another issue dealt with by oppositional frames is the role of data-driven algorithms and their developers in knowledge production about social problems such as crime and risk. In drawing attention to this, oppositional frames echoed the concerns expressed by others who note that data-driven technologies, including the commercial versions, are arbitrarily reframing constructions of crime and deviance (Eaglin, 2017; Ugwudike, 2020). Here, again we witness how the concept of sociotechnical imaginaries can highlight the politics of science and technology innovation. In this case, whilst optimistic frames emphasised positive imaginaries, oppositional frames unravelled the meanings and social implications of technology innovation.

Systemic bias. Some oppositional frames depicted harms such as biased algorithmic outputs as irreducible to technological limitations alone. Instead, the biases were said to reflect and illuminate existing systemic biases or in other words they 'shed new light on an old problem'. Consequently, it was suggested that systemic transformation rather than technical remedies is required to redress algorithmic bias. Others have argued that additional conduits of bias (e.g. the potentially prejudicial choices and preferences infused in technology design) are systemic problems that are also in need of fundamental scrutiny (Eaglin, 2017).

What explains the different frames identified?

Sociotechnical imaginaries and access to digital capital

Optimistic frames manifested imaginaries of data-driven technologies as the panacea for crime control. They did

so through their allusions to the propitious availability, utility and capabilities of datafication, endorsing the merits of the technologies. Problems such as biased predictions were either ignored or subjected to tests designed to disprove them. Crime control imperatives were prioritised over the due process and human rights concerns mooted by sections of the critical scholarship (e.g. Eaglin, 2017; Hamilton, 2015; Starr, 2014; Steinbock, 2005). Neutral frames were vacillatory and just as likely to acknowledge sociotechnical imaginaries as they were to refute them. Unlike these two frames, oppositional frames refuted them entirely. Thus, the different frames evoke potentially competing imaginaries. Whilst the optimistic frames and aspects of the neutral frames emphasise the promise of technical accuracy for efficient criminal justice, oppositional frames focus more on racial and social justice imaginaries. These competing visions have for example been reflected in recent fairness debates concerning deployments of criminal justice algorithms. To cite one example, the risk of recidivism algorithms applied in several western jurisdictions including the UK and the US is considered technically fair if they predict outcomes equally well for all social groups (predictive parity) even if the error rates are higher for some groups than others (Dieterich et al., 2016; Skeem and Lowenkamp, 2020). Hence, predictive parity for efficient criminal justice is prioritised, reflecting the emphasis on technical efficiency associated with optimistic and some neutral frames. But critics of the algorithms adopt oppositional frames that focus on visions or imaginaries of systemic and structural equality, arguing that predictive parity is insufficient and that fairness demands an equal balance of error rates for all racial groups (equalised odds) (see Angwin et al., 2016).

Apart from their differing sociotechnical imaginaries, differences in intellectual (subject knowledge) versus digital capital investments also seem to explain the differences between the authors of the three frames. As noted earlier, digital capital includes the ability to access and capitalise resources linked to digital technology (Van Dijk, 2005). Positive endorsements of such techs stemmed largely from data scientists as well as software developers, mathematicians and statisticians. Their dominance over the design of such techs is indicative of their access to digital capital (in this case, relevant design skills) and means that even if no apparent connections exist between them, together they constitute a heterogenous 'technoepistemic network' of experts (Ballo, 2015: 10) able to inject that their ideals and priorities into technology development, framings and imaginaries. This conjures up an image of a world 'envisioned by actors who have the capacity to materialise their visions' (Ballo, 2015: 11). It is argued that this network is male dominated and lacks representation from socially marginal groups such as racial minorities and deprived communities. This could in part explain the uneven distribution of the risk and harms of criminal justice technologies which as the authors of oppositional frames observe, seem to particularly affect socially marginal groups.

Beyond their access to digital capital, it may be that the proponents of optimistic and neutral frames are also motivated by the promise of professional advancement or commercial and other gains (Caulfield and Condit, 2012). Or it may simply be that their position is inspired by visions of the utopic possibilities offered by new technologies, which emphasises transcendence and a sort of disembodied vision of the relationship between humans and machines (Brydolf-Horwitz, 2018), and this may be more prevalent in individuals with certain disciplinary backgrounds. This could in turn explain the heightened tendency to create imaginaries of 'smart' criminal justice as superior to human competencies and capable of providing efficient and cost-effective crime control. Whatever the reason, the foregoing reveals the usefulness of sociotechnical imaginaries as concept that can be used to reflect on the meanings and motivations underpinning technology development. In any case, we should not forget that policy-makers, law enforcement officers and other practitioners that should use or rely on data-driven technologies do not always have a strong digital capital (Sanchez et al., 2019).

