# ARTIFICIAL INTELLIGENCE AND EVIDENCE

The proliferation and use of artificial intelligence ("AI") systems that are powered by machine learning ("ML") to gather and process information means that admitting such evidence will raise issues not only about the admissibility of electronic evidence but also about the limitations inherent in ML. The treatment of the presumption of reliability of computer systems, including AI systems, as a conclusive, legal presumption fails to understand that software systems can produce subtle mistakes that are not obvious. This is compounded by the fact that the non-procedural nature of ML and AI systems amplifies the difficulty of proving or disproving the reliability of AI systems. The fact that ML and AI systems produce results from datasets that contain embedded human assertions also means that the application of the hearsay rule to AI output may be more apposite than previously thought. The authentication of electronic evidence should be subject to a clear procedure to be developed by the courts, especially in an era of "deepfakes" and other digitally manipulated data. The article concludes with a look at the issues in discovery and disclosure related to voluminous electronic evidence empowered by the use of predictive coding and the need to conduct an "examination" of software code.

Daniel **SENG**
*LLB (Singapore), BCL (Oxford), JSM (Stanford), JSD (Stanford);*
*Advocate and Solicitor (Singapore);*
*Director, Centre for Technology, Robotics, Artificial Intelligence & the Law; Director, LLM Programme in IP and Technology Law,*
*Associate Professor, Faculty of Law, National University of Singapore.*

Stephen **MASON**
*BA (Hons) (CNAA), MA (City), LLM (London), PGCE(FE) (Greenwich);*
*Barrister (Middle Temple);*
*Associate Research Fellow at the Institute of Advanced Legal Studies, School of Advanced Study, University of London.*

## I. Introduction

1 With all the publicity regarding artificial intelligence ("AI") and the fourth industrial revolution, it is important that any discussion regarding AI has to start with a definition, because the term can mean different things to different people. "AI" is defined here as a system that

acts "intelligently" by doing what is appropriate for the circumstances and the purposes assigned to it, including behaving flexibly in changing environments and objectives, learning from experience and making appropriate choices given perceptual limitations and finite computation.[1] AI can be further categorised as:[2]

> (a) **Narrow AI**: AI developed as an aid to human thought, typically through the use of a system that solves tightly constrained problems;
>
> (b) **Strong AI**: AI that attempts to mechanise human-level intelligence, typically when a system is used as a general purpose problem solver; and
>
> (c) **Artificial General Intelligence ("AGI")**: an AI system that greatly exceeds the cognitive performance of humans in virtually all domains of interest.

2 In spite of many of the significant advancements in their sophistication and application today, AGI and even Strong AI systems do not exist yet.[3] Narrow AI, however, does not mean that AI systems that we have built are not powerful or are not able to perform useful tasks. It means that the AI systems that we have designed and are currently using are of narrow application. This includes their widespread use in automatic application processing, anomaly detection, and speech, language and visual pattern recognition for business and public processes. For instance, smart banking systems review and approve credit card and loan applications. Image recognition systems and cameras collect information about individuals and vehicles and track our movements. Electronic commerce platforms process our shopping preferences to generate advertisements and product recommendations. In fact, today, AI systems are gathering, processing and producing much of the information that is integral to the functioning of a modern society.

## II. AI and electronic evidence in context

3 This article concerns evidence gathered and processed using AI systems. AI evidence is first and foremost evidence in electronic form.

---

1 David Poole, Alan Mackworth & Randy Goebel, *Computational Intelligence: A Logical Approach* (Oxford University Press, 1998) at p 1.

2 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014) at p 18 (all three definitions). See also Kathleen Walch, "Rethinking Weak vs. Strong AI" *Forbes* (4 October 2019).

3 "The Life Scientific" *BBC Radio 4* (5 November 2019), presented by Prof Jim Al-Khalili and produced by Anna Buckley, with Demis Hassabis speaking about artificial intelligence.

Because electronic evidence is by its very nature capable of being altered or even deleted, deliberately or inadvertently, the main consideration is to ensure that such evidence and the systems that store, process and analyse them are trustworthy and reliable.[4] This pervades the very nature of the evidence itself – from its inception to its use in legal proceedings – and encompasses the systems in which it is stored, processed and analysed. To take law enforcement as an example, today, police officers equipped with body-worn video cameras and patrol cars equipped with in-car cameras record crucial evidence in real time.[5] For investigative work, police use cell phone tracking software and case-management software to streamline the collection and analysis of collected evidence.[6] And prosecutors, lawyers and judges use case tracking and management systems to manage case filing, information, caseloads and dockets.[7] Even as there is a move away from just automated systems to AI systems by investigative agencies for policing,[8] intelligence gathering[9] and evidential

4 For examples, see *Electronic Evidence* (Stephen Mason & Daniel Seng eds) (Institute of Advanced Legal Studies for the SAS Humanities Digital Library, School of Advanced Study, University of London, 4th Ed, 2017), open access in the Humanities Digital Library <http://ials.sas.ac.uk/digital/humanities-digital-library/observing-law-ials-open-book-service-law/electronic-evidence> (accessed 15 July 2020) (hereinafter "*Electronic Evidence* (4th Ed)").

5 Ben Bowling & Shruti Iyer, "Automated Policing: The Case of Body-worn Video" (2019) 15(2) *International Journal of Law in Context* 140; *Director of Public Prosecutions v Young* [2018] EWHC 3616 (Admin) (accepting body-worn video as evidence). See also Joyce Lim, "Police to Start Wearing Body Cameras from Friday" *The Straits Times* (29 January 2015); Laura Elizabeth Philomin, "Police Get Extra 'Eyes' through 760 Participating In-car Cameras" *Today* (17 May 2015).

6 Kevin Strom, *Research on the Impact of Technology on Policing Strategy in the 21st Century, Final Report* (National Criminal Justice Reference Service, September 2017) <https://www.ncjrs.gov/pdffiles1/nij/grants/251140.pdf> (accessed 15 July 2020).

