BIG BROTHER WATCH
DEFENDING CIVIL LIBERTIES, PROTECTING PRIVACY

Big Brother Watch briefing on Algorithmic Decision-Making in the Criminal Justice System

January 2020

About Big Brother Watch

Big Brother Watch is a cross–party, non-party, independent non-profit organisation leading the protection of privacy and civil liberties in the UK. We expose and challenge threats to people's privacy, freedoms and civil liberties at a time of enormous technological change in the UK.

Introduction

This briefing focuses on use of **algorithms in both policing and judicial decision-making.** New technology and automated systems, encompassing artificial intelligence, machine learning and big data analytics are being used ever more widely in the criminal justice system. These are being used to predict communities to target with greater police resources; to assess individuals' risk of committing crimes in the future; to identify suspects in public places; and to support police investigations.

However, there are significant problems with these systems, including the data being used to build and train these models; the lack of regard those those building these systems have for human rights and data protection. This has resulted in **biased and discriminatory decisions that are in large part automated**, with **serious implications for the fairness of criminal justice.**

There are also a number of ethical concerns raised when decisions are delegated to algorithms, including:

- inconclusive evidence leading to unjustified actions;
- inscrutable evidence leading to opacity;
- misguided evidence leading to bias;
- unfair outcomes leading to discrimination; and
- transformative effects leading to challenges for autonomy and informational privacy.[1]

The protections afforded to individuals subject to these biased, unfair or unjust decisions are minimal. Many will never even be aware they have been subject to such an automated decision.

[1] Mittelstadt et al, 'The ethics of algorithms: Mapping the debate', 1 December 2016, *Big Data & Society* (https://journals.sagepub.com/doi/full/10.1177/2053951716679679)

We urge Parliamentarians to call on the Government to:

1. **Amend the Data Protection Act to ensure that any decisions involving automated processing that engage rights protected under the Human Rights Act 1998 are ultimately human decisions with meaningful human input.**
2. **Introduce a requirement for mandatory bias testing of any algorithms, automated processes or AI software used by the police and criminal justice system in decision-making processes.**
3. **Prohibit the use of predictive policing systems that have the potential to reinforce discriminatory and unfair policing patterns.**

1. Case studies

This briefing will consider the following case studies of algorithmic decision-making systems used by the police and within the criminal justice system:

- **Geographic crime prediction systems:** PredPol and other bespoke police systems
- **Individual-oriented crime prediction:** Durham Constabulary's Harm Assessment Risk Tool (HART)
- **Individual-oriented crime prediction:** National Data Analytics Solution (NDAS)
- **Individual-oriented crime prediction:** Offender Assessment System (OASys) and the Offender Group Reconviction Scale (OGRS)
- **Identification systems:** Live facial recognition surveillance
- **Crime investigation:** artificial intelligence and digital evidence

2. Geographic crime prediction systems: PredPol

2.1 Geographic crime prediction systems use crime data to create future predictions of where and when certain crimes will occur.

2.2 One such example is the commercial geographic crime prediction toll created by the company PredPol, which styles itself as 'The Predictive Policing Company'.[2] The eponymous PredPol product feeds crime and location information into a machine-learning algorithm to calculate predictions of times and locations ('hotspots') where specific crimes are most likely to occur. The algorithm is based on an 'earthquake' model of crime that predicts certain crimes result in further 'aftershock' crimes within the same area.[3] The system uses current and historical police crime data to create its predictions: crime type, crime location and crime date and time.[4]

2.3 PredPol was used by Kent Police for 5 years between 2013 and 2018 before it was scrapped, with a superintendent saying it had been 'challenging' to show whether crime was actually reduced as a result.[5] Greater Manchester Police, West Midlands Police, West

[2] https://www.predpol.com/
[3] https://www.predpol.com/
[4] https://www.predpol.com/technology/
[5] https://www.bbc.co.uk/news/uk-england-kent-46345717

Yorkshire Police and the Metropolitan Police have also either trialled PredPol or other similar geographic crime prediction systems including their own bespoke systems.[6]

2.4 It has been reported that PredPol has a contractual requirement on customers, including police forces, to engage in promotional activities, such as publicly endorsing PredPol as successfully reducing crime,[7] despite the lack of clear evidence to corroborate this.[8] The widely claimed benefits of geographic crime prediction are not independently supported by empirical evidence.[9]

2.5 There are a number of serious issues inherent in geographic crime prediction. Such tools use past crime data from police records to predict future crime patterns – but police records represent the crimes, locations and groups that are policed, rather than the actual occurrence of crime. Police data represents systematic under-reporting and systematic over-reporting of certain types of crime and in certain locations.[10] Police data may represent discriminatory policing practices and societal inequalities, such as those which result in black men being more than 3 times more likely to be arrested than white men in the UK,[11] and black people over 9 times more likely to be stopped and searched than white people.[12]

2.6 This means that the data upon which such models are built are not accurate reflections of the true occurrence of crime and are likely to be skewed towards certain crimes and locations, which may reflect social inequalities or discriminatory policing patterns. The 'hotspot' predictions that PredPol creates are also highly targeted, meaning that even small differences in input probabilities lead to huge differences in these output predictions.

[6] https://www.ibtimes.co.uk/predictive-policing-predpol-future-crime-509891

[7] https://archives.sfweekly.com/sanfrancisco/all-tomorrows-crimes-the-future-of-policing-looks-a-lot-like-good-branding/

[8] https://www.techdirt.com/articles/20131031/13033125091/predictive-policing-company-uses-bad-stats-contractually-obligated-shills-to-tout-unproven-successes.shtml

[9] Albert Meijer & Martijn Wessels (2019) Predictive Policing: Review of Benefits and Drawbacks, International Journal of Public Administration (https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664)

[10] Lum, Kristian, and William Isaac. 2016. 'To Predict and Serve?' Significance 13 (5): 14–19 (https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x); Bennett Moses, L., & Chan, J. (2016). 'Algorithmic prediction in policing: Assumptions, evaluation, and accountability'. *Policing and Society*. (https://www.tandfonline.com/doi/10.1080/10439463.2016.1253695); Barocas, S. and Selbst, A.D., 2016. Big Data's disparate impact. *California law review*, 104, 671. (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899)

[11] Ministry of Justice, 'Black, Asian and Minority Ethnic disproportionality in the Criminal Justice System in England and Wales', 2016 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/639261/bame-disproportionality-in-the-cjs.pdf)

