



# We have to talk about emotional AI and crime

Lena Podoletz<sup>1</sup>

Received: 30 April 2021 / Accepted: 10 March 2022 / Published online: 5 May 2022  
© The Author(s) 2022

## Abstract

Emotional AI is an emerging technology used to make probabilistic predictions about the emotional states of people using data sources, such as facial (micro)-movements, body language, vocal tone or the choice of words. The performance of such systems is heavily debated and so are the underlying scientific methods that serve as the basis for many such technologies. In this article I will engage with this new technology, and with the debates and literature that surround it. Working at the intersection of criminology, policing, surveillance and the study of emotional AI this paper explores and offers a framework of understanding the various issues that these technologies present particularly to liberal democracies. I argue that these technologies should not be deployed within public spaces because there is only a very weak evidence-base as to their effectiveness in a policing and security context, and even more importantly represent a major intrusion to people's private lives and also represent a worrying extension of policing power because of the possibility that intentions and attitudes may be inferred. Further to this, the danger in the use of such invasive surveillance for the purpose of policing and crime prevention in urban spaces is that it potentially leads to a highly regulated and control-oriented society. I argue that emotion recognition has severe impacts on the right to the city by not only undertaking surveillance of existing situations but also making inferences and probabilistic predictions about future events as well as emotions and intentions.

**Keywords** Urban AI · Emotional AI · Policing · Crime · Crime Prediction · Predictive Policing · Public Security

## 1 Introduction

Forms of machine learning and artificial intelligence have already started to be employed in smart cities as well as in policing, crime prevention and security (Mattern 2021, Powell 2021). AI is defined by Bryson and Winfield (2017:116) as 'artefacts that demonstrate' the capacity to perceive context for action, to act, to associate contexts to actions and/or are cognitive. There is a wide range of computational methods and applications within the field of policing that use machine learning, perhaps the two most well-known examples to date are predictive policing and facial recognition. As Babuta and Oswald (2019:4) note, often such technologies are—poorly—being referred to as AI. However, there is a relatively new branch of AI that is aiming to recognise human emotions and, in some cases, to infer intentions and to predict future behaviour. Such a

technology may, therefore, attract the interest of police and law enforcement authorities. This area is known as affective computing, 'computing that relates to, arises from, or influences emotions' (Picard 1995:1). I limit the scope of this paper to the potential applications of emotional AI in policing publicly available urban space. Therefore, I will not explore other potential applications of affect recognition that concern crime, its prevention or its detection, such as identification of victims of domestic violence or bullying based on analysis of social media content. Concepts such as privacy, security, control and public space may be defined and understood differently in different cultures and societies. Similarly, different jurisdictions have different policing traditions and practices as well as cultural attitudes towards policing. For this reason, I write this paper from the perspective of Western liberal democracies, though the generic issues discussed are in theory applicable globally.

Emotional AI is an umbrella term for any technology that uses affective computing and artificial intelligence to make an assessment or a prediction about a person's emotional state or feelings based on data such as 'words, pictures, intonation, gestures, physiology and facial expressions' (McStay

✉ Lena Podoletz  
lena.podoletz@ed.ac.uk

<sup>1</sup> School of Social and Political Science, University of Edinburgh, Edinburgh, UK

2019a:1). To achieve this, techniques and environments such as sentiment analysis, facial coding of expressions, voice analytics, eye-tracking, measuring skin temperature and other physiological states via wearable tools, analysing gestures and behaviour, virtual reality and augmented reality are used<sup>1</sup>. These technologies have been applied in relation to a number of different fields, such as education (Williamson 2017; McStay 2019a), political campaigns (Bakir 2020), social media (Ortigosa et al. 2014; Wang et al. 2014; Yue et al. 2019; Stark 2018) and distribution of information in digital media (Bakir and McStay 2020) or from the perspective of privacy (McStay 2020).

From a criminological perspective the topic of smart cities and urban AI raises questions and suggests linkages with research on policing and surveillance. Cities are often associated with the problems connected to crime and have similarly been the focus of policing and surveillance activities. In the context of policing, cities are the paradigmatic setting both for street crime and perceived anti-social behaviour but also for public assembly and political protest. In liberal countries the history of cities has also long been the history of a struggle between individual rights and state/city control over public spaces (Harvey 2012). Smart technologies deployed in cities have been subjected to criticism on the grounds of ‘data ethics, privacy, mass surveillance, commodification and social control’ (Foth et al. 2021:319) as well as becoming a form of capital and contributing to increasing control over people’s lives through data gathering and analysis (Sadowski 2020). In recent years electronic surveillance technologies of public space such as CCTV have been implemented in this environment. Here, the task of everyday surveillance is ‘increasingly being performed by non-human algorithms’ (Smith 2020:1). The topic of this paper is an emerging new form of surveillance in public urban spaces, namely what has been termed as ‘emotiveillance’ where people’s emotional states may be subject to surveillance and their intentions may be inferred ‘usually for the purposes of influencing and managing people’ (McStay 2016:151). The idea of monitoring the emotions of people in public urban spaces has been suggested before (Guthier et al. 2014; Zeile et al. 2015, Cabalcanti Roza and Postolache 2016, Adikari and Alakahoon 2021) but the particular risks and implications of potential policing applications have not previously been examined in detail. In this article I will engage with this new technology, and with the debates and literature that surround it. Working at the intersection of criminology, policing, surveillance and the study of emotional AI this paper explores and offers a framework of understanding the various issues that these technologies present particularly to liberal democracies. I argue that these

technologies should not be deployed within public spaces because there is only a very weak evidence-base as to their effectiveness in a policing and security context, and even more importantly represent a major intrusion to people’s private lives and also represent a worrying extension of policing power because of the possibility that intentions and attitudes may be inferred. Moreover, I argue that even if these technologies will be proven accurate in recognising emotions and intentions, they still raise the question as to whether they will ever be desirable in liberal democracies that place an emphasis on the privacy and freedom of personal thoughts, feelings and emotions. Further to this, the danger in the use of such invasive surveillance for the purpose of policing and crime prevention in urban spaces is that it potentially leads to a highly regulated and control-oriented society. Smith (2020) notes that forms of urban surveillance, such as automated facial recognition have severe impacts on the right to the city. I argue that emotion recognition takes this notion a step further by not only undertaking surveillance of existing situations but also making inferences and probabilistic predictions about future events as well as emotions and intentions.

I aim to contribute towards understanding of the implications, challenges and limitations of using emotional AI technologies in the field of policing. In order to achieve this, first I provide an overview of the current state of existing applications of emotion recognition technologies, the longstanding tradition of using biometrics in policing and security, how AI is used in policing today and of how the use of cameras has become more prominent in the policing of urban spaces. Then I identify three areas of emotion recognition in which policing and security applications have been either deployed, trialled or prototyped. Finally, I discuss key issues and implications of the use of such technologies related to accuracy, algorithmic bias, accountability, privacy and other human rights. In conclusion, I argue that by understanding the capabilities and limitations of such technologies as well as the societal costs they potentially introduce, as they currently stand, they should not be deployed in policing urban spaces.

## 2 Biometrics, AI and cameras in policing

We have already witnessed attempts to pioneer the use of certain technologies that belong to the field of emotional AI. Albeit not on a large scale, such technologies are already in use in recruitment to assess candidates (Bogen et al. 2018:32,36), in advertising to understand consumer preferences, in education to examine student behaviour, in healthcare to track mental states, and in social media to profile individuals and to examine group behaviour (McStay 2018). In 2018 Gartner, a leading research and advisory company in the field of technology predicted ‘that by 2022

<sup>1</sup> Emotional AI Lab: <https://emotionalai.org/so-what-is-emotional-ai>.

10% of personal devices will have emotional AI capabilities' compared to 'less than 1% in 2018'<sup>2</sup>. We have yet to see whether this prognosis becomes reality or not, especially in light of the emergence of critical voices regarding emotional AI technologies and their legitimacy, accuracy, and effect on privacy and freedom of thought as well as on individual autonomy and human dignity (see for example Chen 2019, Feldman Barrett et al. 2019, Dupré 2020, Article19 2021, Stark and Hoey 2021, Valcke et al. 2021).

In order to understand the environment in which emotion recognition tools may be applied I will briefly visit three topics. First, I situate this new technology within a longer history of using biometrics for investigation, identification and inferring criminal behaviour and contrast it with previous practices. Secondly, I show how the use of cameras have become a widely applied method in policing and thus, may provide a platform within which to integrate emotion recognition technologies. Thirdly, I provide a brief summary of some use cases of AI that has been deployed or trialled in a policing context.

## 2.1 Use of biometric data in policing and security

As was noted above, emotional AI uses data derived from people's voices, body movements, intonation, gestures and facial expressions. Biometrics have been used in policing and have appeared in criminological theories since the late nineteenth century. In this, two main (and quite different) directions of their potential uses have crystallised. The first is where they are used purely for identification, such as the use of fingerprints, DNA or matching a face to another in order to determine if it is the same face. These uses have become everyday practice in policing and in collecting and presenting evidence in criminal cases, even though they have been subject to criticism regarding their accuracy and ability to provide proof beyond reasonable doubt<sup>3</sup>. Such uses of biometric data (see for example Datta 2020) have brought privacy concerns with them as well, for instance

<sup>2</sup> <https://www.gartner.com/smarterwithgartner/13-surprising-uses-for-emotion-ai-technology/>.