7 See, *eg*, Marco Fabri & Francesco Contini, *Justice and Technology in Europe: How ICT Is Changing the Judicial Business* (Kluwer Law International, 2001); April Pattavina, *Information Technology and the Criminal Justice System* (SAGE, 2005); Richard Magnus, "The Confluence of Law and Policy in Leveraging Technology: Singapore Judiciary's Experience" (2004) 12 Wm & Mary Bill Rts J 661.

8 See *Bridges, R (On Application of) v The Chief Constable of South Wales Police* [2020] 1 WLR 672; [2019] EWHC 2341 (Admin) (regarding a challenge to privacy and data protection from police use of automated facial recognition technologies on body-worn videos).

9 Patrick Perrot (Gendarmerie Nationale, Ministry of Interior, Paris, France), "What about AI in Criminal Intelligence? From Predictive Policing to AI Perspectives" (2017) 16 *European Police Science and Research Bulletin* 65; Walter L Perry *et al*, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (Rand Corporation, 2013), <https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf> (accessed 15 July 2020); Albert Meijer & Martijn Wessels, "Predictive Policing: Review of Benefits and Drawbacks" (2019) 42(12) *International Journal of Public Administration* 1031.

analysis,[10] and by the courts for administration, dispute resolution and decision-making purposes,[11] the significant consideration for prosecutors and the courts will always be the trustworthiness and reliability of the evidence and its systems.

4 Lawyers also use AI systems for contract analysis, document review and electronic discovery or disclosure, for legal research and analysis, and in practice management applications, including electronic billing and programs for drafting documents and for redaction. Through legal analytics, lawyers are also using AI systems to predict judicial decisions with a view to maximising their success rates before particular courts on specific matters.[12] However, aside from the matter of electronic discovery, a distinction can be made between the use of information, including information generated by AI systems, to prove matters of dispute in legal proceedings, and the use of such information to manage and administer various litigation, investigation and legal processes. As the latter does not involve issues of evidence, no further reference to such use cases will be made in this article.

## III. Machine learning

5 The bulk of the AI systems today are built on machine learning ("ML") algorithms. Rather than following pre-programmed rules, to carry out complex processes, ML works by "allow[ing] systems to learn

10 Owen Bowcott & Hannah Devlin, "Police Trial AI Software to Help Process Mobile Phone Evidence" *The Guardian* (27 May 2018).

11 Adam Harkens, "Fairness in Algorithmic Decision-Making: Trade-Offs, Policy Choices, and Procedural Protections" (2019) 1(1) *Amicus Curiae* 84; "Algorithm Use in the Criminal Justice System Report" *The Law Society of England and Wales* (4 June 2019) <https://www.lawsociety.org.uk/en/topics/research/algorithm-use-in-the-criminal-justice-system-report> (accessed 15 July 2020); People's Republic of China Supreme People's Court, *Chinese Courts and Internet Judiciary White Paper* (中国法院的互联网司法), <https://file.chinacourt.org/f.php?id=43639&class=file> (accessed 15 July 2020). See generally also Zhang Ni, "Computational Jurisprudence" (2021) 33 SAcLJ 355 and Shaun Lim, "Judicial Decision-Making and Explainable Artificial Intelligence: A Reckoning from First Principles" (2021) 33 SAcLJ 280.

12 Lex Machina, "Using Legal Analytics for Early Case Assessment" (April 2018) <https://lexmachina.com/wp-content/uploads/2018/04/Early-Case-Assessment-UseCase.pdf?utm_source=website&utm_medium=datasheets&utm_campaign=resources> (accessed 15 July 2020); Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiuc-Pietro & Vasileios Lampos, "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective" *PeerJ Computer Science* (24 October 2016) <https://peerj.com/articles/cs-93/> (accessed 15 July 2020), criticised in Frank Pasquale & Glyn Cashwell, "Prediction, Persuasion, and the Jurisprudence of Behaviourism" (2018) 68 U Toronto LJ 63.

directly from examples, data, and experience".[13] There are three main permutations of ML:[14]

(a) **supervised learning**: where the AI system is trained with data that has been labelled, learns how it is structured and uses this to predict categories of new data;

(b) **unsupervised learning**: where the AI system aims to detect characteristics of similarities between data points; and

(c) **reinforcement learning**: where the AI system focuses on learning about the rules of the environment and the consequences of its action from experience by interacting with its environment, and tries to develop strategies for optimising a reward (the reward function) that it is given.

6 Although ML has enabled many impressive advancements and powered many of the innovative uses of AI systems today, it can give results that are not expected or are incorrect.[15] In addition, ML is subject to a number of additional limitations:

(a) Many AI systems, especially supervised ML systems, rely on large amounts of training data given a label by a human being.

(b) ML will learn any biases that are contained in the training data, so (for example) an ML system for determining whether a prisoner should be released by the parole board will exhibit racial bias if it has been trained on data that contains such

13 The Royal Society, *Machine Learning: The Power and Promise of Computers That Learn by Example* (April 2017) at p 19.

14 The Royal Society, *Machine Learning: The Power and Promise of Computers That Learn by Example* (April 2017) at p 20; there is also a difference between offline and online learning.