What are the implications?

As already noted, positive frames and imaginaries of datafication as vital for the efficient maintenance of the administration of criminal justice were dominant. Several implications arise from this and can be conceptualised as the perpetuation of data harms, the prioritisation of systemic or technocratic crime control and the endorsement of expansionary datafication. The first pertains to the social implications whilst the latter two relate to political implications.

Perpetuating data harms. Of the three frames identified, only the oppositional frame fully desisted from proposing AI models as the panacea for crime control. An important point here is that, in some cases, the authors of frames that seemed neutral or solutionist uncritically accepted potentially biased administrative data. Similar to those propagating optimistic frames, they relied mainly on police data such as arrest data and other administrative datasets that can be infused with bias. They used such data to develop what they depicted as improved algorithmic models that were devoid of noted limitations (e.g. privacy violations, racial bias). But their reliance on questionable data calls into question their potential to address these problems, particularly those related to data harms. As already noted, several studies point to the intractability of such harms (Angwin and Larson, 2016; Hao and Stray, 2019; Lum and Isaac, 2016; Starr, 2014). Therefore, the frames may evoke positive sociotechnical imaginaries that nevertheless obscure the well-documented social costs of data harms, particularly the reproduction of historical

inequalities such as systemic biases and the negative labelling of affected groups (Eaglin, 2017; Ugwudike, 2020).

Prioritising crime control. Apart from the social costs, our analysis also revealed the politics of technology development for criminal justice. In particular, some of those proffering optimistic and neutral frames belong to the aforementioned 'technoepistemic network' of experts, who are empowered by their digital capital to create technologies that inform knowledge of how public resources (e.g. criminal justice expenditures) should be deployed. Their technologies can also shape knowledge production pertaining to criminality, riskiness and efficient crime control. This was clear in our analysis which for example revealed how the developers used their digital capital to produce technologies which they depicted as vital for efficient targeting of 'risky' individuals and crime hotspots. They also evinced the apparent focus of state authorities on applications of technology for crime control without parallel consideration of how to mitigate technological harms to which the critical scholarship alludes. Instead, they developed crime control models and offered technocratic justifications for this, particularly the potential to reduce the costs of criminal justice whilst improving its efficiency. Thus, some models were described as useful for efficiently identifying crime hot spots and patterns. This was intended to support optimal foot and motorised patrol allocation and dosage and real-time patrol planning that is adaptable to dynamic hot spots. They could also support expedited response times and efficient officer allocation systems for responding to individual cases. Staff rostering models and performance appraisal systems were described as additional uses. Predictive AI models were described as vital for pre-empting not only crime but also protests. Invasive surveillance models such as those that can improve covert access to data generated by private devices were proposed. For example, there were models for accessing the data stored in private personal computers and surveillance cameras, to enhance the surveillance systems and other investigative devices used by security services including the police.

Models for networked surveillance and control were also developed. For example, there was evidence of data linkage practices whereby data from different agencies were pooled together for automation by predictive algorithms to demonstrate how such data can be used for early intervention activities by authorities. We have already seen that data linkage to inform official decision-making has been described as capable of leading to data breaches and decontextualisation of data that can prompt unwarranted state intervention (Dencik et al., 2018). Alongside this ethical problem, models enabling state authorities such as the police to bypass privacy restrictions were proposed. These models were said to be capable of storing metadata from the datasets that police services were not permitted to keep in storage. An example is evidence relating to child sexual abuse. With the proposed models, the police could store the metadata for future investigations. In proposing these models, the optimistic and neutral frames prioritised crime control imperatives and managerial concerns over concerns about data harms (regardless of the fact that they could be justified under the relevant legal framework, if we consider this issue from a legalistic perspective).

Conclusion

This article presented and discussed the findings of a study that systematically reviewed and compared multidisciplinary academic abstracts on the data-driven tools shaping (or with the clear potential to shape) decision-making across several justice systems. In this context, we proposed three main frames – optimistic, neutral and oppositional – for understanding how relevant technologies are portrayed, debated how notions of sociotechnical imaginaries and access to digital capital are of the upmost importance to explain differences in the frames and conceptualised the main social and political implications of the current datafication revolution in criminal justice.