<sup>3</sup> Methodological inaccuracies in the field of forensic science have been brought to light, for example in connection to the use of hair comparisons evidence in criminal cases. For instance, 268 such cases were reviewed by the FBI and the organisation reported that the probability of a positive match between 'hair evidence and the defendant's hair' was overstated by their 'hair microscopy experts' in 95% of the reviewed cases. (<https://www.innocenceproject.org/overturning-wrongful-convictions-involving-flawed-forensics/>) In some cases similar to this, DNA evidence was able to provide proof of innocence after a wrongful conviction. 'The Innocence Project', for example, assisted to the exoneration of 375 people to date using DNA evidence. (<https://www.innocenceproject.org/dna-exonerations-in-the-united-states/>).

regarding the loss of anonymity, the risk of losing control over sensitive information (Kindt 2013), that it can be used to generate extra information that was not the purpose of the collection or that it can be used to track people (Campisi 2013). The second use of biometrics has emerged amongst criminological theories where scholars have tried to use body measurements to identify 'criminal traits' in people with potentially being interested in predicting criminality in people who possess said 'criminal traits' (see for example Lombroso 2006[1876], Lombroso and Ferrero 2004[1893], Kretschmer 1936[1921], Hooton (1939a, 1939b) and Sheldon 1940). These attempts, however, have been ridden with logical and methodological fallacies, have never presented reliable evidence for the claimed deterministic connection between physiology and criminal behaviour, and thus never have become mainstream theories (see Goring 1913, Merton et al. 1940, Horn 2003, Siegel 2012). Consequently, we need to draw a clear distinction between using biometrics for identification, on the one hand, and as crime explanation and/or for crime prediction, on the other. Given this distinction it appears that emotional AI would be used for the latter as these technologies are used to make probabilistic predictions about emotional states and personality traits. The technology emotional AI promises us is however slightly different from the previous uses of biometric data as its predictions regarding emotional states would not derive from the assessment of a static biological or psychological trait but from a set of dynamic behaviours which would be used to conclude that the person in question was currently in a particular emotional state or had a certain intent. As was briefly described above, in the long history of criminological theories and forensics, the use of biometric data has been present in the past few centuries so it would not be without precedent to again make such attempts with the deployment of emotional AI technologies.

## 2.2 Use of cameras in policing

CCTV cameras and the use of big data in policing is already a reality, and it is easy to see how emotional AI technologies can be included in cameras or the use of open-source databases and social-media platforms for example. Facial recognition technologies are also on the rise, and are commonly deployed at borders, during police checks<sup>4</sup> and in some countries they are reported to have been embedded into CCTV cameras as well<sup>5</sup>. According to a definition provided by tech-giant Amazon, facial recognition technologies

<sup>4</sup> <https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition>.

<sup>5</sup> <https://text.npr.org/788597818>, <https://www.hrw.org/news/2020/10/02/russia-expands-facial-recognition-despite-privacy-concerns>.

refer to a set of computational methods which identify whether and where ‘faces exist in an image or video’ and ‘what attributes those faces have’<sup>6</sup>. This is being used by police forces to identify individuals, for instance to compare an image taken from a CCTV camera and to compare it to a database of known offenders or missing people<sup>7</sup>. However, recent research conducted in the UK came to question its accuracy (Fussey et al. 2021). The application of police body-worn cameras<sup>8</sup> (Ariel et al. 2017, Bowling et al. 2019, Lee et al. 2019, Jones 2021) and the discussions about using UAVs (aerial ‘drones’) for police tasks such as intelligence gathering or looking for missing people<sup>9</sup> when combined with facial recognition would create even more powerful identification systems. This, however, introduces concerns regarding large-scale surveillance and human rights. Automated facial recognition not only enhances the abilities of a CCTV system to ‘gather private information on’ people but also can affect ‘the right to freedom of assembly, freedom of thought belief and religion, the freedom of expression and the freedom of association’ (Surveillance Camera Commissioner 2020:12,20). The embedding of emotional AI technologies is arguably able to amplify these effects and would introduce ‘emotiveillance’ (McStay 2016:151) in public spaces on a mass-scale.

### 2.3 Current use of AI in policing

The two most prominent areas of the use of AI in policing at the moment are predictive policing<sup>10</sup> and the above-mentioned facial recognition. Predictive policing has been defined by Meijer and Wessels (2019:1033) as ‘the collection and analysis of data about previous crimes’. These methods can be applied in order to ‘identify places and times with the highest risk of crime’, individuals who are ‘at risk of being offenders or victims’ or ‘people who most likely committed a past crime’ (RAND 2013:1). The use of machine learning in predictive policing have led to the development of more advanced systems but it also raised several legal and ethical questions, for instance about system bias (Babuta et al. 2019). Live facial recognition have been subject to criticism for racial bias and discrimination

as well as for potential negative impact on privacy, freedom of expression and freedom of assembly (Fussey et al. 2019). Similarly, predictive policing tools have been shown to be biased (Ferguson 2017) and has been shown to (re)produce ‘racialized subjects and spaces’ (Jefferson 2018:12). It was also demonstrated that machine learning in predictive policing can lead to the ‘disproportionate policing of historically over-policed communities’ which may result in discrimination as well as in issues regarding accountability for decisions (Lum et al. 2016). One key reason for these effects was shown to be inaccurate or skewed input data which may have been recorded while the implementation of racially biased or unlawful policing practices (Richardson et al. 2019).

## 3 Emotional AI and crime

For the purposes of this paper, I will separate the current main experimental directions of research and deployment of emotional AI technologies related to crime into the following two distinct types based on the general aim of the tool: (1) Crime prediction or identification; (2) Deception detection. Within the category of crime prediction and identification it is possible to distinguish between two sets of technologies based on what type of data they gather and assess: (a) Behavioural cues; (b) Verbal cues.

### 3.1 Predicting crimes

Predicting crimes, whether it is done by an algorithmic system or a human, and whether it is using a person’s facial expressions or their past criminal activity as predictors is always a complex subject. The reason for this is the fact that crime predictions happen with a particular aim in mind and that goal is crime prevention (Meijer et al. 2019). The suggestion behind these efforts is that if we can successfully predict when, where, by whom and against whom or what object crime is going to be committed, at least some parts of crime prevention become straightforward as it provides the police with targets for intervention (Perry et al. 2013:1, Kaufmann et al. 2019). Sometimes, such as in cases of predictive policing this means to prevent crime from happening under particular circumstances (Perry et al. 2013:17). Other times the prediction is about a particular person who previously committed a crime and a risk assessment determines whether there is a risk of re-offending which then contributes to the decision on what punitive measures to employ in that person’s case (Northpointe 2015). It is important to recognise that these determinations are based entirely on risk and probability, and therefore, the crime predictions made using these methods can, at most, let us know only about risks and are not definitive certainties concerning future events (Strikwerda 2020:423). It is worth noting that many

<sup>6</sup> <https://aws.amazon.com/rekognition/the-facts-on-facial-recognition-with-artificial-intelligence/>.

<sup>7</sup> <https://www.met.police.uk/advice/advice-and-information/fr/facial-recognition/>.

<sup>8</sup> <https://www.met.police.uk/advice/advice-and-information/bwv/body-worn-video-bwv/how-and-when-bwv-cameras-are-used/>.

<sup>9</sup> <https://www.west-midlands.police.uk/frequently-asked-questions/police-drones>, <https://www.lancashire.police.uk/about-us/our-organisation/unmanned-aerial-vehicles-drones-uav/>, <https://www.wiltshire.police.uk/article/795/Unmanned-Aerial-Vehicles-UAVs>.

<sup>10</sup> <https://www.predpol.com>

tools that are being referred to by the umbrella term of being ‘predictive’ perform classification rather than prediction in the traditional sense and the risk assessment step is performed by a human actor (Babuta et al. 2020:11–12). Nevertheless, the process as a whole is predictive in the sense that action is taken a priori based on probabilistic assessments of future events which may never occur. Such phenomena is characterised by Zedner (2007) as pre-crime which runs up against fundamental principles embedded in criminal law and criminal justice processes in which the assumption of innocence is paramount. Policing of pre-crime is consistent with the concept of risk society which is built on the minimisation of risks (Beck 1992:19) where factors that are deemed potentially dangerous are eliminated before crimes can manifest. State reactions become determined by statistical probabilities (Swaaningen 1997:174). As Zedner (2015) further suggests, essential problem is that someone who was merely deemed as carrying a risk of future offending may then be subjected to restrictions or punishment (‘pre-punishment’) as a consequence which are contrary to established principles of justice.

### 3.1.1 Crime prediction/identification based on behavioural cues

When we talk about the use of AI to identify behavioural cues in a crime context, we have to distinguish between technologies that help identifying existing criminal behaviour (for instance acts that have already resulted in violence against people or property) and tools that aim to identify emotions based on facial or behavioural clues that may indicate the possibility of future offending. We do not only have to make this distinction for theoretical reasons but also because these applications potentially call for vastly different methodologies, have different legal consequences, and affect procedural guarantees of law enforcement and the criminal justice system in different ways. For example, if the system output identifies a case of a violent crime such as assault, this result is easier to verify by human actors in law enforcement and/or in the criminal justice system than if the output is a prediction regarding a person’s supposed violent intent. We also need to note that as emotional AI is an umbrella term for technologies that use different types of data as input (McStay 2019a:1), accuracies and concerns may vary vastly depending on the data type.

While it seems logical to assume that there are strong correlations between emotional states and intention (Tistarelli et al. 2012), it is not known at this point whether it is possible to make definitive statements about people’s emotional states based on their behaviour. The idea of being able to predict crime in real-time based on behavioural cues and/or patterns originates in the idea that there are certain signs in the behaviour of offenders that precede the commission

of a particular type of criminal activity, and that these signs can be identified—either by human observation or machine learning—and then can be used to prevent said types of crimes. However, a report on the practices of the Transportation Security Administration (TSA) that operates in the USA indicated that stranger-to-stranger behaviour detection and behaviour detection by human actors does not seem to live up to such expectations (ACLU 2017). Similarly, research on ‘stop and searches’ by human police officers show relatively low percentages when it comes to these actions being followed by arrests or findings of illegal substances or possessions such as firearms (Home Office 2020:2, Ferrandino 2013). These findings suggest either the lack of ability to assess behavioural cues and/or the presence of discrimination and prejudice. What we do not know at this point is whether machine learning and artificial intelligence will ever be able to find a correlation between observable behavioural cues and true intention and whether these correlations will be strong enough to serve as bases for police action and other security measures. Existing studies that examine the possibilities of connecting the identifications of emotions to crime prediction (see for example Li et al. 2021) are scarce. It appears that the area needs a more solid and diverse empirical basis for the classification of emotions (Du et al. 2015) on whether classifying emotions universally is even possible considering the heterogeneity of cultures and personal habits (Jack et al. 2012), and on whether facial expressions can really tell us about a person’s emotions<sup>11</sup>. Moreover, there is a further need for empirical evidence on whether the correlations between the portrayal of a particular emotion and criminal intent is definitive enough for the application of intrusive measures such as stop and searches that belong to the field of law enforcement and similar areas.

