Volume 5, Number 1, Spring 2021

AI, Machine Learning, and International Criminal Investigations

The lessons from forensic science

Karen M. Richmond*

The evolving field of machine learning and artificial intelligence is frequently presented as a positively disruptive branch of data science whose expansion allows for improvements in the speed, efficiency, and reliability of decision-making, and whose potential is impacting across diverse zones of human activity.[1] A particular focus for development is within the criminal justice sector, and more particularly the field of international criminal justice, where AI is presented as a means to filter evidence from digital media, to perform visual analyses of satellite data, or to conduct textual analyses of judicial reporting datasets. Nonetheless, for all of its myriad potentials, the deployment of forensic machine learning and AI may also generate seemingly insoluble challenges. The critical discourse attendant upon the expansion of automated decision-making, and its social and legal consequences, resolves around two interpenetrating issues; specifically, algorithmic bias, and algorithmic opacity, the latter phenomenon being the focus of this study. It is posited that the seemingly intractable evidential challenges associated with the introduction of opaque computational machine learning algorithms, though global in nature, are neither novel nor unfamiliar. Indeed, throughout the past decade and across a multitude of jurisdictions, criminal justice systems have been required to respond to the implementation of opaque forensic algorithms, particularly in relation to complex DNA mixture analysis. Therefore, with the objective of highlighting the potential avenues of challenge which may follow from the introduction of forensic AI, this

* Postdoctoral Research Fellow, Copenhagen University [karen.richmond@jur.ku.dk]

[1] Rhiannon Jackson and Maria McAreavey, 'Black-Box Medicine: Protecting Patient Privacy Without Preventing Innovation' (2019) 3(1) Retskraft – Copenhagen Journal of Legal Studies 68.

study focusses on the prior experience of litigating, and regulating, probabilistic genotyping algorithms within the forensic science and criminal justice fields. Crucially, the study proposes that machine learning opacity constitutes an enhanced form of algorithmic opacity. Therefore, the challenges to rational fact-finding generated through the use of probabilistic genotyping software may be encountered anew, and exacerbated, through the introduction of forensic AI. In anticipating these challenges, the paper explores the distinct categories of opacity, and suggests collaborative solutions which may empower contemporary legal academics – and both legal and forensic practitioners – to set more rigorous and usable standards. The paper concludes by considering the ways in which academics, forensic scientists, and legal practitioners, particularly those working in the field of international criminal justice, might re-conceptualise these opaque technologies, opening a new field of critique and analysis. Using findings from case analyses, overarching regulatory guidance, and data drawn from empirical research interviews, this article addresses the validity, transparency, and interpretability problems, leading to a comprehensive assessment of the current challenges facing the introduction of forensic AI. It builds upon work undertaken at the Nuffield Council on Bioethics *Horizon Scanning Workshop: The future of science in crime and security* (5th July 2019, London).

1. Introduction

Technologies, writes Zuboff, 'define the horizon of our material world, as they shape the limit of what is possible and what is barely imaginable.' Their usage connotes neither neutrality nor objectivity, but rather a contingency that is 'brimming with valence and specificity in the opportunities that it creates and forecloses.'[2] Zuboff's definition encapsulates the contemporary challenges generated by the requirement to standardise, and to regulate, novel forms of machine learning (ML), and artificial intelligence (AI), both of which are the subject of sustained attention from academics, and associated institutional

[2] Shoshana Zuboff, 'Automate/Informate: The Two Faces of Intelligent Technology' (1985) 14(2) Organizational Dynamics 5, 5.

agents.[3] Thus, despite its myriad potentials, this emergent field of data science is characterised as inherently disruptive, and capable of presenting novel, and seemingly insoluble, challenges simultaneously across diverse fields. However, a review of the relevant academic literature suggests that researchers have thus far omitted to consider whether a proportion of the seemingly intractable challenges associated with the introduction of AI are as novel and unfamiliar as is frequently perceived. This article therefore addresses the omission, focusing on the forensic science and legal fields, both of which have been at the forefront of scientific development. The study considers the degree (if any), to which the courts' prior experience of standardising, and regulating, forensic algorithms within the criminal justice system, may generate insights which can aid contemporary legal academics and forensic practitioners in their efforts to set more rigorous and practical standards, with respect of this latest wave of 'disruptive' technology.[4]

The objective of the article is to highlight the implications for rational legal fact-finding, and adjudication, pursuant to the implementation of forensic and investigatory forms of AI within the criminal justice field, consequently its introduction to the international criminal courtroom by way of expert opinion evidence.[5] Whilst the potentials of AI are being explored across diverse national jurisdictions, and in heterogeneous fields such as law enforcement, forensic science, and academic research, it is posited that the international criminal justice arena represents a particularly engaging arena of analysis, given that this sector may invite investigation at a scale most suited to the mobilization of AI-driven

[3] For the purposes of this article, Artificial Intelligence is used to denote all forms of machine learning, utilising artificial neural nets (ANNs) and other forms of algorithmic computation. Machine learning thus forms a subset of artificial intelligence, as commonly understood.

[4] Thomas Buocz, 'Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary' (2018) 2(1) Retskraft – Copenhagen Journal of Legal Studies 41.

[5] See, for example, 'Scientists Developing AI to Spot Paedophiles Just From Images of Their Hands' (*The Week*, 28 February 2020) <https://www.theweek.in/news/sci-tech/2020/02/28/Scientists-developing-AI-to-spot-pedophiles-just-from-images-of-their-hands.html> accessed 27 December 2020.

efficiencies.[6] The receptivity of the international criminal justice (ICJ) sector is further enhanced by both responsiveness of the courts, when presented with evidence drawn from 'open source' data,[7] and the relative lack of procedural safeguards, particularly the absence of a gatekeeping mechanism for expert opinion evidence.[8] Thus, it is posited that the concomitant challenges associated with the deployment of AI may prove particularly impactful in the international justice arena. Nonetheless, the instant study demonstrates that such obstacles constitute a mere extension of those first encountered by national courts in relation to the use of algorithmic DNA analysis software.[9] Further, that the global challenges generated by forensic AI may be resolved in a similar fashion to those generated by the introduction of probabilistic genotyping software, through rigorous validation processes guided by overarching guidelines and regulations.