[12] UK Government Stop & Search facts and figures, February 2019: https://www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/policing/stop-and-search/latest

Resulting predictions are likely to present inaccurate or biased depictions of criminal activity,[13] leading to discriminatory policing interventions.[14]

2.7 As such, these predictive models can expand and entrench the biases represented in the crime data,[15] as a result of self-perpetuating 'feedback loops'. This occurs when previous crime data leads to further location-biased predictions, the dispatch of police resources and further crime recording, which is then fed back into the system. Such data-based predictions risk predicting crime and allocating resources in the same areas, creating self-affirming predictions.[16]

2.8 For example, when policing is disproportionately focused on neighbourhoods with a high black and minority ethnic population, police records will represent higher crime records in those neighbourhoods.[17] There have long been issues with the over-policing of ethnic minorities in the UK, leading to widespread social unrest (for example in St Paul's, Bristol in 1980; Toxteth, Liverpool in 1981; and Broadwater Farm, Tottenham in 1985 and again in 2011.)[18]

2.9 Multiple studies have found that such geographic crime prediction systems, built and trained using historic police crime records, have lead to self-perpetuating feedback loops particularly in areas with low income and black and ethnic minority populations already subject to excessive policing. This risks reinforcing patterns of inequality.[19] One such study on drug crime in Oakland, California, stated that "locations that are flagged for targeted policing are those that were… already over-represented in the historical police data", and concluded that "allowing a predictive policing algorithm to allocate police resources would

[13] Innes, M., Fielding, N., & Cope, N. (2005). 'The appliance of science?': The theory and practice of crime intelligence analysis. *The British Journal of Criminology*, 45, 39–57

[14] Albert Meijer & Martijn Wessels (2019) Predictive Policing: Review of Benefits and Drawbacks, International Journal of Public Administration (https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664)

[15] Lum, Kristian, and William Isaac. 2016. 'To Predict and Serve?' Significance 13 (5): 14–19 (https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x)

[16] Ensign et al, (2017) 'Runaway Feedback Loops in Predictive Policing', Cornell University Library, 29 June 2019 https://arxiv.org/abs/1706.0984); Mohler et al (2011), 'Self-exciting point process modeling of crime', Journal of the American Statistical Association (http://www.stat.ucla.edu/~frederic/papers/crime1.pdf)

[17] Custers, B., 2013. Data dilemmas in the information society: introduction and overview. *In*: B. Custers, T. Calders, B. Schermer and T. Zarsky, eds. *Discrimination and privacy in the information society: data mining and profiling in large databases.*: Springer, 3–26. (https://link.springer.com/chapter/10.1007%2F978-3-642-30487-3_1)

[18] Lewis et al, 'Reading the Riots' (2011), London School of Economics and The Guardian, (http://eprints.lse.ac.uk/46297/1/Reading%20the%20riots(published).pdf); Centre for Crime and Justice Studies, 'Policing the riots: from Bristol and Brixton to Tottenham, via Toxteth, Handsworth, etc', (https://www.crimeandjustice.org.uk/publications/cjm/article/policing-riots-bristol-and-brixton-tottenham-toxteth-handsworth-etc)

[19] Ensign et al, (2017) 'Runaway Feedback Loops in Predictive Policing', Cornell University Library, 29 June 2019 (https://arxiv.org/abs/1706.0984); Lum, Kristian, and William Isaac. 2016. 'To Predict and Serve?' Significance 13 (5): 14–19 (https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x); Lyria Bennett Moses & Janet Chan (2018) Algorithmic prediction in policing: assumptions, evaluation, and accountability, Policing and Society, 28:7, 806-822 (https://www.tandfonline.com/doi/10.1080/10439463.2016.1253695);

result in the disproportionate policing of low-income communities and communities of colour".[20]

2.10 In addition, such predictive algorithmic models typically lack transparency and accountability. Police officers may not be able to fully understand and interpret the outcomes of predictive models, meaning that predictions can dictate decisions rather than meaningfully inform them. This leads to an accountability deficit, where it is not clear if there is any meaningful decision-making input from police who merely act on predictive algorithms without critical analysis.[21] The use of these systems has the potential to create an unchallengeable narrative of criminal communities.

3. Individual-oriented crime prediction: Durham Constabulary's Harm Assessment Risk Tool (HART)

3.1 Durham Constabulary has developed its own machine-learning algorithm, the Harm Assessment Risk Tool (HART), which profiles suspects to predict their risk of re-offending in the future, giving them a risk score: high, moderate or low. This AI-generated risk score is used to advise whether to charge a suspect or release them onto a rehabilitation programme, 'Checkpoint'. If individuals who have been assessed by HART as 'moderate' risks successfully complete the 'Checkpoint' rehabilitation programme, they will not receive a criminal conviction. This system therefore has significant consequences for individuals' criminal justice outcomes. The principle of using historic data about an individual to make predictions about their potential future behaviour also brings into question the presumption of innocence and the right to a fair trial.

3.2 The HART algorithm is based on a random forest model, constructed from 509 separate classification and regression decision trees (CART), which are combined into the forecasting model. HART was built on a dataset using approximately 104,000 custody events over a five year period. It uses 34 different predictor variables to arrive at a forecast, 29 of which focus on the individual's history of criminal behaviour. A further variable is the number of police intelligence reports relating to the individual. The other variables include age, gender and two types of residential postcode.