Identifying crimes-in-progress for timely intervention, investigation, the prevention of further crimes—as stated above—differs from pure crime prediction as here we are dealing with acts that have already happened and the role of emotional AI here could be to differentiate between, for example, a violent robbery and boxing match based on behavioural cues portrayed by the participants. When it comes to the use of emotional AI technologies, two main directions of these efforts seem to be forming in the present time. One is the observation of individual behaviour and interactions between individuals in groups, the other is the observation of crowd behaviour.

The observation of individuals for the purpose of determining whether they are committing a crime or have the intention to commit a crime has long been a part of everyday police patrol operations and other armed forces that are

<sup>11</sup> <https://www.sciencedaily.com/releases/2020/02/200216184520.htm>.

tasked with ensuring the safety of the public, individuals and property (Dahl 1952). During these patrols the police may successfully ‘stumble’ upon criminals in action but as it was pointed out above, detection of suspicious behaviour seems to lack real accuracy. When it comes to the use of machine learning, attempts have been made to identify suspicious behaviour in a specific context, such as the use of cashpoint/ATM machines (Lee et al. 2018) or to detect specific crimes, such as shoplifting (Martínez-Mascorro et al. 2020). One such example is a proposed real-time drone surveillance system which in order to detect violent individuals in public spaces would use aerial images from a drone to estimate human poses and identify which of these seem to be portraying violent behaviour (Singh 2018). This technology is not so much aimed at predicting a crime as at detecting it. But it may not be far-fetched to expect different types of technologies to emerge in its footsteps, ones that instead of merely noticing violent behaviour attempt to predict it, for example by detecting aggression.

Another example in this area is the emotion recognition technology offered by VibraImage. In a nutshell, the technology promises to measure micromovements of a particular person and based on that ‘detect all human emotions’<sup>12</sup>. The technology is used in certain airports in Russia<sup>13</sup> but is has also been deployed in different contexts in countries such as China, Japan and South Korea (Wright 2021). However, as Wright (2021) discovered, details on the technology, training data, relationship between input and output were never properly described or disclosed in any meaningful way. The opaque nature of this tool is enhanced by the fact that it seems that every study or paper that describes its operations simply assumes the fact that the underlying technology is working (Wright 2021).

Crowd control is an important function of the police (Hoggett and Stott 2010; Stott and Kumar 2020). Understanding and assessing the nature of crowd behaviour is a crucial element of designing a police response (Bosch 2013) which could range from tolerant and consensual policing to controlling and repressive. The form of police strategy adopted may further have implications and consequences for police–public relations, trust in the police and perception of police legitimacy. Even though nowadays the type of event where crowd control receives the most attention is protests, it can be a feature of any larger gatherings such as sports, concerts or celebrational public events. However, the public and media attention is not accidental as ‘protesting crowds have their own dynamics’ where various different groups from organisers through businesses to security services interact with each other (Neufeld Redekop 2010). In the past decade

there have been several studies on crowd behaviour analysis (Sánchez and Dencik 2020) on the topics of creating systems that can recognise anomalies or ‘abnormalities’ (not necessarily of criminal nature) in crowds, groups or pedestrian flows (Mahadevan et al. 2010, Chen et al. 2011, Cong et al. 2012, Marsden et al. 2016, Singh 2018, Ullah et al. 2018) and on detection of unusual behaviour (Tung et al. 2011). These technologies can be described as ‘intelligent surveillance’ where the goal is for the system to alert the human actor only when an event (for example unusual behaviour) is detected in order to minimise the need for human participation in the surveillance process (Tung et al. 2011:230).

What is particularly interesting about these above examples from a criminological and legal perspective is that they seem to attribute a certain intent or emotional attitude to certain behavioural patterns and automate this process seemingly without giving much importance to the context in which they are portrayed, the rights and civil liberties of citizens, and the potential effect of surveillance on people exercising their rights. Not only does such surveillance impact privacy in relation to personal information, the person themselves, their personal behaviour and communications, but as Wright and Raab (2014) note, it may violate the privacy of location, of thoughts and feelings and of association. In relation to the importance of context in which behaviour is conducted, it seems that these recognition technologies have a long way to go in interpreting the actual meaning of behaviour as opposed to merely detecting the observable external cues. Just to name one example, it is unclear whether such tool at this stage would be able to differentiate between a fight that qualifies as a violent crime based on legal regulations and a martial arts training session or people acting out a fight scene.

### 3.1.2 Crime identification/prediction based on verbal cues

Another potential source of data for real-time crime identification or prediction is the observation of verbal cues in spoken or written form with the intention of attributing a certain emotion, intention or opinion to these cues. When talking about verbal cues, here we can mean not only the words used but also tone of voice and other signs that can be measured as non-verbal cues in other contexts. One such technique, sentiment analysis, is already being used in marketing for instance (Rambocas et al. 2018). Sentiment analysis aims to determine whether people feel positively or negatively about a product or service for example. It derives ‘subjective information from texts in natural language, such as opinions and sentiments’ in order to generate data that can be used for the purposes of decision-making (Pozzi et al. 2017:1). So far, we have seen models and theoretical ideas for the use of sentiment analysis in a policing and crime prevention context, such as the combined use of sentiment data from Twitter

<sup>12</sup> <https://www.psymaker.com/vibraimage/>.

<sup>13</sup> <https://www.psymaker.com/vibraimage/>.

and weather data (Chen et al. 2015) or the combination of sentiment analysis with other forms of data scanning and data mining techniques that pull data from social media for crime prevention in crisis and post-crisis situations (Domdouzis et al. 2016). A British example for the use of sentiment analysis in law enforcement is from the early 2010 s, when the technology was deployed to observe trends and associations through social media content<sup>14</sup>. At this point, however, we have no empirical data on the accuracy and real effectiveness of such systems for crime prediction or crime identification.

A fast-emerging area is the analysis of online content which can be anything from news articles to social media posts. For instance, EMBERS is a forecasting system that is set out to predict societal events based on open-source data such as tweets, Facebook pages, blog posts, meteorological data, economic indicators and satellite imagery. Some events it focuses on are disease outbreaks, civil unrest and domestic political crises<sup>15</sup>. In case of protests in Brazil in 2013 and in Venezuela in 2014, the system was able to identify indicators of these in social media content and was able to create predictions matching ‘the timing of the events and their trajectory in terms of size and intensity’ (Doyle et al. 2014). This illustrates how these technologies may seem useful for police forces. However, a separate question remains as to how this intelligence would be used in policing practices in line with human rights and civil liberties. This further draws attention to the difference between the potential accuracy and use of predictive analytics and the problems of operationalising them to serve as bases for police action that respects the principles of policing and citizens’ rights that are embedded in liberal democracies.

### 3.2 Deception detection

Deception detection is a subset of techniques and methods in the field of behaviour analysis. Behaviour analysis interviews—which focus on both verbal and non-verbal cues—are very common in police practice. This technique is ‘believed to be one of the two most commonly taught questioning methods in the US’ (Frank Horvath 2006 in Vrij et al. 2007:501). As for the success-rates of humans in these areas, after analysing the results of 20 expert and non-expert comparisons, Bond Jr. et al. (2006: 229) concluded that while experts seem more sceptical than nonexperts when it comes to assuming a person is telling the truth, on average they only ‘achieve less than 55% lie-truth discrimination accuracy’. In their literature review Strömwall et al.

(2004:246) concluded that practitioners or ‘lie experts’ seem to have a similar set of inaccurate beliefs about the non-verbal cues of deception as nonexperts. Police interrogation manuals and police culture serve as the main origins of said beliefs which then are preserved by processes such as ‘cognitive heuristics and biases’ (Strömwall et al. 2004:247). Vrij et al. (2007:510) theorise the possibility that these accuracy rates in experimental studies are partially due to the fact that participants—both non-expert and expert—simply lack ‘knowledge about cues of deception’. This may be true, but we also have to take into account the fact that people with different cultural background and personal characteristics may not share the aforementioned cues of deception or perceptions of deceptive behaviour.

Polygraphs are used in three main areas: screening job-applicants, screening existing employees and in ‘event-specific’ investigations<sup>16</sup>. Some theorise that the use of deception-detection tools in the field of criminal justice is part of the so-called ‘CSI-effect’ where juries in the US have been observed to be keen on seeing more ‘hard evidence’ from both the prosecution and the defence during criminal trials (Chin and Workewych 2016). Chin’s paper notes that this leads to a lot of ‘unnecessary’ deployment of technologies like DNA analysis in situations where from a legal professional point of view there is enough evidence to establish guilt or the lack of it even without these tests (Chin and Workewych 2016). Theoretically, the results of a lie detection test may be used to determine whether a person involved in the investigation or in the criminal justice procedure is telling the truth or not. However, in 2003 a review suggested that the existing previous research conducted on polygraph technology could not provide reliable evidence to show that ‘polygraph tests could have extremely high accuracy’<sup>17</sup>. Currently, standard polygraphs measure certain physiological traits such as cardiovascular activity, respiratory activity and electrodermal activity (Synnott et al. 2015). As opposed to that, emotional AI promises the possibility to assess truth-telling based on facial expressions (Shen et al. 2021) and vocal tones (Marcolla et al. 2020) instead of, or next to, the traditionally observed functions. Whether that task is possible, we do not have enough information to determine at this point as there is only limited evidence for the existence of a correlation between certain micro-facial expressions and deception (Matsumoto et al. 2018). Therefore, it seems certain that we need more research on whether there are certain signs of deception in facial expressions or vocal

<sup>14</sup> <https://www.itnews.com.au/news/police-tap-social-media-in-wake-of-london-attack-344319>.

<sup>15</sup> <https://dac.cs.vt.edu/research-project/embers/>.

<sup>16</sup> The Polygraph and Lie Detection, National Research Council of the National Academies, Washington D.C.: The National Academic Press, 2003, p1.