The prevalence of optimistic frames we have observed is consistent with current criminal justice deployments of data-driven technologies, which continue to proliferate despite well-documented concerns about data harms. What this indicates is that the frames and their accompanying technologies are central to depictions of the technologies as the panacea for crime control regardless of their accompanying negative impacts. Through their digital capital (but in a context where many practitioners who should use or rely on algorithmic tools often do not have a comparable digital capital), the authors of optimistic and neutral frames have been able to develop and test advanced AI models to illustrate their frames and support law enforcement efforts. In the process, however, they have been more likely to depict their products and similar tools only highlighting the positives and not the many negatives that are still present with the current state of technological advancement and available databases. Indeed, by presiding over the discourse about datafication, the domain frames overshadow the caveats provided by the oppositional frames (notably about privacy violations, racial bias, algorithmic opacity, fairness concerns, inadequate legislation, the commercialisation of datafied criminal justice and systemic bias). By paying insufficient attention to these problems and developing data-driven AI models for criminal justice application, the dominant frames are encouraging the confident expansionary rather than the cautionary datafication of criminal justice.

We hope that our contribution will help furthering critical debates on the complex yet fundamental balance needed in improving the efficiency, effectiveness and legitimacy of criminal justice systems on the one hand, whilst avoiding dangerous and unjust data harms on the other. In

a debate still too dominated by a small number of scientific disciplines that have the digital capital to frame the discussion in a certain direction, but that are likely to lack some of the critical expertise traditionally residing in the social sciences and other disciplines, increased cross-disciplinary efforts are a possible solution to diversify ownership of and access to digital capital in order to facilitate the inclusion of a broader range of perspectives and ways of knowing and to stem the dominance of the imaginaries currently overlooking the potentially harmful sociopolitical implications of untamed datafication.

Declaration of conflicting interests

Funding

ORCID iDs

Anita Lavorgna ⓘD https://orcid.org/0000-0001-8484-1613
Pamela Ugwudike ⓘD https://orcid.org/0000-0002-1084-7796

References

AFP (2021) Futures centre – Disruption. Available at: https://www.afp.gov.au/futures/disruption (accessed 6 September 2021).

AJL (2021) Coded bias: A documentary. Available at: https://www.ajl.org/spotlight-documentary-coded-bias (accessed 6 September 2021).

Angwin J and Larson J (2016) Bias in criminal risk scores is mathematically inevitable, researchers say. Available at: https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say (accessed 6 September 2021).

Angwin J, Larson J, Mattu S, et al. (2016) Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. Available at. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 6 September 2021).

Baack S (2015) Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation and journalism. *Big Data & Society* 7(1): 1–11.

Ballo IF (2015) Imagining energy futures: Sociotechnical imaginaries of the future smart grid in Norway. *Energy Research & Social Science* 9: 9–20.

Barocas S and Selbst AD (2016) Big data's disparate impact. *California Law Review* 104: 671–732.

Bennett Moses L and Chan J (2018) Algorithmic prediction in policing: Assumptions, evaluations, and accountability. *Policing and Society* 28(7): 806–822.

Birhane A (2021) Algorithmic injustice: A relational ethics approach. *Patterns* 2(2): 1–9.

Blair PR, Marcus DK and Boccaccini MT (2008) Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Science and Practice* 15(4): 346–360.

Borah P (2011) Conceptual issues in framing theory: A systematic examination of a decade's literature. *Journal of Communication* 61(2): 246–263.

Bourdieu P (1986) The forms of capital. In: Richardson JG (eds) *Handbook of Theory and Research for the Sociology of Education*. Westport: Greenwood Press, 241–258.

Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.

Brantingham PJ, Valasik M and Mohler GO (2018) Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and Public Policy* 5(1): 1–6.

Braun V, Clarke V, Hayfield N, et al. (2019) Thematic analysis. In: Liamputtong P (ed) *Handbook of Research Methods in Health Social Sciences*. Singapore: Springer. 843–860.

Broussard M (2018) *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge: MIT Press.

Brydolf-Horwitz R (2018) Embodied and entangled: Slow violence and harm via digital technologies. *Environment and Planning C: Politics and Space* (online first).

Burrell J (2016) How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society* 10: 1–12.

Burton JW, Stein M and Jensen T (2019) A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioural Decision Making* 33(2): 220–239.