[20] Lum, Kristian, and William Isaac. 2016. 'To Predict and Serve?' Significance 13 (5): 14–19 (https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x)

[21] Bennett Moses, L., & Chan, J. (2016). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society*. (https://www.tandfonline.com/doi/10.1080/10439463.2016.1253695)

3.3 Big Brother Watch's investigation found that one of the postcode variables fed into the HART system was a commercial marketing data product from the global data broker Experian, known as 'Mosaic'.[22] Mosaic is a socio-geodemographic segmentation tool, consisting of postcode stereotypes created from 850 million pieces of data, including census data, ethnicity, health data, employment, GCSE results, child benefits and income support, family and personal names linked to ethnicity, data scraped from online sources including pregnancy advice websites and much more.[23]

3.4 This data is used to profile all 50 million adults in the UK[24] into stereotypes based on their postcodes, creating household profiles which, in 2018, included categories such as "Asian Heritage", "Disconnected Youth", "Crowded Kaleidoscope", "Families with Needs" or "Low Income Workers".[25] Experian's profiles attribute 'demographic characteristics' to each stereotype. For example, 'Asian Heritage' individuals were characterised as being part of "extended families" living in "inexpensive, close-packed Victorian terraces", and that "when people do have jobs, they are generally in low paid routine occupations in transport or food service".[26] 'Crowded Kaleidoscope' were described as "multi-cultural" families likely to live in "cramped" and "overcrowded flats", with names like 'Abdi' and 'Asha'. 'Families with Needs' were profiled as receiving "a range of benefits" with names like 'Stacey', while 'Low Income Workers' were typified as having "few qualifications" and were "heavy TV viewers" with names like 'Terrence' and 'Denise'.[27]

3.5 Durham Constabulary paid £45,913 to Experian for the licencing of their services, including £25,913 for this information,[28] using 'CustodyMosaicCodeTop28', which is described as "the 28 most common socio-geo-demographic characteristics for County Durham",[29] as a predictor in its HART forecasting model, which influenced criminal justice outcomes.

3.6 It is appalling that such discriminatory profiling data was used by police to predict people's "risk", with the potential of affecting potentially life-changing criminal justice decisions. Allowing this kind of profiling data – which includes not only ethnicity data but a whole

[22]https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

[23] Paul Cresswell et al., 'Under the bonnet: Mosaic data, methodology and build', Experian Marketing Services, 1 April 2014. This has since been removed from the Experian website, but we can provide a copy on request.

[24] Mosaic Infographic, Experian, (http://www.experian.co.uk/marketing-services/knowledge/infographics/infographic-new-mosaic.html) Also see Paul Cresswell et al, 'Under the bonnet: Mosaic data, methodology and build', Experian Marketing Services, 1 April 2014, p.7: (http://www.experian.co.uk/assets/marketing-services/presentations/mosaic-data-methodology-and-build.pdf)

[25] https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

[26] https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

[27] https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

[28] Durham PCC Register of Contracts (https://www.durham-pcc.gov.uk/document-library/finance/register-of-contractspcc.pdf)

[29] Sheena Urwin, 'Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model', unpublished thesis, University of Cambridge, 2016, (http://www.crim.cam.ac.uk/alumni/theses/Sheena%20Urwin%20Thesis%2012-12-2016.pdf)

host of other race and socioeconomic proxy information, including postcodes – to be used in public sector algorithms is discriminatory and, in the criminal justice system, will lead to unjust and inaccurate decisions. This AI risk assessment reinforces existing policing biases and social inequalities, instituting a 'postcode lottery' of justice under the banner of innovation.

3.7 One of the academics instrumental to the development of HART stated to Big Brother Watch verbally that in their opinion the Experian Mosaic data was one of the strongest predictor variables and as such had a valid place in the tool. There is no public data in the available literature to evidence this claim – but even if there were, this statement shows a concerning failure to differentiate between correlation and causation, and treats people for whom such generalised interpretations are not valid as simply collateral. This statistical stereotyping leads to unjust and prejudicial treatment that is the very definition of discrimination.

3.8 If the system does produce a discriminatory, inaccurate prediction, it is likely to negatively impact the individual because the system is designed to over-estimate individuals' risk of re-offending:

> "The HART model intentionally favours... cautious errors, where the offenders' levels of risk are over-estimated".[30]

This means that the system will predict a "sizeable proportion" of people as being higher risk than they actually are, with the result that innocent people may be incorrectly profiled and subjected to a prosecution they might otherwise have avoided. Its unacceptable that this model deliberately overestimates 'risk' - in effect, the likelihood of guilt – in a way that is fundamentally incompatible with the rule of law and the right to a fair trial. It is vital that individuals are presumed innocent until proven guilty in our justice system.

3.9 The HART developers' assessment of the system did indeed recognise that "Some of the predictors used in the model... (such as postcode) could be viewed as indirectly related to measures of community deprivation".[31] They also identified the serious potential for the postcode variables to create 'feedback loops' and reinforce biased criminal justice decisions:

[30]Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model, M. Oswald, J. Grace, S. Urwin (Durham Constabulary) & G.C. Barnes, 31 August 2017, (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029345)

[31]Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model, M. Oswald, J. Grace, S. Urwin (Durham Constabulary) & G.C. Barnes, 31 August 2017, (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029345)

> "one could argue that this variable risks a kind of feedback loop that may perpetuate or amplify existing patterns of offending. If the police respond to forecasts by targeting their efforts on the highest-risk postcode areas, then more people from these areas will come to police attention and be arrested than those living in lower-risk, untargeted neighbourhoods. These arrests then become outcomes that are used to generate later iterations of the same model, leading to an ever-deepening cycle of increased police attention."[32]

3.10 Moreover, Durham Constabulary announced an intention for the HART system to expand beyond the current use alongside Checkpoint, "with the forecasts influencing all of the many other decisions that are made in the wake of bringing a suspected offender into police custody".[33]

3.11 Following Big Brother Watch's investigation of the HART system, and the use of Experian's Mosaic stereotyping data, we publicised our findings and called for the Experian Mosaic data to be removed immediately (6 April 2018). Durham Constabulary removed the Experian Mosaic data less than three weeks later (24 April 2018).

3.12 Separately, since our investigation Experian has also rebranded some of the most crudely titled household profiles in Mosaic, for example changing 'Asian Heritage' to 'Large Family Living' and 'Crowded Kaleidoscope" to 'City Diversity'.[34] However, this is a cosmetic change and there is nothing to suggest that the wide range of intrusive underlying data used to create the profiles, including ethnicity data, has changed[35]. Whilst we welcome the removal of overtly offensive stereotype names, we remain deeply concerned about the existence of this profiling data and the role it plays in various areas of public life.