<sup>17</sup> The Polygraph and Lie Detection, National Research Council of the National Academies, Washington D.C.: The National Academic Press, 2003, p212.

tones, whether these are universal or dependent on culture (Rubin 2014; Taylor et al. 2017), and whether they can be observed with enough certainty to give a definitive assessment of truth-telling or deception.

The EU-funded project called iBorderCTRL aimed to create a system that could be applied at the borders of the European Union to ensure faster and more thorough checks for third country nationals at the borders<sup>18</sup>. One of the tools being developed was the ‘Automatic Deception Detection System’ which set out to quantify ‘the probability of deceit in interviews by analysing interviewees’ non-verbal micro-gestures’<sup>19</sup>. Even though the project was trialled at actual real-life borders (e.g. in Hungary), these tools were never used to make assessments about real people who were crossing the border since this was only a research project and based on project material it does not seem that the development is going any further at this point. The overall accuracy of classifications of experimental runs performed on 18 participants (75.6% in case of truthful statements and 73.7% in case of deceptive statements) of the technology called ‘The Silent Talker’ (O’Shea 2018), based on the very small sample they tested it with, appears to be higher than the average results achieved by expert or non-expert human actors described above (54–55%). However, when looking at each participant’s individual case (O’Shea 2018) we can see that even though the accuracy was 100% in several of the cases, there were quite a few where it was significantly below that. A critical examination of the proposal argues that in the context of deploying emotional AI systems, such as deception detection we need to examine how the use of these technologies create a type of governance that can severely impact opportunities and rights (Sánchez-Monedero and Dencik 2020).

## 4 Discussion

Just as ‘regular’ predictive policing uses past—and sometimes even current, live—events as input data in order to make risk assessments regarding future crime events (Leese 2021), emotional AI would use certain types of similar data in order to produce output regarding the risk of criminal behaviour. At this point we can theorise that it would use some data for training and some data of which the prediction needs to be made (which then can also become a part of the training data). To imagine a very simplified example, let us say that the aim of the tool is to determine whether a public gathering should be expected to become violent. In this case the AI will learn what are the typical behaviours that precede

the ‘violent turn’ and will then look for those in a given public event. This might sound very far-fetched and even dystopian but based on current technological developments we have reviewed above and tendencies in crime control in some jurisdictions around the world, it seems a perfectly logical next step to attempt in the efforts to stop crime. For this reason, the potential issues around the introduction of emotional AI in law enforcement and criminal justice need to be raised.

The use of emotional AI technologies in law enforcement and public security presents us with two separate sets of questions. The first one is about its abilities. What can and cannot emotional AI do to prevent crime? The second line of questions concern the issues and challenges that a debate on the deployment of such technologies would bring. What are these issues when it comes to police use? How can we evaluate whether a given technology is mature enough to use and whether it is necessary, proportionate and is worth the costs? In the final section of the present paper I discuss some key themes that emerge from the previous analysis in relation to the questions above, namely the following: (1) Accuracy and performance; (2) Bias; (3) Accountability; (4) Privacy and other rights and freedoms. Each of these dimensions is worth exploring further, not only in the context of policing and criminal justice applications of emotional AI but also regarding emotional AI’s uses in other fields.

### 4.1 Accuracy and performance

Regarding (potential) accuracy and performance, there is a need to distinguish between different uses of emotional AI. In case of environments such as classifications in imbalanced datasets accuracy may not be the best performance-metric (Veale et al. 2018) but for now it appears that the relevant research literature focuses primarily on accuracy in this field. If we take the use of emotional AI in a private home (Mano et al. 2016; Mano 2018) or in a smart health environment (Fernandez-Caballero et al. 2016) and compare it to the use in law enforcement, for instance, it becomes apparent that the amount of data available regarding an individual person may be vastly different. In a private home some users actively consented to the use of emotional AI, wish to get the benefits that can come from the use and are willingly working with the technology on fine-tuning the accuracy when it comes to recognising their personal emotions for instance by providing feedback on accuracy. However, even in private homes there are data subjects who may not be aware of the emotion recognition technology that is being used or not have given their consent to their profiling, such as visitors, children or individuals in domestic situations where they are being spied on by a partner, for instance. In case of a law enforcement there most likely will be only small amount of useable data on the individual person in question so the

<sup>18</sup> <https://www.iborderctrl.eu/The-project>.

<sup>19</sup> <https://www.iborderctrl.eu/Technical-Framework>.



algorithmic decision will be based on whatever training data the system was fed and whatever data it picked up during the course of self-learning. However, that data will be from a pool of individuals that are potentially very different from the person about whom the algorithmic decision is made. For instance, the universalisation<sup>20</sup> of the classification of the correlation between facial expressions and emotional states is subject to criticism and so is the hypothesis that there is a universal understanding of emotions such as anger or fear as these may be highly dependent on cultural context (Russell 1994). The expression and communication of even the basic emotions may vary between different cultures, situations and ‘people in a single situation’ (Feldman Barrett et al. 2019). Studies have shown that despite existing similarities, different cultures can express certain emotions differently (Elfenbein et al. 2007). Even if basic emotions may be recognised in a cross-cultural setting, there are emotions that may not even be recognised across cultural boundaries (Sauter et al. 2010). It is possible that technology and machine learning will be able to provide an answer to this but a recent study comparing eight commercially available automatic classifier tools for facial affect recognition found that their accuracy varies between 48 and 62% (Dupré 2020). Moreover, when such tools classify emotions, they mainly focus on facial expressions which may not provide sufficient data for identifying emotions as a smile does not always mean happiness (Chen 2019). In fact, the same facial expressions and sets of facial movements may be used to express multiple emotion categories, moreover, may even be used to communicate something else than an emotional state (Feldman Barrett et al. 2019). Humans themselves use other information such as body language, the surrounding environment and personal beliefs and expectation when analysing another person’s emotional state which is mostly perceived in a wider context (Calbi et al. 2017). Furthermore, there is the issue of deception as the individuals in the scope of the criminal justice system or law enforcement may not want to actively disclose their emotions.

An important question in relation to accuracy when it comes to algorithmic decisions is whether they need to be subjected to greater scrutiny than human decisions. In other words, for example, if we know that humans cannot successfully determine deception in stranger-to-stranger interactions (as it was stated above, the success-rate of expert actors is around 55% (Bond Jr. et al. 2006:229)), should we be satisfied with similar accuracy rates in the case of computational methods of deception detection? One argument against this is the scale of potential damage, as algorithmic tools and technologies developed most likely will cover cases on a

significantly larger scale than one single human decision-maker, and possibly will be re-deployed in other fields as well once the technology is deemed reliable enough. With algorithmic decision-making systems in general, there is a clear need to distinguish between different tools, technologies and uses as they all vary in accuracy. In this sense, it could be more logical to set a strict but context-dependent minimum standard for accuracy when it comes to the application of emotional AI technologies in the field of policing and criminal justice. The other conclusion we can draw from the diverse range of accuracy rates in AI technologies is that the scope of examination of potential benefits, costs, damages and impacts should go far beyond the study of accuracy alone. We could phrase it the following way: If it is not accurate, it definitely should not be used, but it should not be used only because it is accurate.

## 4.2 Bias

A matter that closely relates to accuracy is the existence of biases in algorithmic decision-making. It has been long known that human decision-making in law enforcement (Spencer et al. 2016) and also in the criminal justice system (Neitz 2013; Mahoney 2015) is ridden with biases. By now there is considerable amount of research indicating that algorithmic decision-making systems in general can be biased, too (see for example Hannák et al. 2017; Bolukbasi et al. 2016; Eslami et al. 2017; Lorenz et al. 2017; Diaz et al. 2018; Dixon et al. 2018; Ekstrand et al. 2018; Shen et al. 2018, Matsangidou et al. 2019, just to name a few). In the field of criminal justice, for example the biased nature of the COMPAS system was shown (Angwin et al. 2016). The aforementioned studies are not specific to the topic of law enforcement and emotional AI technologies as not many of these have been deployed in real life and the ones which have been have not been researched for biases and discrimination. However, there is research that shows the bias in different emotional AI technologies in other contexts, such as racial bias (Rhue 2018) or differences in the ability to recognise emotions in male and female data subjects (Domnich et al. 2021). The sources of such algorithmic bias can be different, such as (training or input) data bias, human bias (which is brought on by inappropriate system use) and algorithmic processing bias (Shulner Tal 2019). While there is a vast amount of efforts to de-bias systems and mitigate bias (see for example Bolukbasi et al. 2016, Veale et al. 2017, Bellamy et al. 2018, Dixon et al. 2018, Zhang et al. 2018, Zhao et al. 2018, Amini et al. 2019, Shulner Tal et al. 2019, Savani et al. 2020), in the field of law enforcement and in criminal justice strong guarantees and safeguards seem to be necessary showing that a system is not biased against a particular group or against persons with a certain characteristic. Biases in algorithmic decision-making can also reinforce existing

<sup>20</sup> The universalisation of emotions is a subject to methodological critique for being Western-centric (McStay et al. 2019).

societal biases and contribute to the reproduction of existing imbalances in the area of application (Barocas et al. 2016, Noble 2018; Packer et al. 2018), for example by recreating past decisions (Rovatsos et al. 2019:62).