Caulfield T and Condit C (2012) Science and the sources of hype. *Public Health Genomics* 15: 209–217.

Citron DK (2007) Technological due process. *Washington University Law Review* 85: 1249–1314.

Citron DK and Pasquale F (2014) The scored society: Due process for automated predictions. *Washington Law Review* 89: 1–34.

Cossins D (2018) Discriminating algorithms: 5 times AI showed prejudice. Available at: https://www.newscientistcom/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/ (accessed 6 September 2021).

Custers BHM (2012) Predicting data that people refuse to disclose. *Privacy Observatory Magazine* 3: 1–4.

Custers BHM (2013) Data dilemmas in the information society. In: Custers BHM, Calders T and Schermer B, et al. (eds) *Discrimination and Privacy in the Information Society*. Heidelberg: Springer. 3–26.

Dearden L (2017) How technology is allowing police to predict where and when crime will happen. Available at: https://www.independent.co.uk/news/uk/home-news/police-big-data-technology-predict-crime-hotspot-mapping-rusi-report-research-minority-report-a7963706.html (accessed 6 September 2021).

Dencik L, Hintz A, Redden J, et al. (2018) Data scores as governance: Investigating uses of citizen scoring in public services. Available at: https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf. (accessed 6 September 2021).

Dieterich W, Mendoza T and Brennan (2016) COMPAS risk scales: Demonstrating accuracy equity and predictive parity. Available at: https://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf (accessed 6 September 2021).

Eaglin JM (2017) Constructing recidivism risk. *Emory Law Journal*: 59–122 (online first).

Ensign D, Friedler SA, Neville S, et al. (2018) Runaway feedback loops in predictive policing. In: *Paper presented at the Machine Learning Research. Conference on Fairness, Accountability, and Transparency* 81:1–12.

Entman RM (1993) Framing: Towards clarification of a fractured paradigm. *Journal of Communication* 43: 51–58.

Fazel S, Singh JP, Doll H, et al. (2012) Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24,827 people: Systematic review and meta-analysis. *British Medical Journal* 245: 1–12.

Ferguson AG (2017) *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: New York University Press.

Fontaine C (2016) The myth of accountability: How data (mis)use is reinforcing the problems of public education. *Data and Society*. Available at: https://datasociety.net/output/the-myth-of-accountability-how-data-misuse-is-reinforcing-the-problems-of-public-education/ (accessed 6 September 2021).

Fussey P and Murray D (2019) *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*. Available at: https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf (accessed 6 September 2021).

González-Bailón S (2013) Social science in the era of big data. *Policy & Internet* 5(2): 147–160.

Guo JW and Iribarren SJ (2014) Reporting quality for abstracts of randomized controlled trials in cancer nursing research. *Cancer Nursing* 37(6): 436–444.

Hamilton M (2015) Risk-needs assessment: Constitutional and ethical challenges. *American Criminal Law Review* 231: 236–239.

Hannah-Moffat K (2018) Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*: 1–18 (online first).

Hao K and Stray J (2019) Can you make AI fairer than a judge? Play our courtroom algorithm game. *MIT Technology Review*. Available at: https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm (accessed 6 September 2021).

Harcourt BE (2015) *Risk as a proxy for race. Federal Sentencing Reporter* 27: 237–243.

Harvard Law Review (2017) State v. Loomis: Wisconsin Supreme Court requires warning before use of algorithmic risk assessments in sentencing. Available at: https://harvardlawreview.org/2017/03/state-v-loomis (accessed 6 September 2021).

Hollywood JS, Woods D, Lauland A, et al. (2018) *Emerging Technology Trends and Their Impact on Criminal Justice*. Santa Monica: RAND Corporation.

Home Office (2002) *Offender Assessment System: OASys User Manual, V2*. London: Home Office.

House of Commons Science and Technology Committee (2018) Algorithms in decision-making: Fourth report of a session 2017-9. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf (accessed 6 September 2021).

Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 1–7.

Jasanoff S (2015) Future imperfect: Science, technology, and the imaginations of modernity. In: Jasanoff J and Kim SH (eds) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press: 1–33.

Jasanoff S, Kim S and Sperling S (2007) *Sociotechnical Imaginaries and Science and Technology Policy: A Cross-National Comparison*. Available at: https://stsprogram.org/admin/files/imaginaries/NSF-imaginaries-proposal.pdf (accessed 6 September 2021).