3.13 In the US, a similar system to HART called COMPAS, which was also designed to assess the risk of reoffending, was found to be evidencing "significant racial disparities". The COMPAS algorithm is trained on police records, and similarly to the information fed into HART via Mosaic, it uses information on an individuals' education, employment, benefits and financial information. COMPAS routinely underestimated the likelihood of white suspects reoffending, even when the suspect's race was not explicitly included in the dataset. The

[32] Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model, M. Oswald, J. Grace, S. Urwin (Durham Constabulary) & G.C. Barnes, 31 August 2017, (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029345)

[33] Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model, M. Oswald, J. Grace, S. Urwin (Durham Constabulary) & G.C. Barnes, 31 August 2017, (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029345)

[34] https://www.experian.co.uk/assets/marketing-services/brochures/mosaic-ps-brochure.pdf

[35] Paul Cresswell et al., 'Under the bonnet: Mosaic data, methodology and build', Experian Marketing Services, 1 April 2014. This has since been removed from the Experian website, but we can provide a copy on request.

opposite was true for black suspects who were generally considered at greater risks of recidivism - the system wrongly labelled them as future criminals at twice the rate of white defendants.[36]

3.14 Durham Constabulary's creation and use of HART exemplifies many of the issues associated with rapid application of algorithms in the justice system: not only profiling, bias and discrimination but also data exploitation, de facto automated decision making, and dubious predictions which have consequences for the presumption of innocence and people's right to a fair trial.

4. Individual-oriented crime prediction: National Data Analytics Solution (NDAS)

4.1 The National Data Analytics Solution (NDAS),[37] created by West Midlands Police in partnership with 8 other police forces, including Greater Manchester Police and the Metropolitan Police,[38] is intended to predict serious violent crime using artificial intelligence. The purpose of such predictions is to promote interventions before crimes have been committed. The NDAS was intended for all police forces to use from March 2019, although its operational implementation has been temporarily delayed.[39]

4.2 West Midlands Police aims for the system to expand to 34 different use cases (e.g. predicting the likelihood of someone to commit violent crime) for all 44 law enforcement agencies (43 forces including the National Crime Agency). The final product will be "a permanent, cloud-hosted analytics platform" running predictive analytics.[40] West Midlands Police has been given £4,465,000 for the National Data Analytics Solution by the UK Home Office's 'Police Transformation Fund' for 2018/19,[41] and another £5,000,000 in 2019/20.[42]

[36] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[37] Formerly known as the National Analytics Solution

[38] Founding partners include Greater Manchester Police, Merseyside Police, Metropolitan Police, Staffordshire Police, Warwickshire Police, West Mercia Police, West Yorkshire Police and an unknown (redacted) other.
See: Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

[39] The West Midlands Police and Crime Commissioner Ethics Committee unanimously voted in favour of being given further information on NAS before it could advise on whether it should go ahead or not.
See: West Midlands Police and Crime Commissioner Ethics Committee, Minutes, 3 April 2019 (https://www.westmidlands-pcc.gov.uk/media/514528/Ethics-Committee-03042019-MINUTES-.pdf)

[40] Founding partners include Greater Manchester Police, Merseyside Police, Metropolitan Police, Staffordshire Police, Warwickshire Police, West Mercia Police, West Yorkshire Police and an unknown (redacted) other.
See: Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

[41] Police Transformation Fund – investments in 2018-19 (https://www.gov.uk/government/publications/police-transformation-fund-investments-in-2018-to-2019)

[42] https://www.gov.uk/government/publications/police-transformation-fund-investments-in-2019-to-2020

4.3 The NDAS intends to legitimise and support pre-emptive policing interventions using big data analytics and machine learning to make predictions about people's potential future actions in order for police to take action before crimes have been committed. West Midland's Police states that the NDAS will "create meaningful insight and identify value driving patterns which should ultimately lead to crime prediction and prevention", enabling police to "make early interventions" and "prevent criminality… by proactively addressing threats".[43] This again raises serious concerns around the presumption of innocence and people's right to a fair trial.

4.4 In addition, there is a significant risk of perpetuating and deepening bias as a result of the data used to train the NDAS and to ultimately make predictions. The NDAS uses data about individuals taken from a number of public and private sources. This includes police records, "data ingested from 9 founding partners' source systems", data from other public bodies including social care services, local authorities, education providers and other emergency services, data from private sector organisations and open source data – including social media data.[44] The private sector data includes the use of Experian's Mosaic,[45] considered above.

4.5 The police records used to train and predict as part of NDAS include CRIMES;[46] Intelligence Management System (IMS);[47] ICIS;[48] Corvus;[49] Prisoner Intelligence Notification System (PINS);[50] Police National Computer (PNC);[51] OASIS;[52] Drug Intervention Programme (DiP);[53] Organised Crime Group (OCG);[54] and Stop and Search records.[55] West Midlands Police combines data from these 9 police systems, using statistical modelling to identify the

[43] Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

[44] Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf); See also: West Midlands Police Police and Crime Commissioner Ethics Stakeholder Engagement Proposal (9 March 2018) (Not publicly available but please request a copy if you would like to see it).

[45] Page 14, Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

[46] Records of crimes committed

[47] Police intelligence reports about events, locations and offenders

[48] Custody information

[49] Intelligence, briefing and tasking system

[50] Prisoner information and notification of release

[51] Information on people, crimes, vehicles and property

[52] Event logging system

[53] Drug intervention programme data

[54] Record and mapping of OCGs in the West Midlands Police area

[55] Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf); https://www.westmidlands-pcc.gov.uk/media/191164/wmpcc_005_2013_technology_task_force_options_paper_appendix.pdf; https://www.westmidlands-pcc.gov.uk/media/473339/SPCB-05-Dec-17-Item-9-WMP-Change-Portfolio.pdf; =

strongest 'predictive' fields that indicate someone's likelihood of involvement in a certain crime.[56]

4.6 There are serious ethical, data protection and rights issues with several of these data sources. First, the use of data from stop and search, a policing tool that has been consistently used in a biased and discriminatory way, to influence future criminal justice outcomes, will clearly result in similarly biased outcomes. In April 2019, it was reported that black people were 5 times more likely than white people to be stopped and searched in the West Midlands Police area, while Asian people were 2.8 times more likely.[57] In 2017/18, nationally, black people were more than 9 times more likely to be stopped and searched than white people (based on Home Office stop and search data).[58] In May 2019, following the increased use of section 60 'suspicionless' stop and search powers, it was reported that black people were 40 times more likely than white people to be stopped and searched across the UK.[59]

4.7 The uncritical general use of crime records within NDAS also embeds biases in policing. As discussed above, police records are not entirely objective and accurate representations of actual criminality and represent societal and structural inequalities as well as recording failures. For example, in 2019, Her Majesty's Inspectorate of Constabulary found that West Midlands Police failed to record more than 16,600 violent crimes each year – 78% of violent crimes and 89% of sexual offences were not recorded when reported.[60] In 2017, HMIC found that West Midlands Police failed to record 38,800 crimes every year – one out of every 6.[61]