### 4.3 Accountability

The issue of accountability is always a complex one when discussing algorithmic decision-making systems. It can be defined as the ‘capacity to assign responsibility to the correct agency’<sup>21</sup> and also be seen as a principle regarding the ‘obligation to justify one’s actions and the risk of sanctions if justifications are inadequate’ (Castelluccia and Le Métayer 2019:III). If there is a human in the loop—such as the acting police officer—then it seems straightforward to suggest that the said human should be responsible for the decision. However, most people who apply algorithmic tools when performing their work duties and also in their everyday lives do not know how the outputs were generated and sometimes do not know of all or even any of the inputs that were used (Smith 2018). If we consider this fact and assume this is true in the case of the acting police officer in our example, it immediately becomes apparent that we are dealing with a difficult question to answer. Moreover, high-stakes algorithmic decisions and decision-support, like usually the ones involving law enforcement, involve various challenges, such as data changes, human actors augmenting outputs, explaining performance (including performance-metrics that differ from accuracy) as well as being characterised by crossed lines of accountability (Veale et al. 2018). It was noted that in a policing context the use of predictive systems can influence the ability of both the individual acting officers and agencies to give an account of decisions that were taken (Bennett Moses et al. 2018). In general, the literature offers several different methods and techniques for increasing accountability (and the related properties of fairness and transparency), such as auditing, explainability management, discrimination discovery and fairness management (Giunchiglia et al. 2019). In the concrete case of predictive policing strict software evaluations have been suggested as full transparency and comprehensibility may not be possible to achieve in this context (Bennett Moses et al. 2018). In any case, it seems that any requirement regarding transparency needs to go further than simply demanding open-source codes and it also needs to consider the apparent trade-off between accuracy and interpretability (Blacklaws 2018). ALGO-CARE offers a decision-making framework for algorithmic assessment tools in law enforcement which suggests a set of questions that should be answered before

deployment (Babuta et al. 2018). As regards to accountability, they recommend clarifying whether there are any restrictions that may limit accountability or proper evaluation, whether the algorithm is transparent and accountable and whether it will be placed under review (Babuta et al. 2018). It seems necessary, however, to clarify the requirements, duties based on due diligence and build clear accountability structures into every procedure that involves algorithmic decision-making and decision-support.

### 4.4 Privacy and other rights and freedoms

The prevalence of rights in law enforcement is a notoriously complex area as this is a field where exercising the full power of the state could result in severe interference with rights and freedoms in the form of restrictions imposed on people during law enforcement procedures (for instance, arrests or searches of property and body) (Ashworth 2012). Among the rights and freedoms impacted can be the freedom of movement, the right to own property, the right to work and freedom of choice in employment, and the direct example, the right to life. On the other hand, in most countries procedures of law enforcement are bound by strict rules and procedural rights of their own which are meant to ensure longstanding legal principles, such as the right to a lawyer<sup>22</sup> or legal aid,<sup>23</sup> the right to be presumed innocent and to be present at one’s own trial,<sup>24</sup> just to name a few. Law enforcement is also a special area of public service considering the nature of duties, tasks and responsibilities in safety and security (Chalom et al. 2001). These characteristics make the fields of law enforcement, criminal justice and other related areas such as border control and counter-terrorism very different from other potential areas of application. Privacy and human rights advocates have drawn attention to the impacts these new, invasive technologies can have on citizens (Wright et al. 2014, Fan 2021). Algorithmic decision-making in this context can affect the right to a fair trial and due process, the freedom of association, the right to effective remedy, the prohibition of discrimination, and the right to privacy, and can have an impact on effective data protection (Council of Europe 2017). Emotional AI technologies can have a significant effect on privacy (such as concerns about ‘self-determination, consent, choice and abuse of personal control’) and the use of such analytic systems raise ethical questions regarding emotional and mental privacy of both

<sup>21</sup> <https://joanna-bryson.blogspot.com/2019/12/responsibility-accountability.html>.

<sup>22</sup> Directive 2013/48/EU of the European Parliament and of the Council.

<sup>23</sup> Directive (EU) 2016/1919 of the European Parliament and of the Council.

<sup>24</sup> Directive (EU) 2016/343 of the European Parliament and of the Council.

individuals and groups (McStay 2019b:6). Moreover, values such as individual autonomy and human dignity can be challenged by the use of automated emotion recognition technologies as well (Valcke et al. 2021). For instance, the way such systems, their data collection methods are designed and what conceptualisations of emotions are used may have a significant effect on whether such systems ‘should be developed and deployed at all’ (Stark and Hoey 2021:790). Aside of rights and freedoms governed by law there are also several ethical principles as well, that characterise the field and are thought to be crucial in connection with, for instance, affective computing, such as the avoidance of deception, the respect of autonomy or the ensuring that the competence of the system in question is understood (Cowie 2015). While assessing the privacy and wider social impact of any new technology is essential, such as by undertaking a ‘Privacy Impact Assessment’ (PIA) or ‘Surveillance Impact Assessment’ (SIA) (Clarke 2009; Wright 2012, Wright and Raab 2012), these need to be undertaken within a wider framework that considers these within the context of other factors including the practical effectiveness of the technology.

## 5 Conclusion

In this article I have introduced and explained the emerging new technologies that have been described as ‘emotional AI’. This paper has considered how they work and some of the issues relating to their effectiveness in practice as well as the implications of their deployment in public spaces in a policing context. As the drive to make cities ever smarter and to embed machine learning technologies within the urban setting continues to gather momentum, it seems likely that policing and law enforcement agencies will soon begin to explore the potential of such systems to address the problem of urban crime. However, these technologies of ‘emotiveillance’ as McStay (2016:151) puts it, raise particularly troubling questions in relation to privacy, rights and freedoms, especially when used in public spaces.

The paper drew attention to how machine learning, cameras used for the surveillance of public spaces and the analysis of biometrics are already established parts of policing practices. This is significant because these phenomena create an environment into which emotional AI systems can be easily embedded from a technological and policing practices perspective. While in some respects emotional AI represents a continuity with police use of technologies for crime prevention and forensics, in other respects it represents the emergence of a new logic of detecting crime, anticipating criminal behaviour and intervening in crime situations as well as a new form of control over urban spaces and behaviour. Following this I introduced some of the existing use systems and tools that have been deployed, trialled

or prototyped in policing contexts, such as crime prevention or crime detection. The analysis showed the probabilistic nature of these tools and how this was in line with Zedner’s (2007) concept of pre-crime in an urban setting. The discussion covered key issues of these technologies regarding accuracy, bias, discrimination, accountability, privacy and human rights. In relation to the urban setting and cities it was noted that amongst the forms of privacy potentially impacted by emotional AI could include the privacy of thoughts and feelings as well as privacy of association.

In conclusion, I argue that even if emotional AI technologies were to become accurate in revealing thoughts, feelings and intentions, their use in a public urban setting for policing purposes should be resisted in democracies because of the technologies’ clash with human rights values and liberties in such societies.

**Acknowledgements** This research has been supported by the UK Research and Innovation’s (UKRI) and the Japan Science and Technology Agency’s (JST) UKRI-JST Joint Call on Artificial Intelligence and Society (Grant number ES/T00696X/1).

**Author contributions** Not applicable.

**Funding** This research has been funded by the UKRI-JST Joint Call on Artificial Intelligence and Society (Grant number ES/T00696X/1).