Whilst the introduction of algorithmically-derived evidence has required the mobilization of diverse bodies of expertise, in both common law and civilian jurisdictions, this study focusses on the comparatively developed and rigorous common law jurisprudence encountered in the United States and United Kingdom, in addition to those regulatory responses and guidelines published by

---

[6] Examples include the use of AI to analyse satellite data to detect the destruction of human settlements, Milena Marin, Freddie Kalaitzis and Buffy Price, 'Using Artificial Intelligence to Scale Up Human Rights Research: a Case Study on Darfur' (*Citizen Evidence Lab*, 6 July 2020) <https://citizenevidence.org/2020/07/06/using-artificial-intelligence-to-scale-up-human-rights-research-a-case-study-on-darfur/> accessed 27 December 2020. A further example is the use of AI to filter evidence from large repositories of open source data, Abishek Kumar, 'Digital Evidence and the Use of Artificial Intelligence' (*International Criminal Court Forum*, 31 May 2020) <https://iccforum.com/forum/permalink/122/33560> accessed 27 December 2020.

[7] Lindsay Freeman, 'Digital Evidence and War Crimes Prosecutions: The Impact of Digital Technologies on International Criminal Investigations and Trials' (2018) 41 Fordham Int'l LJ 283, 283–328; Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Methods for Human Rights Investigations, Advocacy and Accountability* (Oxford University Press 2020).

[8] See n 49.

[9] Julia Gasston and others, 'An Examination of Aspects of the Probabilistic Genotyping Tool: Forensic Statistical Tool' (2020) 2 WIREs Forensic Science e1362.

the European Network of Forensic Science Institutions (ENFSI),[10] the regulatory guidelines published by the Forensic Science Regulator for England and Wales,[11] and the reports of both the United States' Executive Office of the President's Council of Advisors on Science and Technology,[12] and the House of Lords' Science and Technology Select Committee.[13]

Theoretically, this article founds upon scientific theories of evidence interpretation, specifically the the Rationalist Model of Adjudication, as proposed by John Henry Wigmore, and elaborated by William Twining, its most notable contemporary proponent. According to this model,[14] the direct end of adjectival law is rectitude of decision-making through the correct application of valid law, and the accurate determination of true past facts, proved to specified standards, on the basis of careful and rational weighing of reliable evidence, presented to impartial decision-makers. This rigorous formulation forms the backdrop to a careful review of law's instrumentalisation of DNA mixtures analysis software, in its efforts to present information to the court which is beyond the common experience of the trier-of-fact. The review and analysis thus demonstrate the ways in which the introduction of computer-driven probabilistic genotyping methods in 2010 – despite having initially appeared to resolve issues generated by the increased sensitivity of DNA profiling methods – generated significant juridical challenges related to opacity and methodological validity. To add further depth to the analysis, the study draws on qualitative interview data drawn from a study of the perspectives of forensic

[10] European Network of Forensic Science Institutes, 'Guideline for Evaluative Reporting in Forensic Science' (2015) <http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdfEuropean> accessed 27 December 2020.

[11] Forensic Science Regulator, 'Software Validation For DNA Mixture Interpretation' (FSR-G-223 Issue 2, 2020) <https://www.gov.uk/government/publications/software-validation-for-dna-mixture-interpretation-fsr-g-223> accessed 27 December 2020.

[12] President's Council of Advisors on Science and Technology, 'Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods' (2016) <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf> accessed 27 December 2020.

[13] Science and Technology Select Committee, Forensic Science and the Criminal Justice System: a Blueprint for Change (HL 2017–19, 333).

[14] William Twining, *Rethinking Evidence: Exploratory Essays* (2nd edn, Cambridge University Press 2006) 72.

scientists operating within the forensic market in England and Wales. The article maintains a specific focus on the legal challenges mobilised against evidence derived from probabilistic genotyping (PG) software packages, converging on the two related methodological categories of concern; the absence of acceptable standards of validation (both developmental and internal), and the underlying lack of transparency.[15] The study demonstrates how the courts' growing appreciation of these methodological weaknesses necessitated the introduction of novel procedures, and validation protocols. Further, that in a number of instances probabilistic genotyping evidence derived from opaque algorithmic processes was ruled as wholly inadmissible in criminal trials. In substantive terms, the instant paper thus seeks to demonstrate how, and to what extent, problems traceable to a lack of foundational validity, and a lack of transparency, may re-emerge in a heightened form with the proposed implementation of AI within the forensic field. Further, that such evidential problems may become critical, particularly in relation to the deployment of 'opaque AI', since the program's algorithmic base may be manipulated recursively in order for the AI to learn, develop, and build efficiency and accuracy, through a process of trial-and-error. Crucially, this process of manipulation and change occurs beyond the threshold of human perception and control, obstructing reproducibility. When such technologies are introduced into the forensic sphere, as is currently planned, it is posited that their use may present potentially insoluble evidential problems, given that transparency and interpretability are central procedural and legal requirements, necessary in order to establish the validity of novel technologies, and expert opinions, within the courtroom.

2. Opacity

Computational algorithms are now harnessed across all sectors of human endeavour. Their capacity for efficient discrimination, and classification, has enabled them to proliferate in an environment rich in personal and trace data. Algorithms may play either a central or peripheral role, acting singly, or jointly

[15] See, for example, *Commonwealth v Foley* 38 A 3d 882, 2012 Pa Super 31 (Pa Super Ct 2012).

with other algorithms. They enable routine tasks to be performed efficiently, and serve as the engine for mundane data management tasks such as filtering 'spam' and performing internet searches. Algorithms also assume socially consequential roles, where their predictive capacities enable them to make onerous decisions, such as on an applicant's suitability for employment, or ability to repay a loan. In their most advanced iterations, computational algorithms form the cognitive drivers for machine learning systems, as utilized in facial recognition programs, or the autonomous AI of self-driving cars. So too are they deployed throughout the criminal justice sector, where the ability to make accurate categorisations is at a premium. The discriminatory capacities of computational algorithms thus form the basis for a number of forensic technologies, all of which converge around biometric discrimination. The US Government Accountability Office reports that,

> Federal law enforcement agencies … are primarily using three types of forensic algorithms to help assess whether or not evidence collected in a criminal investigation may have originated from an individual: probabilistic genotyping, latent print analysis, and face recognition.[16]

Nonetheless, the harnessing of these technologies has not been unproblematic. Concerns have arisen regarding the potential for algorithms and machine learning systems to exhibit 'algorithmic bias', or to entrench socio-economic and racial inequalities.[17] These analyses view algorithmic decision-making as a distillation of human decision-making. As such, the influence of social inequalities and biases which afflict human decision-making translate to – and are visibly encoded within – the algorithmic system, mediating its outputs. Similar concerns have similarly been raised around the propensity for algorithmic systems to exhibit behaviours which display significant deficiencies

---

[16] See United States Government Accountability Office, 'Forensic Technology: Algorithms Used in Federal Law Enforcement' (GAO-20-479SP, 12 May 2020) <https://www.gao.gov/assets/710/706849.pdf> accessed 28 December 2020.