4.8 These problems are relevant to all police forces. However, it raises particularly serious questions about whether West Midlands Police – or indeed any police - use of data analytics can be credible or fit for purpose when the data they hold is so inaccurate, let alone the fact that police data cannot be considered an accurate record of crime in the first place. In addition, the integration of several 'intelligence' databases (Corvus, IMS) into the NDAS, containing information with potentially questionable or unproven evidential basis, also raises questions about the impartiality and fairness of the system. West Midlands Police have even admitted these problems themselves:

[56] Data Driven Insight & Data Science Capability for UK Law Enforcement (http://www.excellenceinpolicing.org.uk/wp-content/uploads/2017/10/EIP17_2-5_Utilising_Data_Science.pdf)

[57] https://www.westmidlands-pcc.gov.uk/media/514876/SPCB-160419-Item-9a-Stop-and-Search-and-Use-of-Force.pdf

[58] https://www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/policing/stop-and-search/latest

[59] https://www.theguardian.com/law/2019/may/04/stop-and-search-new-row-racial-bias

[60] https://www.bbc.co.uk/news/uk-england-46867657

[61] https://www.bbc.co.uk/news/uk-england-41178872

> *"There is potential for bias to be present in the underlying dataset in terms of the recorded incidents of harmful / most harmful offences and within the intelligence reports."* [62]

4.9 West Midlands Police has said that it intends future partners providing data for the NDAS will include the National Health Service, Department for Education, Department for Work and Pensions, Department for Communities and Local Government.[63] The prospect of police or law enforcement basing criminal justice decisions on information from the health service, education, social welfare, local authorities or other public services information is extremely concerning. People should not be profiled based on this information. Such excessive data sharing raises serious ethical and data protection issues and could have a chilling effect on people's access to vital public services.

4.10 One of the proposed predictive models evidences further problems inherent in this type of predictive analytics. West Midlands Police developed a predictive risk model, using the police records as above, to identify the 32 strongest 'predictive' fields that indicated someone as an 'influencer' of co-offending.[64] These included the number of times an individual was stopped and searched, the number of intelligence reports about an individual (also analysed above), the number of solo crimes committed by nominal associates, and mentions of the individual in drug habit or addiction records.[65] It is clearly wrong to not only take action against people based on predictions using historic data, but to profile and criminalise people based on the actions of others. Recording criminal assumptions about people based on their addictions also raises ethical issues.

4.11 There does not appear to be any provision to inform individuals that they have been subject to a NDAS prediction resulting in intervention or whether they will have any opportunity to object to their data being processed or challenge the prediction.

4.12 An independent review of the National Data Analytics System by the Alan Turing Institute Data Ethics Group (ATI DEG) and Independent Digital Ethics Panel for Policing (IDEPP), based only on a draft police report on the NDAS, concluded that there were "serious ethical

[62] Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

[63] Police Transformation Fund – National Analytics Solution, Final Business Case v6.0 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

[64] Data Driven Insight & Data Science Capability for UK Law Enforcement (http://www.excellenceinpolicing.org.uk/wp-content/uploads/2017/10/EIP17_2-5_Utilising_Data_Science.pdf)

[65] Data Driven Insight & Data Science Capability for UK Law Enforcement (http://www.excellenceinpolicing.org.uk/wp-content/uploads/2017/10/EIP17_2-5_Utilising_Data_Science.pdf)

issues… concerning surveillance and autonomy, as well as the reversal of the presumption of innocence on the basis of statistical prediction”.[66]

4.13 The reviewers questioned whether it was “ethical to use data in order to intervene for the public good against individuals before they have offended even though this approach will single out individuals who, like the public generally, may not have committed a criminal offence, or who will perhaps not go on to commit a future offence”. They also criticised the “reliability or biases in the ‘evidence base’” and noted the consequences for “accuracy as well as the legitimacy of preventive action.”[67] They stated that the NDAS “seeks to legitimise proactive and preventative policing”, “moving law enforcement away from its traditional crime related role and into wider and deeper aspects of social and public policy.”[68]

4.14 The NDAS evidences many of the issues with predictive analytics, predictive policing, and using historic records to make future predictions. The NDAS not only carries out unacceptable and biased profiling using crude Mosaic data and inaccurate police records - it stigmatises people based on the crimes of others and their social networks. The system’s use of biased data and deeply problematic predictors is likely to result in discriminatory feedback loops, reinforcing bias and entrenching structural inequalities. These predictions and the pre-emptive interventions they trigger will result in unfair and unjust criminal justice decisions, reversing the presumption of innocence and possibly infringing people’s right to fair trial.

5. Individual-oriented crime prediction: Offender Assessment System (OASys) and the Offender Group Reconviction Scale (OGRS)

5.1 The Offender Assessment System (OASys) is a “risk and needs” automated assessment tool, developed jointly by the Prison and Probation Services.[69] It aims to assess the risk of harm offenders pose to others and how likely an offender is to reoffend, as well as assessing offender needs. These risk assessments are used to “target interventions” and to

[66] ATI DEG and IDEPP, Ethics Advisory Report for West Midlands Police, July 2017 (https://www.turing.ac.uk/sites/default/files/2018-11/turing_idepp_ethics_advisory_report_to_wmp.pdf)

[67] ATI DEG and IDEPP, Ethics Advisory Report for West Midlands Police, July 2017 (https://www.turing.ac.uk/sites/default/files/2018-11/turing_idepp_ethics_advisory_report_to_wmp.pdf)

[68] ATI DEG and IDEPP, Ethics Advisory Report for West Midlands Police, July 2017 (https://www.turing.ac.uk/sites/default/files/2018-11/turing_idepp_ethics_advisory_report_to_wmp.pdf)

[69] Prison Service Order, Offender Assessment and Sentence Management – OASys (2005) (https://www.justice.gov.uk/downloads/offenders/psipso/pso/PSO_2205_offender_assessment_and_sentence_management.doc); National Offender Management Service, ‘A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013’, (2014) (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf)

influence the sentence plans given to offenders.[70] An electronic version of the tool was rolled-out across both the prison and probation services, with a new single system being implemented in 2013 through the OASys-R project. By the end of March 2014, almost seven million prison and probation assessments had been collated within the central O-DEAT (OASys Data, Evaluation and Analysis Team) database for over one million offenders.[71]