Jewitt C, Price S, Mackley KL, et al. (2020) *Interdisciplinary Insights for Digital Touch Communication*. London: Springer.

Law Society (2019) Algorithm use in the criminal justice system report. Available at: https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report (accessed 6 September 2021).

Lum K and Isaac W (2016) To predict and serve? *Significance* 13: 14–19.

Martín-Martín A, Orduna-Malea E, Thelwall M, et al. (2018) Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics* 12(4): 1160–1177.

Mastrobuoni G (2020) Crime is terribly revealing: Information technology and police productivity. *Review of Economic Studies* 87(6): 2727–2753.

Middleton SE, Lavorgna A, Neumann G, et al. (2020) Information extraction from the long tail: A socio-technical AI approach for criminology investigations into the online illegal plant trade. In: WebSci'20 Workshop, Southampton.

Ministry of Justice (2015) A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf (accessed 6 September 2021).

Ministry of Justice (2019) Risk assessment of offenders: A summary of evidence relating to offender risk assessment, risk of reoffending and risk of serious harm. Available at: https://www.gov.uk/guidance/risk-assessment-of-offenders (accessed 6 September 2021).

Mohler G, Short MB, Malinowski S, et al. (2015) Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association* 110: 1399–1411.

Natale S and Ballatore A (2020) Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence* 26(1): 3–18.

National Offences Management Service (2016) Public protection manual. Available at: https://www.justice.gov.uk/downloads/offenders/psipso/psi-2016/psi-18-2016-pi-17-2016-public-protection-manual.pdf (accessed 6 September 2021).

Olver ME, Stockdale KC and Wormith JS (2014) Thirty years of research on the Level of Service Scales. *Psychological Assessment* 26(1): 156–176.

Oswald M, Grace J, Urwin S, et al. (2018) Algorithmic risk assessment policing models: Lessons from the Durham HART Model and "experimental proportionality". *Information and Communications Technology Law* 27: 223–250.

Pasquale F (2016) *The Black box Society*. Cambridge: Harvard University Press.

Politico (2020) French politicians urge deployment of surveillance technology after series of attacks. Available at: https://

www.politico.eu/article/french-politicians-urge-deployment-of-surveillance-technology-after-series-of-attacks/ (accessed 6 September 2021).

Saint S, Christakis DA, Saha S, et al. (2000) Journal reading habits of internists. *Journal of General Internal Medicine* 15: 881–884.

Sanchez L, Grajeda C, Baggili I, et al. (2019) A practitioner survey exploring the value of forensic tools, AI, filtering, & safer presentation for investigating child sexual abuse material (CSAM). *Digital Investigation* 29: 124–142.

Sanders CB, Weston C and Schott N (2015) Police innovations, "secret squirrels" and accountability: Empirically studying intelligence-led policing in Canada. *British journal of criminology* 55(4): 711–729.

Selbst AD, Boyd D, Friedler SA, et al. (2018) Fairness and abstraction in sociotechnical systems. In: FAT*'19 Conference on Fairness, Accountability, and Transparency, Atlanta, USA.

Skeem J and Lowenkamp C (2020) Using algorithms to address trade-offs inherent in predicting recidivism. *Behav. Sci Law* 38(3): 259–278.

Starr S (2014) Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review* 66(4): 803–872.

Steinbock DJ (2005) Data matching, data mining, and Due process. *Georgia Law Review* 1: 64–81.

Taylor BJ (2017) *Decision Making: Assessment and Risk in Social Work*, 3rd edn University of Ulster (Ireland).

Ugwudike P (2020) Digital prediction technologies in the justice system: The implications of a "race-neutral" agenda. *Theoretical Criminology* (online first).

United Nations (2014) World that counts: Mobilising the data revolution for sustainable development. Available at: https://www.undatarevolution.org/wpcontent/uploads/2014/12/A-World-That-Counts2.pdf (accessed 6 September 2021).

Van Dijk J (2005) *The Deepening Divide: Inequality in the Information Society*. Thousand Oaks: Sage.

Van Dijk J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.

Willoughby A and Nellis M (2016) "You cannot really hide": Experiences of probation officers and young offenders with GPS tracking in Winnipeg, Canada. *Journal Of Technology In Human Services* 34(1): 63–81.

Zarsky TZ (2013) Transparent predictions. *University of Illinois Law Review* 1503: 1553–1568.

Završnik A (2019) Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology* (online first).