[17] See Alexander Babuta, Marion Oswald and Christine Rinik, 'Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges' (RUSI Whitehall Report 3-18, Royal United Services Institute, September 2018).

with regard to discernibility, predictability, and tractability. These crystallise around the concept of 'algorithmic opacity.' As defined by Burrell, algorithms are opaque to the extent that '…if one is a recipient of the output of the algorithm (the classification decision), rarely does one have any concrete sense of how or why a particular classification has been arrived at from inputs.'[18] In terms of rational adjudication, these phenomena are not thereby consonant with the requirement for efficacy and reliability in relation to expert opinion evidence.

Furthermore, those algorithmic inputs may themselves be opaque, or undefined, particularly in relation to that subset of machine learning systems which manipulate their own algorithmic substructure. Opacity is thus often contraposed with the concept of 'algorithmic transparency,' and with calls for the introduction of non-proprietary 'open source' systems. These epistemological issues assume a particular significance within the field of forensic science, and criminal justice, where the 'black-boxing' of algorithmic classifications may require the trier-of-fact to accept expert assertions, absent of meaningful examination and evaluation, whilst simultaneously concealing problems relating to the foundational validity of novel scientific methods. As will be posited in the critique and analysis below, to the extent that these problems remain unaddressed, they threaten to disrupt, or subvert, fundamental principles of the law of evidence, the *ipse dixit* rule, and the overarching right to a fair trial. However, the concept of algorithmic opacity first requires elaboration, alongside an illustrative elaboration of algorithmic typology and mathematical design since, as Burrell contends, 'recognising distinct forms of opacity…is key to determining which of a variety of technical and non-technical solutions could help to prevent harm.'[19] Therefore, in the following section, discussion turns to Burrell's tripartite classification of algorithmic opacity, placing the diverse forms in a rationalist adjudicatory context.

The first category of opacity encountered is 'intentional opacity', designed into the system as a form of proprietary protection, thus intended to help maintain a market position within a competitive field, and to better enable the developer to protect 'trade secrets.' This primary variant of intentional opacity

[18] Jenna Burrell, 'How the Machine 'Thinks:' Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society, 1.
[19] ibid.

has been encountered within marketised segments of the criminal justice sector, and occupies a long-standing area of contention in criminal litigation, particularly in relation to privately developed probabilistic genotyping algorithms. A variant of intentional opacity comprises those covert forms designed to conceal the internal logics of computational algorithms, and deployed as a means to obscure 'sidestepped regulations, the manipulation of consumers, and/or patterns of discrimination.'[20] The deliberate 'black-boxing' of decision-making processes, for commercial interests, militates not only against the rationalist approach to adjudication, and the need for transparency in matters of logical inference: such obfuscation also impacts significantly on the rights of the accused and upon the principle of the equality of arms, the preservation of the latter being paramount wherever technical solutions are deployed in answer to evidentiary challenges.[21] However, the foregoing instances of 'remediable incomprehensibility'- it will be suggested – may be remedied, by the implementation of 'open source' forensic systems even if, as will be demonstrated *infra*, such a solution may offer only partial mitigation.

The secondary variant of algorithmic opacity is 'technical opacity', generated as a by-product of the high degree of specialisation and technical expertise required to design integrated computational systems. The ability to read, and write, computer code clearly requires advanced literacy in programming languages alongside a familiarity with software engineering. Translated to either the national, or international, criminal justice system, it is questionable to what extent many defence practitioners may routinely marshal the necessary skills. Proactive examples will be cited of efforts to reverse-engineer proprietary probabilistic genotyping algorithms using expert programming analysts. However these are the exception, and it is debatable to what degree such expertise is diffused across the criminal justice system. The corollary of the foregoing discussion is that the absence of diffuse expertise may potentially limit

[20] ibid 4.

[21] The Grand Chamber of the ECtHR summarized the principle of 'equality of arms' in *Edwards and Lewis v United Kingdom* (2005) 40 EHRR 24: 'It is in any event a fundamental aspect of the right to a fair trial that criminal proceedings, including the elements of such proceedings which relate to procedure, should be adversarial and that there should be equality of arms between the prosecution and defence'.

the mitigating influence of open-source solutions.[22] A more comprehensive solution may therefore be reached through transparent validation processes or, in the case of commercial suppliers, the commissioning of an independent and confidential review by an external expert.[23] The establishment of foundational validity, or failure thereof, should be the central criterion for courts to determine the reliability of expert scientific opinion, consonant with the need for rectitude of decision-making.[24]

The third variant of algorithmic opacity is 'inherent opacity', which appears as a function of the internal features and operational dynamics of algorithmic systems. It may be otiose to highlight that a number of machine learning systems operate at a scale, and a level of complexity, which renders their overall operations opaque even to those who design the discrete components incorporated within the system.[25] However, it is not the scalar element of machine learning and AI systems which generates the greatest challenges to evidential transparency. Whilst an inability to effectively limn the contours of multi-component systems presents significant obstacles to achieving 'equality of arms', the greatest challenge to tractability derives from the fundamental divergence of human, and machine, logics. Thus, the following critique and analysis must attempt to distinguish between distinct classes of algorithms, and the forms of machine logic particular to each. The first illustration focusses on a visual recognition task using a neural network. The computational algorithms used to perform these 'pattern-matching' tasks display a degree of mimesis with a human neural network, such that a number of input nodes are linked to a central set of nodes called the 'hidden layer', thence to a corresponding set of output nodes. The lines connecting the nodes are ascribed a quantitative value (or weight), and – through a rapid process of trial and error – the machine learns the optimal value for the conjoined matrix of linear weights.

---

[22] Burrell (n 18) 4.

[23] Forensic Science Regulator (n 11) 26.

[24] See, for example, the US Supreme Court Rule 702 (as amended); the English *Criminal Practice Directions 2015* [2015] EWCA Crim 1567, [2015] All ER (D) 134 (Sep), Rule 19A.5.

[25] The prime example is the 'Google' search engine.

However, when set a simple practical task, such as recognizing handwritten numerals, the most salient feature is the marked difference between the dynamics of machine logic and the ways in which human actors might disaggregate the task into a set of intelligible sub-tasks. This fundamental incommensurability between the logic of the 'hidden layer', and human cognition, 'arises from the very notion of computational 'learning.' Machine learning is applied to the sorts of problems for which encoding an explicit logic of decision-making functions very poorly.'[26] Indeed, whilst basic algorithms must be written in a way that is understandable, and logically explicable to those whose task is to develop or maintain the system, the step to machine learning may collapses that division, since the inherent feature of advanced ML and AI is the ability of learning systems to manipulate their algorithmic base. The challenges to transparency are further compounded by a secondary learning process known as back-propagation: '[back-propagation] tweaks the calculations of individual neurons in a way that lets the network learn to produce a desired output.'[27] Clearly, for the ML system, or autonomous AI, explicability – or even intelligibility to human actors – is not a concern. And it is relatively straightforward to discern the central problem: while overarching efficiencies of machine learning may be readily transposed to the criminal justice system, and in particular the forensic science field, it is clear that the inherent opacity of those machine logics may begin to generate unassailable explanatory barriers when implemented in an investigative, classificatory, or evaluative capacity.