5.2 The system collates information on the offenders' previous offences; their education, training and employment; their alcohol and drug misuse; as well as their "attitudes", "thinking and behaviour", "relationships", and "lifestyle". This is done by an assessor who assigns the offender a score based on each category.[72] This data is used alongside the individual's offending record and "offender demographic information" to inform two predictive algorithms: the OASys General reoffending Predictor v.1 (OGP1) and the OASys Violence Predictor v.1 (OVP1).[73] The Offender Group Reconviction Scale (OGRS) is another static actuarial risk assessment tool used to assess and predict an offender's likelihood of reoffending.[74] The OGRS algorithm uses data on the individual's official criminal history, as well as their age and gender, to produce a risk score between 0 and 1 of how likely an offender is to reoffend within one or two years. There have been several iterations of the OGRS since it was first used in 1996; currently OGRS4 is in use.

5.3 A 2014 National Offender Management Service analysis found that the OGP1 and OVP1 predictive algorithms generated different predictions based on race and gender. They found that relative predictive validity "was greater for female than male offenders, for white offenders than offenders of Asian, black and mixed ethnicity, and for older than younger offenders".[75] The most sustained differences were by ethnicity, with both OGP1 and OVP1 "working less well for black offenders and OGP1 also working less well for

[70] Prison Service Order, Offender Assessment and Sentence Management – OASys (2005) (https://www.justice.gov.uk/downloads/offenders/psipso/pso/PSO_2205_offender_assessment_and_sentence_management.doc)

[71] National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2014) (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf)

[72]Non-scored categories: Health and other, emotional wellbeing, financial management

[73] National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2014) (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf)

[74]https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119184256.ch11

[75] National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2014) (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf)

offenders of mixed ethnicity".[76] No assessment of different predictions by ethnicity was carried out in relation to the OGRS4 algorithm. The National Offender Management study from 2014 says there is "a clear need for further studies" to assess, among other things, "whether there are differences… according to age, gender and ethnicity".[77] The recorded disparity in prediction rates between different ethnicities is extremely concerning.

5.4 There does not appear to be any requirement to notify individuals that they have been subjected to this automated risk assessment, nor any mechanism for individuals to challenge the score or the implications it has for their involvement with the criminal justice system.

6. Identification systems: Live facial recognition surveillance

6.1 In the UK, live facial recognition surveillance has been deployed by the Metropolitan Police, South Wales Police, Greater Manchester Police, Leicestershire Police and Humberside Police.

6.2 Live facial recognition cameras scan the faces of every person that walks within the view of the camera; the system creates, even if transitorily, a biometric scan of every viewable person's face, and it compares those biometric scans to a database of images, akin to a fingerprint identification check.

6.3 Police are using this technology without a clear legal basis.[78] When Layla Moran MP posed a written question to the Home Office about current legislation regulating "the use of CCTV cameras with facial recognition and biometric tracking capabilities", Nick Hurd (then Minister for Policing, responding for the Home Office) answered: "There is no legislation regulating the use of CCTV cameras with facial recognition". The Metropolitan Police have also acknowledged that "There is currently no specific legal framework in the use of this technology."[79]

[76] National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2014) (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf)

[77] National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2014) (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf)

[78] Written parliamentary question answered by Mr Nick Hurd MP on 12 September 2017. (https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2017-09-04/8098/)

[79]https://www.london.gov.uk/press-releases/mayoral/independent-panel-delivers-report-on-polices-use

> ***"There is no legislation regulating the use of CCTV cameras with facial recognition".***
>
> Nick Hurd, (then) Minister for Policing – September 2017

6.4 Since 2016, the Metropolitan Police and South Wales Police have deployed this surveillance technology prolifically: at sports matches, concerts, shopping centres and high streets, Notting Hill Carnival, Remembrance Sunday – and even a peaceful demonstration. South Wales Police has received £2m in funding from the Home Office to lead the deployment of automated facial recognition.[80] In 2018, Greater Manchester Police deployed the technology at the Trafford Centre shopping centre for a period of 6 months in 2018 biometrically scanning an estimated 15 million people, before the Surveillance Camera Commissioner intervened.[81]

Inaccuracy

6.5 The police's facial recognition technology has been incredibly inaccurate, and there are serious problems in general with biased identification rates in facial recognition technologies.

6.6 Overall since 2016, figures obtained via freedom of information requests show that the Metropolitan Police's live facial recognition surveillance has been **93% inaccurate.** South Wales Police have deployed live facial recognition surveillance 70 times since June 2017, with **88% of its matches being inaccurate.**

Race and gender bias

6.7 A number of independent studies have found that various facial recognition algorithms have demographic accuracy biases – that is, that they misidentify some demographic groups, particularly women and people of colour, at higher rates than white men. A study found that commercial facial recognition technologies, including those created and sold by Microsoft and IBM, had error rates of up to 35% when identifying the gender of dark-skinned women compared to 1% for light-skinned men.[82] A follow up study found that Amazon's 'Rekognition' software mistook women for men 19% of the time, and darker-skinned women 31% of the time.[83]

[80] South Wales Police and Crime Commissioner, 'Medium Term Financial Strategy 2017-2021', 28 December 2016 https://pcclivewww.blob.core.windows.net/wordpress-uploads/2016-12-28-Final-Medium-Term-Financial-Strategy.pdf

[81]Working together on automatic facial recognition – Tony Porter, Surveillance Camera Commissioner, 10 October 2018 - https://videosurveillance.blog.gov.uk/2018/10/10/working-together-on-automatic-facial-recognition/

[82] http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

[83] http://www.aies-conference.com/wp-content/uploads/2019/01/AIES-19_paper_223.pdf

6.8 The Biometrics and Forensics Ethics Group warned that UK police's use of live facial recognition technology has the "*potential for biased outputs and biased decision-making on the part of system operators*".[84]

6.9 The Metropolitan Police has been aware of these concerns since 2014, when it was raised during an Association of Chief Police Officers (ACPO) 'Facial Imaging Working Group'.[85] There still remains no independent testing of demographic bias with the live facial recognition algorithm used by the Metropolitan Police. We find this unacceptable.