15 There are multiple reasons for incorrect or unexpected results, for a broad discussion, see Will Knight, "AI's Language Problem" *MIT Technology Review* (9 August 2016) <https://www.technologyreview.com/s/602094/ais-language-problem/> (accessed 15 July 2020).

bias.[16] And correlations discovered through ML do not equate to causality.[17]

(c) Datasets will invariably contain hidden biases, as would the choice and use of ML algorithms.[18] This is because the development of datasets and algorithms will involve decisions by humans, who, apart from their own qualifications (or lack thereof) and inherent biases, will have to consider compromises and trade-offs.[19]

(d) There are many constraints on the real world that we know from natural laws (such as physics and mathematics) or logic. It is not straightforward to include these constraints with ML methods.[20]

(e) When our expertise fails, humans fall back on "common sense". But current ML systems do not define or encode this

16 *State v Loomis* 881 NW 2d 749 (Wis, 2016), cert denied, 137 S Ct 2290 (2017); Susan Nevelow Mart, "The Algorithm As a Human Artifact: Implications for Legal [Re]Search" (2017) 109 Law Libr J 387; Anupam Chander, "The Racist Algorithm?" (2017) 115 Mich L Rev 1023; Molly Griffard, "A Bias-Free Predictive Policing Tool?: An Evaluation of the NYPD's Patternizr" (2019) 47 Fordham Urb LJ 43; Aylin Caliskan, Joanna J Bryson & Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases" (2017) 356(6334) *Science* 183; Jieyu Zhao *et al*, "Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-level Constraints" (Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing) (Association for Computational Linguistics, 2017) <https://www.aclweb.org/anthology/D17-1323.pdf> (accessed 15 July 2020).

17 And the correlation can be spurious, for which see Frank Pasquale & Glyn Cashwell, "Prediction, Persuasion, and the Jurisprudence of Behaviourism" (2018) 68 U Toronto LJ 63 at 75.

18 See Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing Group, 2017); "Book Reports" (2017) 14 *Digital Evidence and Electronic Signature Law Review* 95; Stella Lowry & Gordon Macpherson, "A Blot on the Profession" *British Medical Journal* (5 March 1988) at p 657 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2545288/> (accessed 15 July 2020).

19 David Lehr & Paul Ohm, "Playing with the Data: What Legal Scholars Should Learn about Machine Learning" (2017) 51 UC Davis L Rev 653 at 655. See also Solon Barocas & Andrew D Selbst, "Big Data's Disparate Impact" (2016) 104 Cal L Rev 671, which is highly recommended by Lehr & Ohm at 665 and Deven R Desai & Joshua S Kroll, "Trust But Verify: A Guide to Algorithms and the Law" (2017) 31 Harv JL Tech 1 at 23–35.

20 See, *eg*, Jorge Villagra *et al*, "Smooth Path and Speed Planning for an Automated Public Transport Vehicle" (2012) 60(2) *Robotics and Autonomous Systems* 252; and Yinhao Zhu *et al*, "Physics-Constrained Deep Learning for High-dimensional Surrogate Modeling and Uncertainty Quantification without Labeled Data" (2019) <https://arxiv.org/abs/1901.06314> (accessed 15 July 2020).

behaviour.[21] This means that when they fail, they may fail in a serious or brittle manner. In particular, an ML system may be unstable when presented with novel combinations of data, so even if it has been trained on past decisions that have been separately verified by experts, that may not be enough to justify high confidence in a subsequent decision.[22]

(f) Humans are good at transferring ideas from one problem domain to another. This remains challenging for computers even with the latest "transfer learning" ML techniques.[23]

(g) Related to this is the challenge of interpretability – the need to represent knowledge encoded in the learning system in a form that is comprehensible by humans. For instance, it is not easy for even programmers and analysts who train neural networks[24] that are typically used in deep learning ML algorithms to identify or explain the factors or weights that have been used by the networks to arrive at a decision in a particular case.[25]

7 All these special characteristics of AI translate into real and substantial issues regarding the admission of electronic evidence, either in the form of real evidence or records produced by AI systems. It challenges presumptions in evidence about the reliability of automated systems, questions the characterisation of records from AI systems as real evidence or as hearsay, deepens the analysis of such evidence on grounds of authenticity and even goes to the issue of whether such evidence can be the proper subject of legal disclosure or discovery. It is to an examination of these issues that this article now turns.

21 See also paras 65–67 below for a discussion about transparency and regulatory oversight for possible solutions.

22 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014) at p 14.

23 Wikipedia, "Transfer Learning" <https://en.wikipedia.org/wiki/Transfer_learning> (accessed 15 July 2020).

24 For an indication of how much humans are involved in designing and teaching a system, see Riccardo Miotto *et al*, "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records" *Scientific Reports* (17 May 2016) <https://www.nature.com/articles/srep26094> (accessed 15 July 2020). Deep learning techniques depend on neural networks. See Wikipedia, "Deep Learning" <https://en.wikipedia.org/wiki/Deep_learning> (accessed 15 July 2020).

25 E-mail communication, Stephen Mason with Professor Martyn Thomas CBE; see also Gary Marcus, "Deep Learning: A Critical Appraisal" (2017) <https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf> (accessed 15 July 2020).

## IV. Presumption of reliability

8 At common law, a rule of significant concern that governs the admissibility of electronic evidence is the presumption that computer systems are "reliable". In England and Wales, this presumption states that: "In the absence of evidence to the contrary, the courts will presume that mechanical instruments were in order at the material time."[26]

9 While Commonwealth jurisprudence has gradually abandoned the requirement that computer systems are "reliable" as a precondition for the admissibility of electronic evidence,[27] and by corollary, evidence produced by AI systems, the requirement that computer systems be shown to be "reliable" still underpins many exclusionary rules of evidence such as the best evidence rule,[28] the business records exception to the hearsay rule,[29] and the authentication evidence rule.[30] Who has the burden of proving (or disproving) the reliability of the computer system? In the context of AI systems, this issue is especially pertinent because AI systems in general, and certain ML algorithms in particular, can operate in ways that are opaque and not obvious, even to their programmers and operators.

10 In practice, in considering the "reliability" of computers, many judges have not always taken relevant expert advice, or consulted the technical literature on the topic of the "reliability" of computers in reaching their conclusions on this issue.[31] These judges conclude that because these systems do what was expected of them, notwithstanding the opponent's challenge, they are satisfied that these systems are "reliable".[32] In so doing, the Bench has erroneously elevated the presumption into

---

26 United Kingdom, The Law Commission, *Evidence in Criminal Proceedings: Hearsay and Related Topics* (CP No 138, 1997) at para 13.13. This presumption has been approved by the Divisional Court in *Castle v Cross* [1984] 1 WLR 1372 at 1377B, *per* Stephen Brown LJ. There is also a statutory presumption, for which see *Electronic Evidence* (4th Ed) at paras 6.186-6.191. This presumption is often also expressed as the Latin maxim *praesummuntur omnia rite esse acta*.