Burrell cites a second example of the inherent opacity of machine learning systems, in this instance programs tasked with filtering 'spam' messages.[28] This model utilises algorithmic modules known as Support Vector Machines (SVMs) in order to differentiate 'spam' messages from 'non-spam', through a linear regression process. The training module learns a set of words and ascribes a weighting to each. Once again, however, it is the incommensurability of the machine logic, when performing these protocols, which generates inherent

[26] ibid 6.
[27] Will Knight, 'The Dark Secret at the Heart of AI' (*MIT Technology Review*, 11 April 2017)<https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/> accessed 28 December 2020.
[28] Burrell (n 18) 7.

opacity and diverges from human norms, since the computational algorithm is blind to any natural semiotic configuration between words, phrases, and narratives. Further, the ML does not attempt to reason with regard to the presence or absence of certain words, but rather aggregates the weightings associated with all of the words contained in a given sample. This relatively simple example once more demonstrates the counterintuitive nature of machine logic, whose inherent opacity may impact not only on our ability to explain classifications when applied to practical tasks within the legal and forensic fields, but potentially circumscribe legal and forensic research based upon discourse, and narrative, analyses. These challenges increase exponentially when opaque algorithms are incorporated into a multi-dimensional model working across a multitude of features. It is posited that Burrell's tripartite classification, as developed above, serves as a useful typology with which to analyse specific extensions of algorithmic computation, particularly the use of machine learning and AI in international criminal investigations. However, discussion first turns to the use of proprietary forensic DNA profiling algorithms, and the challenges which these generated, in order to discern ascertain whether the solutions arrived at by the courts – and allied institutional agents – may offer practical insights, whose application might reduce those risks associated with the use of opaque ML and AI systems.

3. DNA Profiling and the Criminal Justice System

The criminal justice system has been one of the foremost sectors willing to embrace the efficiencies of algorithmic and machine learning classification. Indeed, the forensic science field has, for the past decade, been at the forefront of testing and adapting innovative methods, in an effort to harness the discriminatory potentials of automated computation. One area of rapid development involves the automated interpretation and evaluation of complex DNA profiles, including DNA mixtures, degraded DNA, and trace samples. This contentious area has generated a body of criminal litigation and a rich seam of academic comment. It is posited that the creative tensions between the legal and forensic science fields, which emerged in relation to the issue of probabilistic genotyping, form a cogent base for further discussion regarding algorithmic

opacity, and the potentials of forensic AI, and machine learning. However, before proceeding with this wider critique it is first necessary to establish the underlying conceptual foundations relative to DNA profiling and analysis.

It is generally accepted that the palette of forensic techniques which together go under the term 'forensic science' do not all enjoy equal merit, exhibit similar levels of foundational validity, or are accorded comparative scientific status. Of all of these techniques – ballistics, fingerprinting, and the like – DNA profiling alone has been accorded the epistemic status of research science, a standing acknowledged by forensic scientists, academic commentators,[29] and members of the public alike.[30] Indeed, the US National Academy of Science (NAS) committee, when delineating the ambit of their 2009 study, and explaining the absence of DNA profiling within their review, noted that forensic DNA had previously been subject to two landmark studies, which had settled 'the DNA wars' and had firmly established the pedigree of forensic DNA profiling.[31] As Murphy observes, running counter to the ascendancy of DNA profiling,

---

[29] A review of the literature demonstrates that, beyond the core-set of forensic-scientific practitioners (and associated institutional actors), DNA-profiling techniques have been accorded an exceptional – if not unassailable – epistemological status. Evidence derived from DNA-profiling has been described by defence lawyers as 'infallible', or as furnishing 'irrefutable proof' [see Barry C Scheck, 'Preventing the Execution of the Innocent: Testimony Before the Senate Judiciary Committee' (2001) 29 Hofstra Law Review 1165]; by judges as a 'truth machine', or 'revelation machine' [Helena Machado and Rafaela Granja, 'Police Epistemic Culture and Boundary Work with Judicial Authorities and Forensic Scientists: the Case of Transnational DNA Data Exchange in the EU' (2019) 38 New Genetics and Society 289]; and by a prison inmate as 'God's signature'; [Michael Lynch, 'God's Signature: DNA Profiling, the New Gold Standard in Forensic Science' (2003) 27 Endeavour 93]. Such epistemic exceptionalism is not uncommon amongst the academic literature, and associated publications, devoted to forensic DNA profiling.

[30] The epistemological privileging of knowledge claims derived from such techniques is not limited to the claims of institutional actors. Prison inmate Loyd, E-J., is quoted as stating that 'DNA – deoxyribonucleic acid – is God's signature. God's signature is never a forgery.' See Jodi Wilgorin, 'Confession Had His Signature; DNA Did Not' *New York Times* (New York, 26 August 2002) A 1.

[31] Committee on Identifying the Needs of the Forensic Sciences Community, 'Strengthening Forensic Science in the United States: A Path Forward' (National Academy of Sciences 2009).

> … the traditional forensic disciplines that had long served as the backbone of scientific evidence in the courtroom, and continued to make up the majority of the scientific evidence in criminal cases, went largely ignored despite loud pleas from a dedicated coterie within the scholarly and scientific community.[32]

Thus, forensic DNA was presented as the paradigm forensic technique, uniquely scientific, the benchmark forensic science discipline, and the purpose of the NAS report was therefore to provide the groundwork for the residuary categories of forensic techniques to meet the scientific standards set by DNA, in order that they might establish similarly robust epistemic credentials. Murphy rightly highlights the difference between 'first generation' pattern-matching techniques, and 'second generation' bio-identification sciences, and sheds light on the way in which DNA became to be regarded as a '*sine qua non*'. With regard to single source DNA, this is a convincing analysis. However, when probabilistic genotyping of mixed samples is factored into this analysis, the picture changes. Absent from Murphy's critique as presented here (though the subject of trenchant analysis throughout her work) is the conception that DNA may itself be fallible, affected by technological developments, or influenced by alterations to overarching governance structures. Indeed, it is necessary to stress that later iterations of DNA profiling techniques must continue to establish a basic foundational validity which meets legal standards and the overarching objectives of the NAS Report.