6.10 Nevertheless, in the Metropolitan Police's written evidence to the Science and Technology Committee, the force has admitted there are issues:

> "*The MPS is cognisant of the concern over the system response with respect to different demographics. We are working to further mitigate potential impact of this within the operational context, where it should be noted, additional checks and balances are in place and the final decision is by a human operator."*[86]

They continued, "*The MPS plans to continue to test demographic differences"* - a long overdue and confusing commitment, given that MPS has never before tested demographic differences and has thus far resisted all of our calls to do so.

6.11 The Metropolitan Police has noticed the need to "mitigate" the discriminatory impact, despite the fact that this has never been formally tested by them. However, they claim that a human review of a match prior to stopping someone is sufficient to mitigate the risk of ethnic minorities being discriminated against, which is plainly an untrue and unacceptable position.

6.12 In a presentation at University College London on 29 May 2019 about live facial recognition, the Metropolitan Police Senior Technologist, Johanna Morley, admitted that they had found significant gender bias in their technology – that it misidentified women at higher rates than men.[87]

Consequences of misidentifications

[84] Biometrics and Forensics Ethics Group, Interim report, February 2019 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/781745/Facial_Recognition_Briefing_BFEG_February_2019.pdf)

[85] Obtained through Freedom of Information Requests.

[86] Written evidence submitted by Metropolitan Police Service (WBC0005), 19 March 2019: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/the-work-of-the-biometrics-commissioner-and-the-forensic-science-regulator/written/97851.pdf

[87] https://www.ucl.ac.uk/jill-dando-institute/events/2019/may/just-looking-learning-police-trials-live-facial-recognition

6.13 Big Brother Watch has witnessed several incidents that evidence the serious and harmful potential of police live facial recognition misidentifications. At a deployment at Notting Hill Carnival in 2017, we witnessed several innocent women being misidentified as wanted men on the police watchlist. At a deployment in Romford in February 2019, a 14 year old black school child, in school uniform, was wrongly identified by the facial recognition system and subsequently surrounded by four plainclothes police officers. He was pulled onto a side street, his arms held, questioned, asked for his phone, asked for identification, and was then fingerprinted. After ten minutes of this ordeal he was released, when police realised the facial recognition 'match' was in fact a misidentification.

Infringement of fundamental rights to privacy, freedom of expression and freedom of association

6.14 Even if live facial recognition technology improves in demographic and general accuracy, which it likely will, it remains too great a risk to civil liberties to be acceptable as a general public surveillance tool in a democratic society. Live facial recognition surveillance dangerously imbalances power between citizen and state, and constitutes a fundamental threat to the right to privacy.

6.15 It is our view that police use of live facial recognition surveillance is incompatible with fundamental rights protected by the Human Rights Act 1998. Live facial recognition cameras, acting as biometric identification checkpoints, are a clear threat to both individual privacy and privacy as a social norm, as well as people's freedom of expression and association. It is plainly disproportionate to deploy a public surveillance technology by which the face of every passer-by is analysed, mapped and their identity checked. Furthermore, a facial recognition match can result in an individual being stopped in the street by the police and asked to prove their identity and thus their innocence.

6.16 We are concerned that the use of live facial recognition with CCTV has a chilling effect on people's attendance of public spaces and events, and therefore their ability to express ideas and opinions and communicate with others in those spaces. The London Policing Ethics Panel report on police live facial recognition surveillance found that 38% of 16-24 year-olds would stay away from events or places where facial recognition surveillance was being used, as well as high numbers of Black, Asian and Minority Ethnic people.[88]

Action on facial recognition surveillance

[88] http://www.policingethicspanel.london/uploads/4/4/0/7/44076193/lfr_final_report_-_may_2019.pdf

6.17 Both the police and Government have so far failed to take action to resolve any of the issues around live facial recognition surveillance, with the Government consistently stating that this is "an operational matter for the police".[89] However, the Biometrics Commissioner has rightly said that "*deciding what is proportionate should not be left to those who seek to benefit from the use of the biometric.*"[90] Meanwhile, the Science and Technology Committee has called for a moratorium on the police's use of the technology:

> "We call on the Government to issue a moratorium on the current use of facial recognition technology and no further trials should take place until a legislative framework has been introduced and guidance on trial protocols, and an oversight and evaluation system, has been established."[91]

6.18 In addition, **26 rights, race equality and technology groups,** as well as **cross party MPs** including David Davis MP, Diane Abbott MP, Ed Davey MP, and Caroline Lucas MP, have **called for an "immediate stop" to facial recognition surveillance.**[92]

Metropolitan Police operational use of live facial recognition surveillance

6.19 Despite the serious concerns around the infringements of fundamental rights to privacy, free expression and association; the likelihood for bias and discrimination, and the significant inaccuracy and related misidentifications and wrongful stops, **on 23rd January 2020 the Metropolitan Police announced that it is rolling out live facial recognition surveillance across London.**

6.20 The Met has also published new 'Standard Operating Procedures' which allows the force to put people on their watchlists who are not wanted but merely 'of interest' - a person does not even need to have a custody image to be put on a watchlist.[93]

6.21 We urge Parliamentarians to **immediately stop police and private companies using live facial recognition surveillance.**

7. Artificial intelligence and digital evidence

[89] Layla Moran, Written Parliamentary Question, 4th May 2018 (https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2018-05-04/141377/)

[90] Biometrics Commissioner, Annual Report 2017 (June 2018)

[91] https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/1970/1970.pdf

[92] https://bigbrotherwatch.org.uk/wp-content/uploads/2019/09/Statement-to-stop-live-facial-recognition-surveillance-BBW-September-2019.pdf

[93] https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/mps-lfr-sop-v1-0.pdf

7.1 It has been reported that UK police are trialling the use of AI to analyse digital evidence.[94] The Metropolitan Police has confirmed[95] that it has been exploring Cellebrite's 'Analytics Enterprise' artificial intelligence tool, which claims to "detect and match objects within images and video such as weapons, money, nudity and more", use "automatic facial detection", and "analyse links… to reveal hidden connections… and communication patterns".[96]

7.2 As this is proprietary technology created by a private for-profit company, there is very little information in the public domain about exactly how the system works or its true capabilities, such as how the system draws such 'links' within communication patterns.