27 See, *eg*, the repeal of ss 35 and 36 of the Singapore Evidence Act (Cap 97, 1997 Rev Ed) and s 69 of the UK Police and Criminal Evidence Act 1984 (c 60), and similar reform with respect to the England and Wales Civil Evidence Act 1995 (c 38).

28 See, *eg*, s 64 Explanation 3 of the Singapore Evidence Act (Cap 97, 1997 Rev Ed).

29 See, *eg*, s 32(*b*) of the Singapore Evidence Act (Cap 97, 1997 Rev Ed); and *Mitfam International Ltd v Motley Resources Pte Ltd* [2014] 1 SLR 1253.

30 See, *eg*, s 9 of the Evidence Act (Cap 97, 1997 Rev Ed) (facts necessary to explain or introduce relevant facts).

31 For a detailed analysis of each of the mechanisms listed, together with relevant case law, see *Electronic Evidence* (4th Ed) ch 6; see also Bryan H Choi, "Crashworthy Code" (2019) 94 Wash L Rev 39.

32 See, *eg*, the conclusion by Walsh J in *Her Majesty the Queen v Dennis James Oland* 2015 NBQB 245.

a legal presumption that reallocates the burden of proof, not on the proponent of the electronic evidence, but on its opponent.[33]

11 This can be usefully illustrated by reviewing the prosecutions conducted by the UK Post Office Limited ("POL"), against its former sub-postmasters and sub-postmistresses ("SPMs") for fraud, theft and false accounting after POL installed a new computer system in 1999 for managing local post office finances. Known as the Horizon IT system, these prosecutions started when the SPMs reported shortfalls in the accounts of the post offices which they managed as reflected in the Horizon system. When POL prosecuted its SPMs to recover the account discrepancies, the SPM defendants challenged the reliability of the Horizon system. In *R v Seema Misra*,[34] for instance, defence counsel sought information about 2½ years' worth of transaction information for the post office managed by the defendant. The Prosecution challenged this request, on the basis that this was not a sufficiently focused disclosure, that it would be "enormously expensive" to provide this material and that any computer problem "should be visible, at least the symptoms of it to the user of the computer [which the defendant] was using … every single day, probably perhaps hundreds of times a day".[35] Despite the defence counsel making a formal application for disclosure, especially since the defendant's expert had just been shown by the Prosecution's expert an extract of the "known errors log" for the Horizon system (which would have recorded all errors that had taken place on the system),[36] the judge

33 In the Singapore Evidence Act (Cap 97, 1997 Rev Ed), a legal presumption is defined in s 4(2) as follows: "Whenever it is directed by this Act that the court shall presume a fact, it shall regard such fact as proved unless and until it is disproved." This is to be contrasted with s 4(1), which has been interpreted to refer to an evidential presumption, that reads as follows: "Whenever it is provided by this Act that the court may presume a fact, it may either regard such fact as proved unless and until it is disproved, or may call for proof of it."

34 T20090070, in the Crown Court at Guilford, 2010, before his Honour Judge N A Stewart and a jury. For the transcript of the trial, see (2015) 12 *Digital Evidence and Electronic Signature Law Review* 44, Introduction; Documents Supplement. Tim McCormack, "The Post Office Horizon System and Seema Misra" (2016) 13 *Digital Evidence and Electronic Signature Law Review* 133; "'Remote Chance' Horizon Scandal" *Private Eye* (No 1438, 24 February–9 March 2017) at p 37 (this article quotes from the transcript of the trial published in the *Digital Evidence and Electronic Signature Law Review*).

35 Day 1, Monday 11 October 2010, at 5B–5E ((2015) 12 *Digital Evidence and Electronic Signature Law Review*, Document Supplement). This observation appears to parallel what is known as the "Tapper Condition", described in Colin Tapper, "Discovery in Modern Times: A Voyage around the Common Law World" (1991) 67 Chi-Kent L Rev 217 at 248.

36 Day 1, Monday 11 October 2010, at 16 ((2015) 12 *Digital Evidence and Electronic Signature Law Review*, Document Supplement). See also Tim McCormack, "The Trial of Seema Misra" *ProblemswithPOL* (19 January 2020) <https://problemswithpol.wordpress.com/2020/01/> (accessed 15 July 2020).

denied the application, holding that the Defence had ample material to test the integrity of the system.

12 The defendant Misra was subsequently found guilty by a jury and convicted, as were many other ex-SPMs.[37] The dissatisfied SPMs formed an alliance[38] and took civil proceedings against POL. Despite POL's contention that "the Horizon system is and was very robust (being more robust than most comparable systems)",[39] after a lengthy trial in 2019 undertaken with expert forensic analysis of the software,[40] Frazer J in *Bates v The Post Office Ltd (No 6: Horizon Issues) (Rev 1)*[41] unequivocally rejected POL's contention and concluded that there were indeed "a large number of bugs, errors and defects … far larger [in] number than were admitted by the Post Office at the beginning of the [trial], and a far larger number than ought to have been present in the system if [the Horizon system] were to be considered sufficiently robust such that they were extremely unlikely to be the cause of [account] shortfalls in branches".[42] On the basis of this finding, POL acknowledged that it had "got things wrong in our dealings with a number of these postmasters" and in 2020, the CCRC referred 47 Post Office cases on the abuse of process to the Court of Appeal.[43]

### *A. Evidential presumption* versus *legal presumption*

13 The authors are therefore of the view that the presumption of reliability should operate only to place an evidential burden on the party opposing the presumption. As a presumption of evidential convenience, that is prompted only when its premises are met. In its statutory form, the Singapore Evidence Act[44] requires that there be appropriate scientific evidence that the device or process in question "is one that, or is of a kind

37 See T20090070, in the Crown Court at Guilford, 2010, before his Honour Judge N A Stewart and a jury. For the transcript of the trial, see (2015) 12 *Digital Evidence and Electronic Signature Law Review*, Documents Supplement, sentencing: transcript of proceedings 11 November 2010.