4. Mixtures and Low Template DNA

At this stage, it should be re-iterated that the basic DNA profiling protocols, on which the above perceptions are based, had been subject to thorough validation and accreditation procedures, and had established reliable scientific underpinnings. In contrast, even though pattern-matching techniques present their conclusions in terms of a 'match/non-match', such unique categorisations

[32] Erin Murphy, 'What "Strengthening Forensic Science" Today Means for Tomorrow: DNA Exceptionalism and the 2009 NAS Report' (2010) 9 Law, Probability and Risk 7.

lack a scientific basis, being non-probabilistic, open to significant bias, and unable to articulate established error rates. The reason that 'single-source and simple-mixture sample analyses are considered highly reliable [is] because each of the steps involved in the analysis is 'repeatable, reproducible, and accurate.' This trio of requirements is referred to as 'foundational validity.'[33] However, the same foundational validity, based on a high degree of trust in the accuracy of results, is neither exhibited by first generation techniques, nor capable of extension to more complex processes, such as those involving minute traces of 'low template' DNA, or degraded DNA, especially where these involve the interpretation of 'DNA mixtures' drawn from a number of individuals.

The occurrence of DNA mixtures has risen sharply since the introduction of sensitive testing protocols (such as DNA-17 and Globafiler-24, both of which replaced the less sensitive SGM Plus system).[34] These protocols are now capable of picking up trace amounts of 'low template' DNA, their use leading to the routine reporting of mixed DNA profiles. Complex mixtures undergo the same forms of processing as simple, or single-source DNA samples. In short, the sample is stabilised, and amplified. Scientists then use standardised procedures to count the numbers of Short Tandem Repeats (STRs are polymorphisms, or areas which exhibit a high degree of variation) at a number of loci, or sites, on the DNA. A graphical output displays each loci as a peak whose height is a product of the number of STRs at that site. Together these peaks create a DNA profile which can be rendered numerically, for statistical analysis against background population data.

However, in the case of DNA mixtures, these require deconvolution, and the interpretation of the results may display significant levels of variation, not least as the set of superimposed peaks require to be carefully evaluated in order to

[33] Katherine Kwong, 'The Algorithm Says You Did It: The Use of Black Box Algorithms to Analyze Complex DNA Evidence' (2017) 31 Harv JL & Tech 275, 277.

[34] See, for example, Matthew J Ludeman and others, 'Developmental Validation of GlobalFiler™ PCR Amplification Kit: A 6-Dye Multiplex Assay Designed for Amplification of Casework Samples' (2018) 132 International Journal of Legal Medicine 1555.

determine whether a suspect profile is included.[35] This can be achieved manually, and mathematically. Alternatively, probabilistic genotyping (PG) programs may be utilised. These computerised mathematical models and simulations estimate the likelihood that a particular individual's DNA is part of the mixture present in the sample. Although the preponderance of PG systems (and subsequent cases cited) emanate from the United States, it should be noted that the issues raised affect forensic practice in a multitude of jurisdictions. For example, empirical research in the UK revealed similar concerns regarding the use of probabilistic genotyping algorithms to de-convolute mixed DNA profiles as those raised in the literature, particularly with regard to validation.

> There are two different types. Cellmark uses David Balding's [open source LikeLTD] system. LGC developed LiRA. These systems can deal with two or more people, though for a while Balding's system wasn't validated – it is now. There are differences between the systems but the same system can deliver different answers depending on how the question is formed.
>
> (Interview with Lead Scientist: Oxford, 2015)

This typical response (drawn from 33 semi-structured interviews with DNA profiling scientists and allied institutional agents), supports the claim of levels of scepticism amongst groups of experts with regard to the scientific validity and operational dynamics of algorithmic forms of probabilistic genotyping. Such scepticism also focusses on the need to establish foundational validity within the courtroom. Further, to ensure that the operator inputs – including the framing of propositions – are explicitly noted in order to facilitate transparency and reproducibility.[36] The following section elaborates on these concerns, analysing

---

[35] See Rich Press, 'DNA Mixtures: A Forensic Science Explainer' (National Institute of Standards and Technology, 3 April 2019) <https://www.nist.gov/featured-stories/dna-mixtures-forensic-science-explainer> accessed 28 December 2020.

[36] Whilst a variation in output consequent to a variation in input is hardly problematic, within the forensic and legal context, the propositions on which probabilistic

the use of PG software in the courtroom with reference to a number of case studies, and utilising Burrell's tripartite classification in order to discern the forms of opacity encountered therein. It goes on to evaluate the implications of the generation of particular forms of opacity for the exercise of rational fact-finding and legal adjudication.

5. Probabilistic Genotyping Software Case Studies

The first example of forensic-algorithmic opacity focusses on the use of a probabilistic genotyping package known as the Forensic Statistical Tool (FST). This software system was developed by the New York City Office of the Chief Medical Examiner (OCME). Introduced in 2010, the OCME began to routinely use the FST in tandem with high sensitivity testing (HST) in cases which involved mixed, trace, and/or degraded, samples. Indeed, the laboratory stated that it had used High Sensitivity Testing (HST) in 3450 cases between 2006, and 2017. Further, that it had used the Forensic Statistical Tool in 1350 cases between 2011 and 2017. However, for nearly six years, between 2010 and 2016, defense requests to conduct independent expert witness reviews of this in-house proprietary software (including the source code, supporting development material, and executable software versions) were denied, even where the request involved an audit under protective order. When, in 2016, the source code was first reviewed, several problems were encountered, not least a previously undisclosed data-dropping function which discarded evidence of potential value to the defence. In later studies, which focused on the quantitative impact of the undisclosed function on the original validation data of 439 samples, it was found that the data-drop was triggered in 23.7% of cases (104 samples). The overall

calculations proceed must be addressed carefully in order to elicit an accurate answer to the particular question which is being asked in relation to the evidence eg whether the DNA sample was deposited by a particular source, as opposed to through a particular activity. That process must meet the same requirement for transparency as that pertaining to the calculation itself. See Forensic Science Regulator (n 11) 17.

effect was 'to skew results towards false inclusion for individuals whose DNA was not present in the evidence sample.'[37]

A landmark case involving the FST followed an assault on an individual in Brooklyn, New York, in 2013.[38] In the wake of a brawl in a Hasidic Jewish district,[39] during which an African American male was seriously injured by a number of assailants, a shoe was recovered, and sent to the NYC Medical Examiner's office for testing. When an area of the shoe was swabbed, a mixed DNA sample from two individuals was recovered. The sample size was 97.9 picograms, which was below the lower limit for standard DNA processing (100pg).[40] Therefore the sample was also subjected to high-sensitivity testing (HST), which extrapolated the size of the sample by reproducing it. Ordinarily samples underwent 28 cycles of amplification. However, HST samples underwent 31 cycles. This boosted the sample size but also served to amplify any latent defects and artefacts. The resulting sample was then subjected to probabilistic genotyping, analysed using the FST. The OCME stated that the two-person mixture contained both the victim's DNA, and that of the accused, with an attendant probabilistic determination of 133 to 1. The accused was convicted but the verdict was overturned on appeal, the evidence from the FST being described as 'less than convincing.' The reasoning was based on the OCME's combining two testing methods which both lacked foundational validity. Further, the unsuitability of the FST calculations when applied to a suspect drawn from a genetically homogenous population. Thirdly, due to the fact that the technician had altered the testing parameters. For the purposes if the instant study, it should be noted that throughout this case the OCME

---

[37] Jeanna Matthews and others, 'The Right to Confront Your Accusers: Opening the Black Box of Forensic DNA Software' in American Association for Artificial Intelligence and Association for Computing Machinery, *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (United States Association for Computing Machinery 2019) 321.