**Data availability** Not applicable.

**Code availability** Not applicable.

## Declarations

**Conflict of interest** I declare that I have no conflict of interest.

**Ethical approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- ACLU (2017) Bad Trip: Debunking the TSA's 'Behavior Detection' Program. <https://www.aclu.org/report/bad-trip-debunking-tsas-behavior-detection-program> Accessed 21 Apr 2021
- Adikari A, Alahakoon D (2021) Understanding Citizens' emotional pulse in a smart city using Artificial Intelligence. *IEEE Trans Ind Inform*. <https://doi.org/10.1109/TII.2020.3009277>
- Amini A, Soleimany AP, Schwarting W, Bhatia SN, Rus D (2019) Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure. AIES'19: Proc 2019 AAAI/ACM Conf on AI, Ethics Soc Assoc for Comput Mach, N Y. <https://doi.org/10.1145/3306618.3314243>
- Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine bias. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed 21 Apr 2021
- Ariel B, Sutherland A, Henstock D, Young J, Drover P, Sykes J, Megicks S, Henderson R (2017) Contagious accountability: a global multisite randomized controlled trial on the effect of police body-worn cameras on citizens' complaints against the police. *Crim Justice Behav* 44(2):293–316. <https://doi.org/10.1177/0093854816668218>
- Ashworth A (2012) Negotiating the fundamental right to personal liberty: four problem cases. F W guest memorial lecture: 14 March 2012. *Otago Law Review*. <http://www.nzlii.org/nz/journals/OtaLawRw/2013/1.html>. Accessed 21 Apr 2021
- Babuta A, Oswald M (2018) Machine Learning Algorithms and Police Decision-Making. RUSI. [https://www.excellenceinpolicing.org.uk/wp-content/uploads/2018/09/1-4\\_NewTech\\_Law\\_Privacy\\_Ethics.pdf](https://www.excellenceinpolicing.org.uk/wp-content/uploads/2018/09/1-4_NewTech_Law_Privacy_Ethics.pdf). Accessed 21 Apr 2021
- Babuta A, Oswald M (2019) Data analytics and algorithmic bias in policing. RUSI briefing paper. Royal united services institute for defence and security studies. [https://static.rusi.org/20190916\\_data\\_analytics\\_and\\_algorithmic\\_bias\\_in\\_policing\\_web\\_0.pdf](https://static.rusi.org/20190916_data_analytics_and_algorithmic_bias_in_policing_web_0.pdf). Accessed 8 Dec 2021
- Babuta A, Oswald M (2020) Data analytics and algorithms in policing in England and Wales: towards a new policy framework (RUSI Occasional Paper). Royal United Services institute for defence and security studies. [https://rusi.org/sites/default/files/rusi\\_pub\\_165\\_2020\\_01\\_algorithmic\\_policing\\_babuta\\_final\\_web\\_copy.pdf](https://rusi.org/sites/default/files/rusi_pub_165_2020_01_algorithmic_policing_babuta_final_web_copy.pdf). Accessed 21 Apr 2021
- Bakir V (2020) Psychological operations in digital political campaigns: assessing cambridge analytica's psychographic profiling and targeting. *Front Political Commun* 2020:67. <https://doi.org/10.3389/fcomm.2020.00067>
- Bakir V, McStay A (2020) Empathic Media, emotional ai and the optimization of disinformation. In: Boler M, Davis E (eds) *Affective Politics of Digital Media: Propaganda by Other Means*. Routledge, New York, pp 263–279
- Barocas S, Selbst AD (2016) Big Data's Disparate impact. *Calif Law Rev* 104(3):671–732. <https://doi.org/10.15779/Z38BG31>
- Beck U (1992) *Risk society: towards a new modernity*. Sage, London
- Bellamy RKE, Dey K, Hind M, Hoffmann SC, Houde S, Kannan K, Lohia P, Martino J, Mehta S, Mojsilovic A, Nagar S, Ramamurthy KN, Richards J, Saha D, Sattigeri P, Singh M, Varshney KR, Zhang Y (2018) AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. Preprint at <https://arxiv.org/abs/1810.01943v1>
- Bennett Moses L, Chan J (2018) Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Pol Soc* 28(7):806–822. <https://doi.org/10.1080/10439463.2016.1253695>
- Blacklaws C (2018) Algorithms: transparency and accountability. *Philos Trans R Soc*. <https://doi.org/10.1098/rsta.2017.0351>
- Bogen M, Rieke A (2018) Help wanted: an examination of hiring algorithms, equity, and bias. <https://www.upturn.org/reports/2018/hiring-algorithms/>. Accessed 21 Apr 2021
- Bolukbasi T, Chang KW, Zou JY, Saligrama V, Kalai AT (2016) Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in neural information processing systems*, 4349–4357. Preprint at <https://arxiv.org/abs/1607.06520v1>
- Bond CF Jr, DePaulo BM (2006) Accuracy of deception judgements. *Pers Soc Psychol Rev* 10(3):214–234. [https://doi.org/10.1207/s15327957pspr1003\\_2](https://doi.org/10.1207/s15327957pspr1003_2)
- Bosch C (2013) Crowd theory and the management of crowds: a controversial relationship. *Curr Sociol* 61(5–6):584–601. <https://doi-org.ezproxy.is.ed.ac.uk/10.1177%2F0011392113486443>. Accessed 21 Apr 2021
- Bowling B, Shruti I (2019) Automated policing: the case of body-worn video. *Int J Law Context* 15(2):140–161. <https://doi.org/10.1017/S1744552319000089>
- Bryson JJ, Winfield AFT (2017) Standardising ethical design for artificial intelligence and autonomous systems. *Computer* 50(5):116–119. <https://doi.org/10.1109/MC.2017.154>
- Calbi M, Heimann K, Barratt D, Siri F, Umiltá MA, Gallese V (2017) How context influences our perception of emotional faces: a behavioural study on the kuleshov effect. *Front Psychol* 8:1864. <https://doi.org/10.3389/fpsyg.2017.01684>
- Campisi P (2013) Security and privacy in biometrics: towards a holistic approach. In: Campisi P (ed) *Security and Privacy in Biometrics*. Springer, London
- Castelluccia C, Le Métayer D (2019) Understanding algorithmic decision-making: opportunities and challenges. Study: panel for the future of science and technology. European Union, Brussels. [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS\\_STU\(2019\)624261\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf). Accessed 21 Apr 2021
- CavalcantiRoza VC, Postolache OA (2016) Citizen emotion analysis in smart city. 2016 7th Int Conf Inf, Intell, Syst Appl(IISA). <https://doi.org/10.1109/IISA.2016.7785335>
- Chalom M, Léonard L, Vanderschueren F, Vézina C (2001) Urban safety and good governance: the role of the police. International Centre for the Prevention of Crime, Montreal. <https://www.un.org/ruleoflaw/files/UNH%20Role%20of%20Police%20Publication.pdf>. Accessed 21 Apr 2021
- Chen DY, Huang PC (2011) Motion-based unusual event detection in human crowds. *J Vis Commun Image Represent* 22(2011):178–186
- Chen X, Cho Y, Jang S (2015) Crime prediction using Twitter sentiment and weather. 2015. *Syst Inf Eng Des Symp, SIEDS*. <https://doi.org/10.1109/SIEDS.2015.7117012>
- Chen A (2019) Computers can't tell if you're happy when you smile. *MIT Technology Review*. <https://www.technologyreview.com/2019/07/26/238782/emotion-recognition-technology-artificial-intelligence-inaccurate-psychology/>. Accessed 21 Apr 2021
- Chin JM, Workewych L (2016) The CSI Effect. *Oxford Handbooks Online, Scholarly Research Reviews*. <https://doi.org/10.1093/oxfordhb/9780199935352.013.28>
- Clarke R (2009) Privacy impact assessment: its origins and development. *Comput Law Secur Rev* 25(2):123–135. <https://doi.org/10.1016/j.clsr.2009.02.002>
- Cong Y, Yuan J, Liu J (2012) Abnormal event detection in crowded scenes using sparse representation. *Pattern Recogn* 46(7):1851–1864. <https://doi.org/10.1016/j.patcog.2012.11.021>
- Council of Europe (2017) Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications. <https://rm.coe.int/>

- [algorithms-and-human-rights-en-rev/16807956b5](#). Accessed 21 Apr 2021
- Cowie R (2015) Ethical issues in affective computing. In: Calvo R, D’Mello S, Gratch J, Kappas A (eds) *The oxford handbook of affective computing*. <https://doi.org/10.1093/oxfordhb/9780199942237.013.006>. Accessed 15 Apr 2022
- Dahl RA (1952) Importance of observation in law enforcement. *J Crim Law Criminol* 43(1):103–113. <https://scholarlycommons.law.northwestern.edu/jclc/vol43/iss1/12>. Accessed 15 Apr 2022
- Datta A (2020) Self(ie)-governance: technologies of intimate surveillance in india under COVID-19. *Dialogues Hum Geogr* 10(2):234–237. <https://doi.org/10.1177/2043820620929797>
- Diaz M, Johnson I, Lazar A, Piper AM, Gergle D (2018) Addressing age-related bias in sentiment analysis. *Proc of the 2018 ACM CHI Conf Hum Factors Computing Syst*. <https://doi.org/10.1145/3173574.3173986>
- Dixon L, Li J, Sorensen J, Thain N, Vasserman L (2018) Measuring and mitigating unintended bias in text classification. *Proc of the 2018 AAAI/ACM Conf on AI, Ethics, and Soc (AIES’18)*. <https://doi.org/10.1145/3278721.3278729>
- Domdouzis K, Akhgar B, Andrews S, Gibson H (2016) A social-media and crowd-sourcing data mining system for crime prevention during and post-crisis situations. *J Syst Inf Technol* 18(4):364–382. <https://doi.org/10.1108/JSIT-06-2016-0039>
- Domnich A, Anbarjafari G (2021) Responsible AI: gender bias assessment in emotion recognition. *Computer vision and pattern recognition*. Preprint at <https://arxiv.org/abs/2103.11436v1>
- Doyle A, Katz G, Summers K, Ackermann C, Zavorin I, Lim Z, Muthiah S, Butler P, Self N, Zhao N, Lu CT, Khandpur RP, Fayed Y, Ramakrishnan N (2014) Forecasting significant societal events using the embers streaming predictive analytics system. *Big Data* 2(4):185–195. <https://doi.org/10.1089/big.2014.0046>
- Du S, Martinez AM (2015) Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues Clin Neurosci* 17(4):443–455. <https://doi.org/10.31887/DCNS.2015.17.4/sdu>
- Dupré D, Krumhuber EG, Küster D, McKeown GJ (2020) A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLoS ONE* 15:4. <https://doi.org/10.1371/journal.pone.0231968>
- Ekstrand MD, Tian M, Azpiazu IM, Ekstrand JD, Anuyah O, McNeill D, Pera MS (2018) All the cool kids, how do they fit in?: popularity and demographic biases in recommender evaluation and effectiveness. *Proc 1st Conf Fairness, Account Transpar(PMLR)* 81:172–186. <https://doi.org/10.1822/B2GM6F>
- Elfenbein HA, Beaupré M, Lévesque M, Hess U (2007) Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion* 7(1):131–146. <https://doi.org/10.1037/1528-3542.7.1.131>
- Eslami M, Vaccaro K, Karahalios K, Hamilton K (2017) ‘Be careful; Things can be worse than they appear’ – understanding biased algorithms and users’ behaviour around them in rating platforms. In: *Proceedings of the 11th international conference on web and social media (ICWSM 2017)*, pp 62–71
- Fan MD (2021) Privacy, public disclosure, and police body-worn camera footage. In: Newell BC (ed) *Police on camera*. Routledge, London, pp 182–199
- Feldman Barrett L, Adolphs R, Marsella S, Martinez AM, Pollak SD (2019) Emotional expressions reconsidered challenges to inferring emotion from human facial movements. *Psychol Sci Public Interest* 20(1):1–68. <https://doi.org/10.1177/2F1529100619832930>
- Ferguson AG (2017) *The rise of big data policing: surveillance*. NYU Press, Race and the Future of Law Enforcement
- Fernandez-Caballero A, Martínez-Rodrigo A, Pastor JM, Castillo JC, Lozano-Monator E, López MT, Zangróniz R, Latorre JM, Fernández-Sotos A (2016) Smart environment architecture for emotion detection and regulation. *J Biomed Inform* 64:55–73. <https://doi.org/10.1016/j.jbi.2016.09.015>
- Ferrandino J (2013) The efficiency of frisks in the NYPD, 2004–2010. *Crim Justice Rev*. <https://doi.org/10.1177/2F0734016812466570>
- Foth M, Anastasiu I, Mann M, Mitchell P (2021) From automation to autonomy: technological sovereignty for better data care in smart cities. In: Wang BT, Wang CM (eds) *Automating Cities*. Springer, pp 319–343
- Fussey P, Murray D (2019) Independent report on the london metropolitan police service’s trial of live facial recognition technology. University of Essex, Human Rights Centre
- Fussey P, Davies B, Innes M (2021) ‘Assested’ facial recognition and the reinvention of suspicion and discretion in digital policing. *Br J Criminol* 61(2):325–344. <https://doi.org/10.1093/bjc/azaa068>
- Giunchiglia F, Otterbacher J, Batsuren K, Bogina V, Hartman A, Kleanthous S, Kuflik T, Orphanou K, Shulner Tal A (2019) Literature review and bibliographic referencing system. [http://www.cycat.io/wp-content/uploads/2020/06/D3.1\\_Literature\\_Review\\_M12.pdf](http://www.cycat.io/wp-content/uploads/2020/06/D3.1_Literature_Review_M12.pdf). Accessed 15 Apr 2022
- Goring C (1913) *The English convict: a statistical study*. His Majesty’s Stationery Office, London
- Guthier B, Alharthi R, Abaalkhail R, El Saddik A (2014) Detection and visualization of emotions in an affect-aware city. In: *EMASC’14: proceedings of the 1st international workshop on emerging multimedia applications and services for smart cities*. 2014 (November) 23–28
- Hannák A, Wagner C, Garcia D, Mislove A, Strohmaier M, Wilson C (2017) Bias in online freelance marketplaces: evidence from Taskrabbit and Fiverr. *Proc of the ACM Conf on Comput Supported Cooperative Work and Soc Comput*. <https://doi.org/10.1145/2998181.2998327>
- Harvey D (2012) *Rebel cities: from the right to the city to the urban revolution*. Verso, New York
- Hoggett J, Stott C (2010) The role of crowd theory in determining the use of force in public order policing. *Polic Soc* 20(2):223–236. <https://doi.org/10.1080/10439461003668468>
- Hooton EA (1939a) *Crime and the Man*. Harvard University Press
- Hooton EA (1939b) *The American Criminal: an anthropological study*. Harvard University Press
- Horn DG (2003) *The Criminal Body: Lombroso and the Anatomy of Deviance*. Routledge, London
- Jack RE, Garrod OGB, Yu H, Caldara R, Schyns PG (2012) Facial expressions of emotions are not culturally universal. *Proc Natl Acad Sci USA* 109(19):7241–7244. <https://doi.org/10.1073/pnas.1200155109>
- Jefferson BJ (2018) Predictable policing: predictive crime mapping and geographies of policing and race. *Ann Am Assoc Geogr* 108(1):1–16
- Jones R (2021) *Theorising police body-worn cameras*. In: Newell BC (ed) *Police on camera*. Routledge, London, pp 38–47
- Kaufmann M, Egbert S, Leese M (2019) Predictive policing and the politics of patterns. *Br J Criminol* 59(3):674–692. <https://doi.org/10.1093/bjc/azy060>
- Kindt EJ (2013) *Privacy and data protection issues of biometric applications: a comparative legal analysis*. Springer, Netherlands, Dordrecht
- Kretschmer E (1936[1921]) *Physique and Character: an investigation of the nature of constitution and of the theory of temperament*. K. Paul, Trench, Trubner and co., London
- Lee WK, Leong CF, Lai WK, Leow LK, Yap TH (2018) ArchCam: real time expert system for suspicious behaviour detection in ATM site. *Expert Syst Appl* 109:12–24. <https://doi.org/10.1016/j.eswa.2018.05.014>