38 They formed the Justice for Subpostmasters Alliance: <https://www.jfsa.org.uk/>.

39 *Bates v Post Office Ltd (No 6: Horizon Issues) (Rev 1)* [2019] EWHC 3408 (QB) at [440].

40 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) and *Bates v the Post Office Ltd (No 6: Horizon Issues) (Rev 1)* [2019] EWHC 3408 (QB).

41 [2019] EWHC 3408 (QB), see Appendix 2.

42 *Bates v Post Office Ltd (No 6: Horizon Issues) (Rev 1)* [2019] EWHC 3408 (QB) at [434].

43 See Criminal Cases Review Commission, "The CCRC Refers Eight More Post Office Cases for Appeal – Bringing Total to 47 So Far", press release (3 June 2020) <https://ccrc.gov.uk/the-ccrc-refers-eight-more-post-office-cases-for-appeal-bringing-total-to-47-so-far/> (accessed 29 Dec 2020).

44 Cap 97, 1997 Rev Ed.

that, if properly used, ordinarily produces or accurately communicates an electronic record" before it is presumed that the electronic record in question is produced or accurately communicated by that device or process.[45] Although this is a low threshold requirement, it is not non-existent: when the Prosecution claims that a system is "robust", or that a system is of "high quality", it should be supported by facts and not by mere assertion.[46] The Evidence Act does not prescribe how this requirement can be met, but evidence of the correct, accurate and proper operation of the system can *ex facie* come from users who have used the system[47] and developers or experts who opine that the users' methods are apt to produce the correct result.[48] Nor is this requirement only to be met in relation to evidence about the actual system itself: evidence about a related class or similar kind of device or process may suffice.

14 If the presumption is so established, the consequence is that there is an evidential presumption that the system in question is reliable, and nothing more. The presumption does not overturn the basic rule of evidence that the burden of proof remains with the proponent of electronic evidence to prove the evidence.[49] The proponent of the evidence generated by the system still has to discharge the legal burden in relation to the reliability of the machine, and likewise, the authenticity or integrity and the trustworthiness of the evidence.[50] In this regard, it is critical to note the equivocal language of the presumption where it is encoded in statutory form. In the Singapore Evidence Act, the presumption may be easily refuted where there is "[adduced] evidence sufficient to raise doubt

45 Evidence Act (Cap 97, 1997 Rev Ed) s 116A(1) Illustration. See also *Electronic Evidence* (4th Ed) at para 6.27.

46 For which see Peter Bernard Ladkin, "Robustness of Software" (2020) 17 *Digital Evidence and Electronic Signature Law Review* 15.

47 The person using the system must be sufficiently knowledgeable about how it works. The evidence from England and Wales is that lay witnesses are permitted to indicate that a computer system is reliable. But a whimsical answer such as, "Touch wood, no. I have never known [the computer to] break down since we have had it" hardly qualifies as adequate evidence of the reliability of the system. See *Electronic Evidence* (4th Ed) at para 10.12.

48 *Mehesz v Redman (No 2)* (1980) 26 SASR 244 at 248, *per* King CJ. For an example as to how the s 116A(1) presumption is established, see *Telemedia Pacific Group Ltd v Credit Agricole (Suisse) SA* [2015] 1 SLR 338 at [242]–[260]. Criticism, however, can be made, in that the person who testified as a user could only testify as to what the records generation machine had done – he was in no position to testify as to what the machine failed to do, which was the nub of the defendant's case – that he had not received the records purportedly generated by the records generation machine.

49 For which see the erudite article by Nigel Bridge, "Presumptions and Burdens" (1949) 12 Mod L Rev 273.

50 Evidence (Amendment) Bill 2012 (Bill 2 of 2012) Explanatory Statement, cl 13. See also *Electronic Evidence* (4th Ed) at paras 6.192 and 6.196.

about the presumption".[51] The Australian Commonwealth Evidence Act uses a similar formulation in relation to its reliability presumption.[52] That the UK Law Commission in 1997 intended for the presumption to operate as an evidential presumption is clear as follows:[53]

> Even where the presumption applies, it ceases to have any effect once evidence of malfunction has been adduced. The question is, what sort of evidence must the defence adduce, and how realistic is it to suppose that the defence will be able to adduce it without any knowledge of the working of the machine? On the one hand the concept of the *evidential burden* is a flexible one: a party cannot be required to produce more by way of evidence than one in his or her position could be expected to produce. It could therefore take very little for the presumption to be rebutted, if the party against whom the evidence was adduced could not be expected to produce more. [emphasis added]

15 Nor should the absence of evidence of any computer failure suggest system reliability. After all, "the fact that a class of failures has not happened before is not a reason for assuming that it cannot occur".[54] This "absence of evidence of failure" may be because such failures are never recorded in the first place. It is precisely to avoid such types of inferences from false negatives that licensing regulations for autonomous vehicles have required that such systems keep records of sensor and other telemetric data to enable the circumstances surrounding vehicle accidents to be reconstructed.[55] The absence of any kind of error logging feature or even a failure to produce such a log for review for a sophisticated AI system should itself be a red flag as to the system design error when considering the overall reliability of the system.[56]

---

51 Evidence Act (Cap 97, 1997 Rev Ed) s 116A(1). In contrast, s 116A(2) uses the language "evidence to the contrary is adduced", which enables the operation of the presumption in s 116A(2) as a "legal presumption" as the legal burden is on the opponent of the presumption to adduce evidence to refute the presumption. See also Daniel Seng & Sriram Chakravarthi, *Computer Output As Evidence* (Singapore Academy of Law, Consultation Paper, 2003) at p 140.