[38] *People v Herskovic* 2018 NY Slip Op 06763.

[39] The ethnicity of the victim and accused is an important consideration, when attempting to derive a statistical output from a DNA profile measured against a population database.

[40] A picogram (pg) is one trillionth of a gram, or 0.000000000000001 kilogram (SI unit).

vigorously opposed examination of its FST source code. Nonetheless, a comprehensive code audit was later conducted, which unearthed significant problematic features.

The cases involving the FST, and the opaque features exposed by the subsequent quantitative code audit, exhibit the first, second, and third, categories of algorithmic opacity, relating respectively to intentional, technical, and inherent opacity. Firstly, while it must be noted that the OCME was not operating within a competitive market, and had no commercial proprietary interest in the FST, significant efforts were made to avoid regulatory, and legal, oversight. That regulatory oversight would have required the independent validation and adversarial testing of the software (and development material) and publication of results. Next, the FST cases provide an example of technical opacity, deriving – initially – from the comparative lack of technical awareness and literacy amongst defendants, and public defenders, compounded with a dearth of resources necessary to address these issues. Lastly, the FST case displayed a form of inherent opacity. This related to a data-discard function which had been introduced during development, as an improvised solution to resolve other software issues, and contravened both the published methodology of the FST, and that promulgated in oral evidence.

These themes, involving lack of validation and opposition to oversight, would recur in subsequent cases involving commercial PG software packages, as detailed below. However, it is first necessary to place the foregoing analysis in a legal and regulatory context. As stated, *supra*, this analysis gauges the purported validity of PG software variants (and prospective forensic AI developments) in correspondence with rationalist evidentiary norms, instantiated through the comprehensive regulatory requirements laid down by the US PCAST report, the ENFSI '*Guidelines for Evaluative Reporting in Forensic Science'* and the UK Forensic Science Regulator's *'Guidance on Software Validation for DNA Mixture Interpretation.*'[41] The FSR guidance[42] offers a number of solutions aimed at ensuring that the development, validation, and use, of proprietary forensic software conforms to the highest standards. The guidance now requires oversight

[41] Forensic Science Regulator (n 11).

[42] The FSR guidance is itself based upon the preceding PCAST report, see n 12.

involving routine operating quality checks and addresses data input considerations. Thus, minimum standards are now specified for a DNA profile to be considered suitable for interpretation, and criteria for reports now requires that all relevant information used in the calculations be included, in addition to 'the alternative scenarios considered to facilitate checking, auditing and defence review, and the reproduction of results.'[43] Further, the population genetic issues which surfaced in the *Herskovic* case have been addressed, the guidance stating that, '…in relation to population genetic issues, the ability to specify a range of ethnic databases is essential.'[44] In procedural terms, this stipulation answers the need to provide comprehensive background data in relation to those variables which may influence the result of a particular forensic calculation. In summary, these technical requirements together constitute a quality management framework which embeds transparency into all stages. Further, it ensures that technical opacity is addressed through stringent reporting requirements which, also known error rates. Discussion now turns to the legal and regulatory responses triggered by the paradigm example of intentional opacity in proprietary forensic software.

The zenith of protection of proprietary interest protectionism was reached in the case of *Commonwealth v Foley*,[45] the first case to challenge the foundational validity and scientific pedigree of a commercial PG software system. This case involved the assault and murder of a dentist at his home. A mixed sample of DNA from two individuals – presumably the victim and the murderer – was recovered from under the victim's fingernails. Three experts testified that the DNA was consistent with that of the accused, a state trooper who had been living with the victim's estranged wife. However, the experts' probabilistic determinations differed by several orders of magnitude, ranging from 1 in 13,000 to 1 in 189 billion. The latter statistic was arrived at by using a proprietary software package (TrueAllele) marketed by Cybergenetics, a company owned by one of the reporting scientists. The defence challenged the expert's testimony on the grounds that this automated PG approach constituted

[43] Forensic Science Regulator (n 11) 16.
[44] ibid 17.
[45] *Commonwealth v Foley* (n 15)

a novel and unproven method.[46] Further, they requested the release of the source code in order to conduct validation tests. The courts ruled against the *Frye* challenge and denied access to the proprietary algorithms on commercial grounds, stating that, 'TrueAllele is proprietary software. It would not be possible to market TrueAllele if it were available for free.'[47]

Further, the court in *Commonwealth v* Foley, stated that scientists were not in any case prevented from assessing the reliability of a software package absent the release of the source code, accepting the argument proffered by the makers of TrueAllele that the publication of the results of internal validation studies in peer-reviewed journals signaled that the scientific community had debated, and accepted, the scientific foundations of the PG package. Thus, TrueAllele was held to have met the US *Daubert* test[48] for expert scientific evidence. However,

[46] The US courts introduced the Frye standard (*Frye v United States*, 293 F 1013 (DC Cir 1923)) in order to determine the admissibility of expert opinion evidence. This test holds that expert testimony based upon scientific techniques is only admissible when these techniques have become generally accepted within the relevant scientific community. It has now been superseded in the preponderance of US states by the *Daubert* test, discussed *infra.*

[47] *Commonwealth v Foley* (n 15) 889.

[48] Following the judgment in *Daubert v Merrel Dow Pharmaceuticals* 509 US 579 (1994), the Supreme Court amended Rule 702 (regarding the use of expert testimony) to introduce the *Daubert* admissibility test. Within the preponderance of US states, all expert opinion evidence must now meet the *Daubert* standard, measured against five criteria. Daubert requires that, in judging the admissibility of expert evidence, the court must look to the underlying methods used, in order to assess: whether a method can or has been tested; the known or potential rate of error; whether the methods have been subjected to peer review; whether there are standards controlling the technique's operation; and, the general acceptance of the method within the relevant community. Thus, the judge exercises a gate-keeping function, and must now ensure that all expert testimony 'proceeds from scientific knowledge''. It should also be noted that the UK now employs an 'enhanced *Daubert*' test, see Tony Ward, 'An English *Daubert*? Law, Forensic Science and Epistemic Deference' (2015) 15(1) Journal of Philosophy, Science and Law 26. See also, Karen M Richmond, 'The Forensic Regulator Bill: Articulating Normative Standards in a Forensic Market' in K Jakobs and D-H Kim, (eds), *Proceedings of the 25th EURAS Annual Standardisation Conference: Standards for Digital Transformation: Blockchain and Innovation* (Verlag Mainz 2020) 245–59.

as Kwong[49] argues (elaborating upon oblique criticisms contained in a PCAST report),[50]

> … having internal validation studies published in peer-reviewed journals does not mean that the scientific community has debated and accepted the science involved; it merely indicates that the peer reviewers did not identify any disqualifying characteristics of the study as it was described by the paper, such as obvious methodological errors or inaccurate analysis [of the reported results].