- Lee M, Taylor E, Willis M (2019) Being held to account: detainees' perceptions of body-worn cameras. *Aust N.Z. J Criminol* 52(2):174–192. <https://doi.org/10.1177/2F0004865818781913>
- Leese M (2021) Security as Socio-Technical Practice: Predictive Policing and (Non-)Automation. *Swiss Political Sci Rev* 27(1):150–157. <https://doi.org/10.1111/spsr.12432>
- Li Z, Zhang T, Jing X, Wang Y (2021) Facial expression-based analysis on emotion correlations, hotspots, and potential occurrence of urban crimes. *Alex Eng J* 60(1):1411–1420. <https://doi.org/10.1016/j.aej.2020.10.061>
- Lombroso C, Ferrero G (2004[1893]) *Criminal Woman, the Prostitute, and the Normal Woman*. Duke University Press
- Lombroso C (2006[1876]) *Criminal Man*. Duke University Press
- Lorenz F, Yuan J, Lommatzsch A, Mu M, Race N, Hopfgartner F, Albayrak S (2017) Countering contextual bias in Tv watching behaviour: introducing social trend as external contextual factor in TV recommenders. *Proc of the 2017 ACM Int Conf on Interact Exp for TV and Online Video(TVX'17)*. <https://doi.org/10.1145/3077548.3077552>
- Lum K, Isaac W (2016) To predict and serve? Significance. <https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- Mahadevan J, Li W, Bhalodia V (2010) Vasconcelos N (2010) anomaly detection in crowded scenes. *IEEE Computer Soc Conf Computer Vis Pattern Recognit 2010*:1975–1981. <https://doi.org/10.1109/CVPR.2010.5539872>
- Mahoney K (2015) Judicial bias: the ongoing challenge. *J Dispute Resolut* 2015(1):43–69
- Mano LY, Faical BS, Nakamura LHV, Gomes PH, Libralon GL, Meneguete RI, Filho GPR, Giancristofaro GT, Pessin G, Krishnamachari B, Ueyama J (2016) Exploiting IoT technologies for enhancing health smart homes through patient identification and emotion recognition. *Comput Commun* 89–90:178–190. <https://doi.org/10.1016/j.comcom.2016.03.010>
- Mano LY (2018) Emotional condition in the health smart homes environment: emotion recognition using ensemble of classifiers. 2018 *Innov Intell Syst Appl (INISTA Conference)*. <https://doi.org/10.1109/INISTA.2018.8466318>
- Marcolla FM, de Santiago R, ScarantoDazzi RL (2020) Novel Lie speech classification by using voice stress. *Proc of the 12th Int Conf on Agents and Artificial Intelligence* 2:742–749
- Marsden M, McGuinness K, Little S, O'Connor NE (2016) Holistic features for real-time crowd behaviour anomaly detection. Preprint at <https://arxiv.org/abs/1606.05310v1>
- Martínez-Mascorro GA, Orzís-Bayliss JC, Aberu-Pederzini JR, Terashima-Marín H (2020) Suspicious behavior detection on shoplifting cases for crime prevention by using 3D convolutional neural networks. Preprint at <https://arxiv.org/abs/2005.02142v1>
- Matsangidou M, Otterbacher J (2019) What is Beautiful Continues to Be Good. In: Lamas D, Loizides F, Nacke L, Petrie H, Winckler M, Zaphiris P (eds) *Human-Computer Interaction INTERACT 2019: INTERACT 2019 Lecture Notes in Computer Science*, vol 11749. Springer, pp 243–264
- Matsumoto D, Hwang HC (2018) Microexpressions differentiate truths from lies about future malicious intent. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2018.02545>
- Mattern S (2021) *A city is not a computer: other urban intelligences*. Princeton University Press, Princeton, NJ
- McStay AJ (2016) *Digital advertising*, 2nd edn. Palgrave, London
- McStay A (2019a) Emotional AI and EdTech: serving the public good? *Learn Media Technol* 45(3):270–283. <https://doi.org/10.1080/17439884.2020.1686016>
- McStay A (2018) *Emotional AI: the rise of empathic media*. Sage Publications, New York, NY
- McStay A (2019b) *The Right to Privacy in the Age of Emotional AI*. [https://www.ohchr.org/Documents/Issues/DigitalAge/ReportPrivacyinDigitalAge/AndrewMcStayProfessor%20of%20Dig](https://www.ohchr.org/Documents/Issues/DigitalAge/ReportPrivacyinDigitalAge/AndrewMcStayProfessor%20of%20Digital%20Life,%20BangorUniversityWalesUK.pdf)
- McStay A (2020) Emotional AI, soft biometrics and the surveillance of emotional life: an unusual consensus on privacy. *Big Data Soc*. <https://doi.org/10.1177/2F2053951720904386>
- Meijer A, Wessels M (2019) Predictive policing: review of benefits and drawbacks. *Int J Public Adm* 42(12):1031–1039. <https://doi.org/10.1080/01900692.2019.1575664>
- Merton RK, Ashley-Montagu MF (1940) Crime and the Anthropologist. *Am Anthropol New Series* 42(3):384–408
- Neitz MB (2013) Socioeconomic bias in the judiciary. *Cleveland State Law Rev* 61:137–165
- Neufeld Redekop V, Paré S (2010) Beyond control: a mutual respect approach to protest crowd-police relations. Bloomsbury Academic, New York, NY
- Noble SU (2018) *Algorithms of oppression: how search engines reinforce racism*. NYU Press, New York, NY
- Northpointe (2015) *Practitioner's Guide to COMPAS Core*. <https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>. Accessed 21 Apr 2021
- O'Shea J, Crockett K, Khan W, Kidynis P, Antoniadis A, Boultaakis G (2018) Intelligent deception detection through machine based interviewing. *Int Joint Conf Neural Netw (IJCNN)*. <https://doi.org/10.1109/IJCNN.2018.8489392>
- Home Office (2020) *Police powers and procedures England and Wales, year ending 31 March 2020 – Second Edition*. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/935355/police-powers-procedures-mar20-hosb3120.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/935355/police-powers-procedures-mar20-hosb3120.pdf). Accessed 15 Apr 2022
- Ortigosa A, Martín JM, Carro RM (2014) Sentiment analysis in Facebook and its application to e-learning. *Comput Hum Behav* 31:527–541. <https://doi.org/10.1016/j.chb.2013.05.024>
- Packer B, Halpern Y, Guajardo-Céspedes M, Mitchell M (2018) Text embedding models contain bias. Here's why that matters. Google AI: <https://developers.googleblog.com/2018/04/text-embedding-models-contain-bias.html>. Accessed 15 Apr 2022
- Perry WL, McInnis B, Price CC, Smith SC, Hollywood JS (2013) Predictive policing: the role of crime forecasting in law enforcement operations. RAND Corporation, USA. [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR200/RR233/RAND\\_RR233.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf). Accessed 15 Apr 2022
- Picard RW (1995) *Affective computing*. M.I.T. media laboratory perceptual computing section technical report No. 321. <https://affect.media.mit.edu/pdfs/95.picard.pdf>. Accessed 15 Apr 2022
- Powell AB (2021) *Undoing optimisation: civic action in smart cities*. Yale University Press, New Haven, Connecticut
- Pozzi FA, Fersini E, Messina E, Liu B (2017) Challenges of Sentiment Analysis in Social Networks: An Overview. In: Pozzi FA, Fersini E, Messina E, Liu B (eds) *Sentiment Analysis in Social Networks*. Elsevier, Cambridge, pp 1–11
- Rambocas M, Pacheco BG (2018) Online sentiment analysis in marketing research: a review. *J Res Interact Mark* 12(2):146–163. <https://doi.org/10.1108/jrim-05-2017-0030>
- RAND (2013) Predictive Policing: Forecasting Crime for Law Enforcement. [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR200/RR233/RAND\\_RR233.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf). Accessed 15 Apr 2022
- Rhue L (2018) Racial influence on automated perceptions of emotions. Available at SSRN. <https://doi.org/10.2139/ssrn.3281765>. Accessed 15 Apr 2022
- Richardson R, Schulz J, Crawford K (2019) Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems and justice. *N.Y.U. Law Rev* 94:192–233
- Rovatsos M, Mittelstadt B, Koene A (2019) Landscape summary: bias in algorithmic decision-making. Centre for Data Ethics and Innovation. <https://assets.publishing.service.gov.uk/gover>