7.3 We are extremely concerned that such sensitive police work is being outsourced to experimental systems, with little or no consideration of the myriad transparency, accountability and privacy issues involved. AI analysis is even being trialled to sift through victims' and witnesses' digital information, which is increasingly collected in disproportionate volumes. This raises the prospect of a victim of a sexual offence having their digital device and deeply personal information examined and analysed by an experimental, faceless AI system.

7.4 Police should not be using artificial intelligence systems to conduct such sensitive investigations.

[94] https://www.theguardian.com/uk-news/2018/may/27/police-trial-ai-software-to-help-process-mobile-phone-evidence
[95] https://www.theguardian.com/uk-news/2018/may/27/police-trial-ai-software-to-help-process-mobile-phone-evidence
[96] https://www.cellebrite.com/en/products/analytics-enterprise/

8. Solutions

8.1 A number of serious issues have been identified in this briefing in relation to bias, data protection, automated decision-making, and fundamental human rights.

8.2 Data is frequently imbued with the prejudices of prior decision makers, and these prejudices will be coded into the decisions of algorithm built using this data.[97] Discrimination can occur because the data being used represent historical patterns of discrimination – and there is no easy method to adjust historical data to rid it of this bias.[98] Even when identifiably biased data is removed from a dataset or algorithm, this does not necessarily remove bias, as other variables can introduce bias into the system by proxy. For example, postcodes are often a proxy for race and socioeconomic status.

8.3 There are so many opportunities for bias in data that it has been argued that it is unreasonable to say it can be removed. If it is decided that a system is to be used, developers should at the very least attempt to identify such issues with source datasets, consider their appropriateness, and build tools into models to identify and, if possible, mitigate that bias.[99]

8.4 Algorithms being used with significant effect in the public sector should be transparent, with auditable processes and explainable decisions so that they can be understood and challenged by those affected.

Data Protection Act 1998

8.5 Current data protection law in the UK does not adequately protect individuals against automated decision-making systems, such as those considered in this briefing.

8.6 The GDPR protects individuals against significant decisions based solely on automated processing.[100] However, the UK's Data Protection Act 2018 makes exemptions from this important GDPR right, including the right not to be subject to an automated decision set out in the GDPR – and as such fails to sufficiently protect citizens' rights.[101]

[97] Barocas, S. and Selbst, A.D., 2016. Big Data's disparate impact. *California law review*, 104, 671. (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899)

[98] Barocas, S. and Selbst, A.D., 2016. Big Data's disparate impact. *California law review*, 104, 671. (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899)

[99] Professor Suresh Venkatsubramanian presentation at the Amnesty International Expert Meeting on Predictive Policing, 20 May 2019; See also: Friedler, Scheidegger, Venkatasubramanian, Choudhary, Hamilton, Roth (2019). A comparative study of fairness-enhancing interventions in machine learning. In *ACM Conference on Fairness, Accountability and Transparency* (https://arxiv.org/abs/1802.04422)

[100] GDPR, Article 22

[101] Data Protection Act 2018, Section 49(1)

8.7 Section 14 of the Data Protection Act 2018 permits purely automated decisions with legal or similar significant effects to be made about a subject, in absence of the subject's consent – so long as the subject is notified that the decision was purely automated after the fact. The subject is then to be afforded just one month to request a new decision if they wish.

8.8 However, we are not aware of individuals being notified of purely automated decisions by any public authority, despite the prevalence of automated-decision-making systems in use described above.

8.9 This is likely because under section 14 of the Data Protection Act 2018, automated decisions that have significant legal or similar effects on a subject are not necessarily classified as "purely automated" if a human has administrative input. For example, if a human merely ticks to accept and thus enact a serious automated decision, the decision would not need to be classified as "purely automated" under law and as such, the minimal safeguards of notification and re-evaluation would not even apply.

8.10 Therefore, predictive decisions could be being made that are for all intents and purposes automated decisions, without individuals being notified of this fact or of their right to appeal. We raised concerns about this during the passage of the (then) Data Protection Bill in 2018, which were echoed by the Deputy Counsel to the Joint Committee on Human Rights who warned, *"There may be decisions taken with minimal human input that remain de facto determined by an automated process"*.[102]

8.11 The Data Protection Act 2018 in fact throws open the door for authorities to make significant decisions about people based on big data and automated processing – and weak legal definitions mean that the few safeguards there are may not even apply.

8.12 This means that UK police and other agencies in the criminal justice system are allowed by data protection law[103] to subject individuals to purely automated decisions that engage and affect people's rights. For example, current data protection law permits Durham Constabulary's HART system to not only influence decisions but to effectively make decisions about risk and thus prosecution.

[102] Note from Deputy Counsel, 'The Human Rights Implications of the Data Protection Bill', 6 December 2017 (https://www.parliament.uk/documents/joint-committees/human-rights/correspondence/2017-19/Note_Deputy_Counsel_DPBill.pdf)

[103] Data Protection Act 2018, Part 3

8.13 Big Brother Watch campaigned for amendments to the Data Protection Bill that would have ensured human decisions were ultimately required where any automated decision-making systems engage human rights.[104] We believe that such amendments would significantly improve protections for individuals against some of the issues identified in this briefing.

8.14 Two amendments are required to the Data Protection Act 2018. First, decisions that engage individuals' human rights must never be purely automated decisions. Second, automated decisions should be more clearly defined as those lacking *meaningful* human input.

We urge Parliamentarians to call on the Government to:

1. **Amend the Data Protection Act to ensure that any decisions involving automated processing that engage rights protected under the Human Rights Act 1998 are ultimately human decisions with meaningful human input.**
2. **Introduce a requirement for mandatory bias testing of any algorithms, automated processes or AI software used by the police and criminal justice system in decision-making processes.**
3. **Prohibit the use of predictive policing systems that have the potential to reinforce discriminatory and unfair policing patterns.**

[104] Big Brother Watch, 'Big Brother Watch's Briefing on the Data Protection Bill for Committee Stage in the House of Commons', March 2018 (https://bigbrotherwatch.org.uk/wp-content/uploads/2018/03/Big-Brother-Watch%E2%80%99s-Briefing-on-the-Data-Protection-Bill-for-Committee-Stage-in-the-House-of-Commons.pdf) See also: Griff Ferris, "We Must Protect Our Rights From Automated Decisions", The Huffington Post, 14 March 2018 (https://www.huffingtonpost.co.uk/entry/the-future-is-now-we-must-protect-our-rights-from_uk_5aa91fb5e4b0dccc83c1ed5b)