Utilising Burrell's typology of algorithmic opacity, the cases involving TrueAllele can be said to exhibit intentional opacity, deployed both to maintain market position, and to avoid legal oversight and review. Indeed, the *Foley* case is most notable for the placing of proprietary interests above the rights of the accused. However, it was far from a sole instance of private interests trumping fundamental rights. As of 2017, all defence requests to view the TrueAllele source code had been defeated, or were being vigorously opposed.[51] As for the inherent opacity of the TrueAllele system, it should be noted that the validation studies for this PG package only accounted for use within narrow, pre-defined parameters. However, the system has subsequently been operated outside the validation parameters. Thus, development, application, and a concomitant extension beyond the validated methodological boundaries can, in this instance, be seen to generate inherent opacity. Further, whilst the designers of TrueAllele have claimed that it is 'impossible' for the package to return a false positive,[52] others have been more circumspect about the possibility of error.[53]

---

[49] Kwong (n 33) 289.
[50] President's Council of Advisors on Science and Technology, 'Forensic Science in Criminal Courts' (n 12).
[51] Kwong (n 33) 292.
[52] See Exec. Office of the President, President's Council of Advisors on Science and Technology, *An addendum to the PCAST Report on Forensic Science in Criminal Courts* 8 (2017) at 8President's Council of Advisors on Science and Technology, 'An Addendum to the PCAST Report on Forensic Science in Criminal Courts' (2017) 8.
[53] Kwong (n 33) 290.

The legal and regulatory guidance specified in relation to the FST, *infra*, remains pertinent. In this instance, the guidance places an onus upon the developer to explicitly acknowledge errors and mistakes, particularly in relation to the overall error rate, and to analytical mistakes, for example: whether the model on which the software is based rests on unjustifiable assumptions; and whether mistakes in software coding result in inaccuracy and unreliability of function.[54]

The requirement of transparency is placed within a framework for end-to-end validation, which encompasses both conceptual, and end-user, validation. The process commences with the requirement to establish conceptual validity which states that, when publishing developmental studies,

> ideally the underlying data on which conclusions are based should also be made available, for example, as supplementary material within the journal or access provided online to downloadable material including all data and a full statistical description. This enables other scientists in the field to inspect it independently and verify the results obtained in order to enable general acceptance of the model concept within the scientific community. Such transparency is essential for any software used within the CJS, for which there can be no 'secret science'.[55]

At the other extreme, the guidance requires end-user validation from the court reporting officers, who need to be satisfied, through the provision of full validation documentation – plus formal assessment and authorisation by their respective organisations – that the software they are relying upon to provide expert opinion is fit for purpose and will not result in misdirection of the court.[56] Indeed, some developers of proprietary software systems have striven to meet the required levels for transparency, and to address known errors in their source code. A notable example occurred in relation to STRMix (a proprietary software package designed by New Zealand's Crown Research Institute, in collaboration

[54] Forensic Science Regulator (n 11) 24.
[55] ibid 26.
[56] This requirement is encapsulated in the Criminal Practice Directions Rule 19A.6(b), and the Federal Rules of Evidence Rule 702.

with Forensic Science South Australia), whose makers drew attention to two coding errors, the inclusion of which had affected the results of DNA analyses in a significant proportion of criminal cases.[57] Further, STRMix has released its source code to defense teams for inspection subject to a confidentiality agreement. Whilst this provides a rare instance of intentional transparency, it nonetheless supports the apprehension of inherent opacity, as endemic to complex algorithmic systems. In the final section, discussion turns to the legal implications of such algorithmic opacity, and discusses the implications for forensic AI packages.

6. Legal Implications and Solutions

As demonstrated above, the introduction of 'black-boxed' algorithmic decision-making systems have given rise to a number of inter-related legal issues. These crystallise around one question, appositely framed by Jeanna Matthews; 'In a society that purports to guarantee defendants the right to face their accusers and confront the evidence against them, what then is the role of black-box forensic software systems in…decision-making in forensic science?'[58] The question surfaces the inherent tensions between resort to algorithmic efficiency, and the paramount importance of established legal principles: the right to a fair and public trial; the rights of accused persons to review and confront the evidence against them; and the right to equal justice under the law. It is argued that there are few circumstances which might be envisaged in which the former should supercede the latter. Indeed, as has been demonstrated, such supersession may run counter not just to legal principle, but to the procedural rules of evidence. As Murphy argued in relation to the courts' protection of proprietary interests in the TrueAllele cases, 'courts would not accept opinions from witnesses not shown to have the qualifications as an expert, so, too, should courts not accept opinions from digital 'experts' without probing the 'qualifications' of the technology.'[59] It may be further argued that the true issue extends beyond the

[57] Kwong (n 33) 292.
[58] Matthews and others (n 37) 321
[59] Erin Murphy, *Inside the Cell: The Dark Side of Forensic DNA* (Nation Books 2015).

'qualifications' of digital experts, which may have been widely accepted within the scientific community, and whose use may have been uncontroversial, at least to the extent that they remained unchallenged. Rather, in relation to AI and advanced machine learning systems, the 'opinions' of algorithmic experts are a direct product of their opaque underlying methodologies. As such these outputs constitute a 'digital *ipse dixit*.' The *ipse dixit* rule, a prohibition of arguments from authority and unsupported expert opinion, extends across multiple legal systems and domains, restricting experts from offering unsupported opinion evidence.[60] The principle focusses neither on the expertise, nor the experience, of the witness but rather on the underlying methodology on which the expert claims are based. Thus, claims from expertise and experience may be validly proffered, provided that such claims are supported by a clear explanation of how experience leads to conclusion; why experience is a sufficient basis for such testimony; and how said experience may be reliably applied to the facts.[61] In the context of algorithmic decision-making, and forensic AI, it is posited that this elementary duty to provide support for an assertion cannot be discharged, or avoided, absent of the rigorous validation processes detailed, *infra*.