- ment/uploads/system/uploads/attachment\_data/file/819055/Landscape\_Summary\_-\_Bias\_in\_Algorithmic\_Decision-Making.pdf. Accessed 15 Apr 2022
- Rubin VL (2014) Pragmatic and cultural considerations for deception detection in asian languages. *ACM Trans on Asian Lang Inf Process*. <https://doi.org/10.1145/2605292>
- Russell JA (1994) Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychol Bull* 115(1):102–141. <https://doi.org/10.1037/0033-2909.115.1.102>
- Sadowski J (2020) Too smart: how digital capitalism is extracting data. MIT Press, Controlling Our Lives and Taking Over the World
- Sánchez FL, Hupont I, Tabik S, Herrera F (2020) Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects. *Inf Fusion* 64:318–335. <https://doi.org/10.1016/2Fj.inffus.2020.07.008>
- Sánchez-Monedero J, Dencik L (2020) The politics of deceptive borders: ‘biomarkers of deceit’ and the case of iBorderCtrl. *Inf Commun Soc*. <https://doi.org/10.1080/1369118X.2020.1792530>
- Sauter DA, Eisner F, Ekman P, Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc Natl Acad Sci USA* 107(6):2408–2412. <https://doi.org/10.1073/pnas.0908239106>
- Savani Y, White C, Govindarajulu NS (2020) Intra-processing methods for debiasing neural networks. <https://arxiv.org/abs/2006.08564v2>
- Sheldon WH (1940) The varieties of human physique: an introduction to constitutional psychology. Harper and Brothers Publishers, New York and London
- Shen JH, Fratamico L, Rahwan I, Rush AM (2018) Darling or Babygirl? Investigating Stylistic Bias in Sentiment Analysis. 5<sup>th</sup> Workshop on Fairness, Accountability and Transparency in machine Learning. [https://www.fatml.org/media/documents/darling\\_or\\_babygirl\\_stylistic\\_bias.pdf](https://www.fatml.org/media/documents/darling_or_babygirl_stylistic_bias.pdf)
- Shen X, Fan G, Niu C, Chen Z (2021) Catching a liar through facial expression of fear. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2021.675097>. Accessed 15 Apr 2022
- Siegel D (2012) The Methods of Lombroso and Cultural Criminology. In: Knepper P, Ystehede PJ (eds) *The Cesare Lombroso Handbook*. Routledge, Abingdon
- Singh A, Patil D, Omkar SN (2018) Eye in the sky: real-time drone surveillance system (DDS) for violent individuals identification using scatternet hybrid deep learning network. *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW)*. <https://arxiv.org/abs/1806.00746v1>
- Smith A (2018) Many Facebook users don’t understand how the site’s news feed works. <https://www.pewresearch.org/fact-tank/2018/09/05/many-facebook-users-dont-understand-how-the-sites-news-feed-works/>. Accessed 15 Apr 2022
- Smith GJD (2020) The politics of algorithmic governance in the black box city. *Big Data Soc*. <https://doi.org/10.1177/2F2053951720933989>
- Spencer KB, Charbonneau AK, Glaser J (2016) Implicit bias and policing. *Soc Personal Psychol Compass* 10(1):50–63. <https://gspp.berkeley.edu/assets/uploads/research/pdf/SpencerCharbonneauGlaser.Compass.2016.pdf>. Accessed 15 Apr 2022
- Stark L (2018) Facial recognition, emotion and race in animated social media. *First Monday* 23:9. <https://doi.org/10.5210/fm.v23i9.9406>
- Stark L, Hoey J (2021) The ethics of emotion in artificial intelligence systems. *FACt’21: Proc of the ACM Conf on Fairness, Account and Transpar*. <https://doi.org/10.1145/3442188.3445939>
- Stott C, Kumar TKV (2020) Far from the ‘madding crowd’: an introduction to the special issue of policing crowd event. *Policing: J Policy Pr* 14(4):860–864
- Strikwerda L (2020) Predictive policing: the risks associated with risk assessment. *Police J: Theory, Pr Princ* 94(3):422–436. <https://doi.org/10.1177/2F0032258X20947749>
- Strömwall LA, Granhag PA, Hartwig M (2004) Practitioners’ beliefs about deception. In: Granhag PA, Strömwall LA (eds) *The Detection of Deception in Forensic Contexts*. Cambridge University Press, pp 229–250
- Surveillance Camera Commissioner (2020) Facing the camera. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/940386/6.7024\\_SCC\\_Facial\\_recognition\\_report\\_v3\\_WEB.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940386/6.7024_SCC_Facial_recognition_report_v3_WEB.pdf). Accessed 15 Apr 2022
- Swaaningen R (1997) *Critical Criminology: Visions from Europe*. Sage Publications, London
- Synnott J, Dietzel D, Ioannou M (2015) A review of the polygraph: history, methodology and current status. *Crime Psychol Rev* 1(1):59–83. <https://doi.org/10.1080/23744006.2015.1060080>
- Taylor PJ, Larner S, Conchie SM, Menacere T (2017) Culture moderates changes in linguistic self-presentation and detail provision when deceiving others. *Royal Soc Open Sci* 4:6. <https://doi.org/10.1098/rsos.170128>
- Tistarelli M, Barrett SE, O’Toole AJ (2012) Facial Recognition, Facial Expressions and Intention Detection. In: Mordini E, Tzovaras D (eds) *Second generation biometrics: the ethical, legal and social context*. The international library of ethics, law and technology, vol 11. Springer, Dordrecht
- Tung F, Zelek JS, Clausi DA (2011) Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance. *Image Vis Comput* 29:230–240. <https://doi.org/10.1016/j.imavis.2010.11.003>
- Ullah H, Altamimi AB, Uzair M, Ullah M (2018) Anomalous entities detection and localisation in pedestrian flows. *Neurocomputing* 290:74–86. <https://doi.org/10.1016/j.neucom.2018.02.045>
- Valcke P, Clifford D, Steponėnaitė VK (2021) Constitutional challenges in the emotional AI Era. In: Micklitz HW, Pollicino O, Reichman A, Simoncini A, Sartor G, De Gregorio G (eds) *Constitutional challenges in the algorithmic society*. Cambridge University Press, Cambridge
- Veale M, Binns R (2017) Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. *Big Data Soc*. <https://doi.org/10.1177/2F2053951717743530>
- Veale M, Van Kleek M, Binns R (2018) Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proc of the 2018 CHI Conf on Hum Factors in Comput Syst (CHI’18)*. <https://doi.org/10.1145/3173574.3174014>
- Vrij A, Mann S, Kristen S, Fisher RP (2007) Cues to Deception and ability to detect lies as a function of police interview styles. *Law Behav* 31(5):499–518. <https://doi.org/10.1007/s10979-006-9066-4>
- Wang N, Kosinski M, Stillwell DJ, Rust J (2014) Can well-being be measured using Facebook status updates? Validation of Facebook’s gross national happiness index. *Soc Indic Res* 115(1):483–491. <https://doi.org/10.1007/s11205-012-9996-9>
- Williamson B (2017) Moulding student emotions through computational psychology: affective learning technologies and algorithmic governance. *Educ Media Int* 54(4):267–288. <https://doi.org/10.1080/09523987.2017.1407080>
- Wright D (2012) The state of the art in privacy impact assessment. *Comput Law Secur Rev* 28(1):54–61. <https://doi.org/10.1016/j.clsr.2011.11.007>
- Wright D, Raab C (2012) Constructing a surveillance impact assessment. *Comput Law Secur Rev* 28(6):613–626. <https://doi.org/10.1016/j.clsr.2012.09.003>
- Wright D, Raab C (2014) Privacy principles, risks and harms. *Int Rev Law, Comput Technol* 28(3):277–298. <https://doi.org/10.1080/13600869.2014.913874>
- Wright J (2021) Suspect AI: vibraimage, emotion recognition technology, and algorithmic opacity. *Sci, Technol Soc*. <https://doi.org/10.1177/2F09717218211003411>

- Yue L, Chen W, Li X, Zuo W, Yin M (2019) A survey of sentiment analysis in social media. *Knowl Inf Syst* 60(2):617–663. <https://doi.org/10.1007/s10115-018-1236-4>
- Zedner L (2007) Pre-crime and post-criminology. *Theor Criminol* 11(2):261–281
- Zedner L (2015) Pre-crime and pre-punishment: a health warning. *Crim Justice Matters* 2015(102):23–24
- Zeile P, Resch B, Exner JP, Sagl G (2015) Urban Emotions: Benefits and Risks in using Human Sensory Assessment for the Extraction of Contextual Emotion Information in Urban Planning. In: Geertman S, Ferreira J, Goodspeed R, Stillwell J (eds) *Planning Support Systems and Smart Cities*. Springer, pp 209–225
- Zhang BH, Lemoine B, Mitchell M (2018) Mitigating unwanted biases with adversarial learning. *Proc of the 2018 AAAI/ACM Conf on AI, Ethics, and Soc(AIES'18)*. <https://doi.org/10.1145/3278721.3278779>
- Zhao J, Wang T, Yatskar M, Ordonez V, Chang KW (2018) Gender bias in coreference resolution: evaluation and debiasing methods. *Proc of the 2018 Conf of the North Am Chapter of the Assoc for Comput Linguist: Hum Lang Technol* 2:15–20. <https://doi.org/10.18643/v1/N18-2003>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.