52 Commonwealth Evidence Act 1995 (Cth) ss 146 and 147. In contrast, the equivalent provision in the Canadian Evidence Act (RSC, 1985, c C-5), s 21.3(a), appears to be formulated in an equivocal way such that it can be read as either a legal presumption or an evidential presumption.

53 United Kingdom, The Law Commission, *Evidence in Criminal Proceedings: Hearsay and Related Topics* (CP No 138, 1997) at para 13.18.

54 United Kingdom, The Law Commission, *Evidence in Criminal Proceedings: Hearsay and Related Topics* (CP No 138, 1997) at para 13.18.

55 Road Traffic (Autonomous Motor Vehicles) Rules 2017 (S 464/2017) r 17(4) (requiring data recorder on autonomous vehicle to be kept for at least three years from date of recording, and to record, among others, the vehicle speed, location, status, operator override history, sensor data and camera or video footage).

56 See, *eg*, Paul Marshall, "The Harm That Judges Do – Misunderstanding Computer Evidence: Mr Castleton's Story: 'An Affront to the Public Conscience'" (2020) 17 *Digital Evidence and Electronic Signature Review* 25 at 32.

16 Therefore, the evidential presumption can be refuted, for instance, with any evidence of any previous computer failure[57] as this would raise doubt about the presumption of reliability. The proponent may object to this as not being material to the type of evidence in question, or that the failure relates to another category of evidence. In this regard, it behoves the judge to be careful when evaluating such evidence and the underlying objective to be supported by the presumption.

17 The central problem (implicit and explicit) for overstating the presumption is the failure to understand that software systems can produce subtle mistakes that are not obvious. Contrary to what has been opined,[58] it is not generally true that "most computer error is either immediately detectable or results from error in the data entered into the machine".[59] Most of such blatant errors will be caught in evaluation, beta testing and unit tests, and will not make their way into production. And very few systems have computer codes that are provably reliable and correct (this is only with the use of special programming tools such as functional programming[60] coupled with formal mathematical proofs).[61] This also flies in the face of industry knowledge that modern software is complex: it depends on a myriad of other software systems and components; it may include components developed by third parties, whose software components may in turn depend on components developed by third parties and so on; it may be linked to via networks or the Internet or interact with some other systems (such as mobile or embedded devices), which may behave in unpredictable ways; if so, it will have to deal with other systems which may operate their own separate and independent platforms, environments and systems. In fact, Frazer J's detailed summary of the errors in the Horizon system strongly suggests that many of the software errors stem from the complex software interaction between the front-end, middleware and backend components of the post office system architecture and the failure to account for communication, handshaking, synchronisation, device and user-interface errors, leading to corruption

57 Peter Bernard Ladkin *et al*, "The Law Commission Presumption Concerning the Dependability of Computer Evidence" (2020) 17 *Digital Evidence and Electronic Signature Law Review* 1.

58 Colin Tapper, "Discovery in Modern Times: A Voyage around the Common Law World" (1991) 67 Chi-Kent L Rev 217 at 248.

59 Colin Tapper, "Discovery in Modern Times: A Voyage around the Common Law World" (1991) 67 Chi-Kent L Rev 217 at 248. For a critique of this assertion, see Peter Bernard Ladkin *et al*, "The Law Commission Presumption Concerning the Dependability of Computer Evidence" (2020) 17 *Digital Evidence and Electronic Signature Law Review* 1.

60 Wikipedia, "Functional Programming" <https://en.wikipedia.org/wiki/Functional_programming> (accessed 15 July 2020).

61 Wikipedia, "Formal Verification" <https://en.wikipedia.org/wiki/Formal_verification> (accessed 15 July 2020).

of the accounting records and the overall integrity of the accounts.[62] It is therefore of no surprise that almost every non-customised software licence includes an explicit disclaimer of warranty of merchantability and fitness for purpose and seeks to exempt or limit any liability for any direct or indirect damage caused by any malfunctioning of software.

### *B. What about AI systems?*

18 The complexity of AI systems takes this issue of proving (or disproving) the reliability of AI systems to the next level. The knowledge and processes encoded by way of ML in AI systems are not in the form of procedural steps but are in the weights of the features and the architecture of the different components of the system. This means that traditional methods of testing software for mistakes such as unit tests and the unit test platforms themselves[63] are unlikely to be very helpful. This is especially the case when AI systems have errors embedded in them that will activate only in exceptional circumstances. AI systems operating in open rather than closed environments can be exposed to an infinite variety of such exceptional circumstances or "corner cases".[64] Research into the use of adversarial attacks indeed represent an attempt at exploring the weaknesses of AI systems when dealing with corner case situations.[65]

19 An example of a well-investigated instance of an "unreliable" AI system operating in such an environment is the case of the Uber autonomous vehicle that killed a pedestrian. The case attracted much attention because it was the first recorded case of a pedestrian fatality involving an autonomous vehicle. Investigations by the US National Transportation Safety Board based on the recorded telemetry and sensor data showed that the main problem was the AI system for environmental perception, which had difficulty correctly classifying the victim as a pedestrian. She was initially classified as an unknown object, then as a vehicle, and finally as a bicycle, each of which had a different predicted

62 See, *eg*, *Bates v Post Office Ltd (No 6: Horizon Issues) (Rev 1)* [2019] EWHC 3408 (QB) at [440].

63 The authors thank Andrew Sheldon MSc and Professor Martyn Thomas CBE for these points.

64 A corner case is a problem or situation that occurs outside of normal operating parameters. See Wikipedia, "Corner Case" <https://en.wikipedia.org/wiki/Corner_case> (accessed 15 July 2020).