It remains to consider the implications, and possible solutions, for forensic ML, and AI, applications. Given the above, it is clear that the introduction of machine learning processes within the forensic, and (international) criminal justice fields, may compound the problems already posed by tertiary forms of algorithmic opacity. This applies to both pattern-matching techniques, which lack the foundational validity of DNA profiling, and to attempts to use quantitative analyses, or visual recognition, in order to process mixed DNA profiles, or to filter open source data. A number of solutions legal and technical solutions present themselves. First, the use of such approaches may be controlled by way of legislative intervention, aimed at limiting or regulating their use.

---

[60] Michael J Saks, 'Banishing *Ipse Dixit*: The Impact of *Kumho Tire* on Forensic Identification Science' (2000) 57 Wash & Lee L Rev 879.

[61] See *United States v Frazier* 387 F 3d, 1244 (11th Circuit 2004) (*en banc*), in which scientific opinion evidence was excluded, the forensic specialist having failed to establish the methodological reliability of his opinion.

Indeed, the European Commission White Paper on Artificial Intelligence[62] makes a number of recommendations in this area. These recommendations reflect seven key requirements listed by the High-Level Expert Group.[63] Of these seven, the Commission identifies a lack of transparency in AI as a particular risk, positing that existing EU, and national, legislative frameworks could be improved in order to address the current lack of oversight in this area. The Commission expressed particular concerns over the use of opaque AI in the private sphere, stating that,

> The lack of transparency (opaqueness of AI) makes it difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation. Therefore, in order to ensure an effective application and enforcement, it may be necessary to adjust or clarify legislation in certain areas.[64]

The Commission uses the term 'high-risk AI systems' when addressing those systems whose capabilities, functional protocols, and limitations are not explicitly articulated.[65] It is proposed that the legal response may be extended to the international criminal justice arena. However, as discussed, *supra*, softer legal and regulatory responses have been promulgated, such as the use of software audits, and open source systems. However, these solutions may be limited by a lack of requisite expertise, and a lack of diffuse experience across legal jurisdictions. In addition, more general developments in legal and forensic training, might serve to address the need for improved interdisciplinary communication, and the need to compass the normative and epistemological

---

[62] Commission, 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust' COM (2020) 65 final.

[63] The 2019 experts group lists seven key requirements under the following heads: Human agency and oversight; Technical robustness and safety; Privacy and data governance; Transparency; Diversity, non-discrimination and fairness; Societal and environmental wellbeing, and; Accountability.

[64] COM (2020) 65 final (n 62) 14.

[65] See Riikka Koulu, 'Human Control over Automation: EU Policy and AI Ethics' (2020) 12(1) European Journal of Legal Studies 9.

requirements of allied fields.[66] Technological 'solutions' – for example a resort to 'constrained AI'[67] – may be attempted. However, these involve a significant compromise in efficiency whilst failing to eliminate the risks exposited above. In conclusion, none of these solutions should be approached in isolation. Indeed, Matthews recommends that 'both in research and in casework, an emphasis should be placed on comparisons between multiple reasonable systems' evaluations of the same input data.'[68]

The comparative lack of diffuse expertise within the international criminal justice sector may cause further complications. In relation to evidence handling, the ICJ sector is notable for a marked spatial and temporal divergence separating evidence collection, stabilization, evaluation, and reporting. In a domestic jurisdiction these processes are approached holistically, through the joint efforts of forensic experts and allied institutional agents, who together shape the evidential trajectory. However, in the context of alleged international crimes there exists a fundamental bifurcation between collection and stabilisation of evidence – particularly in relation to open source data collected and filtered by members of the public and NGOs – and its subsequent evaluation and reporting by prosecution experts. Whilst proponents of open source investigation may highlight the potentials of emergent open source data collection and processing systems to furnish the international courts with evidence, in light of the foregoing discussion it may be stated with relative certainty that by placing forensic AI systems in the hands of uncertified volunteers, their functions are comparatively less amenable to control. Therefore, to conform with regulatory

[66] See Chris Lawless, 'A Curious Reconstruction? The Shaping of "Marketized" Forensic Science' (2010) CARR Discussion Paper 63; Christopher James Lawless, 'Policing Markets: The Contested Shaping of Neo-Liberal Forensic Science' (2011) 51 British Journal of Criminology 671; Sally F Kelty, Roberta Julian and Alastair Ross, 'Dismantling the Justice Silos: Avoiding the Pitfalls and Reaping the Benefits of Information-Sharing between Forensic Science, Medicine and Law' (2013) 230 Forensic Science International 8; The Rt Hon the Lord Thomas of Cwmgiedd, 'The Legal Framework for More Robust Forensic Science Evidence' (2015) 370 Philosophical Transactions of the Royal Society B 20140258, 1.

[67] Constrained AI founds on paramaterised algorithms operating within limits set by the operator. These are utilized in an attempt to increase the tractability of machine learning and AI processes.

[68] Matthews and others (n 37) 322.

guidance, levels of access should be imposed such that only the input variables can be defined by the operator, 'whilst access to files that define the analytical parameters would require a higher level of authorisation. System access logs, settings changes and parameters used for past tests should be auditable.'[69]

This leads to a broader issue, which goes beyond the fundamental need for transparency and accuracy in forensic reporting In the context of a criminal investigation, a calculation should proceed only if the software is capable of aiding a meaningful interpretation. It should be borne in mind that while the efficiencies offered by machine learning may prove increasingly attractive to researchers and practitioners, academics have aptly demonstrated that efficiencies drawn from mathematical expertise and human endeavor are still capable of delivering the most accurate and transparent efficiencies.[70] Thus, the international criminal justice system should be particularly circumspect in its engagement with novel but opaque technologies whose underlying methodologies resist exegesis. In the allied fields of international criminal justice, legal research, and forensic science – where the interpretability of results, and the explicability of propositional foundations, are at a premium – the utilisation of machine learning, and AI systems, should be exercised with caution, particularly in respect of thee more complex, and comparatively opaque, instantiations. The efficient processing of data must be tempered by 'healthy skepticism about the design, development, and use of complex software systems used in criminal justice.'[71] Otherwise, the established principles of rational inference, rectitude of adjudication, and legal order, negotiated collectively over centuries, could be fatally undermined by the introduction of automated systems whose logics cannot be explained.

---

[69] Forensic Science Regulator (n 11) 19.

[70] Therese Graversen and Steffen Lauritzen, 'Computational Aspects of DNA Mixture Analysis' (2015) 25 Statistics and Computing 527. See Faculty of Science, 'Danish DNA Detective Helps English Police in Homicide Cases' (*University of Copenhagen*, 23 May 2018) <https://www.science.ku.dk/english/press/news/2018/danish-dna-detective-helps-english-police-in-homicide-cases/> accessed 17 January 2021.

[71] Matthews and others (n 37) 322.