65 Wikipedia, "Adversarial Machine Learning" <https://en.wikipedia.org/wiki/Adversarial_machine_learning> (accessed 15 July 2020).

path according to the collision detection logic.[66] When emergency braking was determined to be necessary and required, only 1.3 seconds prior to impact, the vehicular control system did not make an emergency stop on its own accord and did not alert the operator, because this was disabled by Uber to reduce the potential for erratic vehicle behaviour.[67] Other human factors such as the distracted safety driver and the jaywalking pedestrian at night also played a part in the accident.[68]

20 Proof of the reliability of AI systems will therefore involve a combination of many mechanisms. The testing will have to encompass each and every subcomponent of the AI system – many of which are ML systems in themselves. The testing environment for certifying AI systems will have to test for basic correctness of such systems that integrate all the ML subsystems based on known environments or parameters that have correct or known results. For instance, autonomous vehicles will have to demonstrate that they can comply with traffic signs and traffic rules. This test will exercise not only the environmental perception systems of autonomous vehicles, but also the navigation systems and overall vehicular control systems.

21 The testing environment will have to simulate various real-time conditions, ranging from different physical environmental conditions, different road conditions and different situations, *eg*, emergencies, police interventions. The robustness of an AI system will largely depend on how many of these variations in environments it has been tested in and for which it has been validated. Since there is no limit to the type of exceptions which AI systems can be exposed to, and since AI systems cannot be formally proved to be accurate, answering this question will involve reviewing any evidence of any "errors" which AI systems cannot handle, including the number of such errors, their frequencies and the nature of these errors.[69] Based on this information, a court can consider whether the evidence sought to be admitted could be materially changed

66 National Transportation Safety Board, *Preliminary Report, Highway, HWY18MH010* (24 May 2018) at p 3 <https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf> (accessed 15 July 2020).

67 National Transportation Safety Board, *Preliminary Report, Highway, HWY18MH010* (24 May 2018) at p 3 <https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf> (accessed 15 July 2020).

68 Wikipedia, "Death of Elaine Herzberg" <https://en.wikipedia.org/wiki/Death_of_Elaine_Herzberg> (accessed 15 July 2020).

69 Peter Bernard Ladkin *et al*, "The Law Commission Presumption Concerning the Dependability of Computer Evidence" (2020) 17 *Digital Evidence and Electronic Signature Law Review* 1 at 7.

by such computer errors.[70] In this regard, the court could be guided in its assessment as to whether such systems are reliable by whether such systems comply with any relevant standards, and if so, the level of conformance to such standards.[71]

## V. Hearsay

22 In the leading case of *Subramaniam v Public Prosecutor*[72] ("*Subramaniam*"), the Privy Council on appeal from Malaysia defined the hearsay rule as follows:

> Evidence of a statement made to a witness by a person who is not himself called as a witness may or may not be hearsay. It is hearsay and inadmissible when the object of the evidence is to establish the truth of what is contained in the statement. It is not hearsay and is admissible when it is proposed to establish by the evidence, not the truth of the statement, but the fact that it was made.

23 Following *Subramaniam*, in Singapore law, any assertion – express or implied or intended or unintended – is *prima facie* treated as hearsay if it is tendered for the content of its assertion. Whether it can otherwise be admitted not as evidence of its assertion but as something else (and therefore not prompt the hearsay rule) will depend on what is proposed to be established by its admission, and thus, on the contexts in which the assertion is made and the facts in issue or relevant facts sought to be established by the "assertion".

24 While the hearsay rule in Singapore has survived as part of the common law of evidence and is operating within the strictures of the Evidence Act and the Criminal Procedure Code,[73] in contrast, the hearsay rule has seen general retreat in many common law jurisdictions. With a view to enabling the admission of hearsay that is relevant and probative, statutory reforms have either limited the scope of the rule to statements of express assertions and conduct within which an intention to assert could be established (thereby excluding implied or unintended

70 Peter Bernard Ladkin *et al*, "The Law Commission Presumption Concerning the Dependability of Computer Evidence" (2020) 17 *Digital Evidence and Electronic Signature Law Review* 1 at 8.

71 Peter Bernard Ladkin *et al*, "The Law Commission Presumption Concerning the Dependability of Computer Evidence" (2020) 17 *Digital Evidence and Electronic Signature Law Review* 1 at 8.

72 [1956] 1 WLR 965.

73 Cap 68, 2012 Rev Ed. See, *eg*, s 268 of the Criminal Procedure Code (providing that hearsay evidence in criminal proceedings is admissible as evidence of any fact stated therein to the extent that it is so admissible by the Code, the Evidence Act and any other written law).

assertions),[74] or provided statutory mechanisms to admit the hearsay statement by notice[75] or in the interests of justice.[76] This is usually in addition to the existing exemptions to the hearsay rule such as admissions and confessions,[77] and business records.[78]

25 All these developments suggest that the hearsay rule has little or no relevance to electronic evidence in general, let alone to AI-generated or AI-produced evidence in particular. Such an analysis would be substantially wrong.

26 While the purpose or rationale of the hearsay rule and its manifold exceptions have been the subject of various academic debate and critique, the underlying premise of the rule – that any evidence must be "testable and reliable" – has remained a universal tenet for both the common law and civil law systems. As explained by Stein, this affords a party affected by the evidence protection against any adverse inferences drawn from such evidence, unless the evidence has been exposed to and survived "maximal individualized testing". This includes giving the adversely-affected party "every practical possibility of testing the applicability of the inference in question to [that party's] case".[79]

27 Today, almost everybody uses electronic evidence and through our interactions, we consciously or unconsciously make assessments to decide whether to rely on it. We constantly reject information as phishing e-mails, mobile telephone scams, false news and clickbait. In many instances, these records, messages and posts are generated and disseminated automatically through the use of automated and AI-driven systems. As such information finds its way as evidence in proceedings, this raises questions as to its correct treatment in a court of law. For instance, in *Bucknor v R*,[80] where the defendant was charged with committing murder as member of a gang, the trial judge admitted as evidence photographs from a social networking site that portrayed the defendant

74 See, *eg*, s 115(3) of the UK Criminal Justice Act 2003 (c 44); s 4 of the New Zealand Evidence Act 2006; and s 59(1) of the Australian Commonwealth Evidence Act 1995.

75 See, *eg*, s 2 of the England and Wales Civil Evidence Act 1995 (c 38); s 22 of the New Zealand Evidence Act 2006; and s 67 of the Australian Commonwealth Evidence Act 1995.

76 See, *eg*, s 114(1)(d) of the UK Criminal Justice Act 2003 (c 44).

77 See, *eg*, ss 17–23 of the Evidence Act (Cap 97, 1997 Rev Ed).

78 See, *eg*, s 32(1)(*b*) of the Evidence Act (Cap 97, 1997 Rev Ed).

79 Alex Stein, "The Refoundation of Evidence Law" (1996) 9 *Journal of Law and Jurisprudence* 279 at 326–327 and 331. See also George L Paul, "Systems of Evidence in the Age of Complexity" (2014) 12 *Ave Maria Law Review* 173; and Steve W Teppler, "Testable Reliability: A Modernized Approach to ESI Admissibility" (2014) 12 *Ave Maria Law Review* 213.

80 [2010] EWCA Crim 1152.

as a member of an organised criminal gang, and a YouTube video that portrayed the gang as violent. The jury were directed to consider this as "background information" if they conclude that such content originated from the defendant, who denied all knowledge of it. The English Court of Appeal quashed the defendant's conviction, noting that the site contents and video were hearsay, especially since the maker of such content was not identified. Even if the content, assuming it to be true, had probative value, the court held that a failure to consider how reliable the maker of the contents was and how many levels of hearsay were involved meant that no consideration was given to the reliability of such content.

28 Such a sentiment expressed by the court is especially prescient in today's information era, and highlights the important filtering role served by the hearsay rule, made especially pertinent with the flood of phishing e-mails, mobile telephone scams and false news on social media.

### *A. Three categories of electronic devices*

29 AI systems can produce many different types of evidence. Voice recognition systems can be automatically activated and store conversation snippets it has recorded.[81] Image recognition systems such as those found on traffic enforcement cameras can capture photographs of vehicles[82] and generate traffic violation tickets when linked to number plate recognition systems.[83] Fraud detection systems can monitor credit card transactions and identify anomalous transactions for further investigation. These examples are just illustrative of the wide range of information generated by AI systems that may be admitted in court as relevant and material evidence. But is such evidence allowed under the hearsay rule?

30 In 1997, one of the authors proposed the following classification for testing electronic evidence under the hearsay rule:[84]

81 See, *eg*, Zack Whittaker, "Judge Orders Amazon to Turn over Echo Recordings in Double Murder Case" *TechCrunch* (15 November 2018) <https://techcrunch.com/2018/11/14/amazon-echo-recordings-judge-murder-case/> (accessed 15 July 2020); and Jon Fingas, "Florida Police Obtain Alexa Recordings in Murder Investigation" *engadget* (2 November 2019) <https://www.engadget.com/2019-11-02-florida-police-obtain-alexa-recordings-in-murder-case.html> (accessed 15 July 2020).

82 See, *eg*, *Jackson v R* [2011] EWCA Crim 1870; *Attorney General's Reference No 114–115 of 2009* [2010] EWCA Crim 1459, Annexure A; *Najib v R* [2013] EWCA Crim 86 at [9]; *Khan v R* [2013] EWCA Crim 2230; and *Welsh v R* [2014] EWCA Crim 1027 at [2].

83 See David Pitt, "Iowa Court: Automated Speeding Tickets Not a Public Record" *AP News* (4 January 2020).

84 Daniel Seng, "Computer Output As Evidence" [1997] Sing JLS 130 at 173.

> Computers which are used as data processing devices can be classified into the following categories:
>
> [category 1] devices which accept human-supplied input and produce output,
>
> [category 2] self-contained data processing devices which obtain input or take recordings from the environment without human intervention, and
>
> [category 3 devices which are] a hybrid of the two.

31 Since it was first proposed more than 20 years ago, the analysis of electronic evidence using this classification has been accepted in academic writings and in legal texts.[85] With advancements of ML and increasing acceptance of AI systems, it is timely to review this classification in relation to AI-generated evidence.

### *B. The treatment of AI evidence*

32 In the context of AI systems, Category 1 devices are used primarily to store and record content that is written and spoken by one or more persons. They operate primarily to receive human input and record that same input for subsequent retrieval. When the information is retrieved and is used for the content of the human input, that input is used testimonially. Any ML-driven processing or ingestion is limited to activation of the device for recording, and transcribing or indexing the recording to facilitate its retrieval. Evidence produced by Category 1 evidence therefore is generally hearsay. Here, the AI system and the data and software code that go into it can be separated from the content input by the maker of the statement.[86] The focus thus of any use of such evidence is directed at an investigation into the identity of the person who made the assertion and the reliability of such an assertion. In such a case, the same conclusion as that reached in *Aw Kew Lim v Public Prosecutor*,[87] that the computerised storage of company registration records, regarding the identities and addresses of the defendants, did not materially alter the content of the assertions (as to identities and addresses), would require such evidence be treated as hearsay.

33 Evidence produced by Category 2 devices is, simply put, evidence that is substantially the product of automation and is not used

---

85 See, *eg*, Stephen W Teppler, "Testable Reliability: A Modernized Approach to ESI Admissibility" (2014) 12 *Ave Maria Law Review* 213 at 255; *Archbold: Criminal Pleading, Evidence and Practice 2016* (James Richardson ed) (Sweet & Maxwell, 64th Ed, 2016) at paras 9-11–9-14; and *Electronic Evidence* (4th Ed) at paras 5.19 *ff*.

86 *Electronic Evidence* (4th Ed) at paras 5.21–5.24 for a discussion about whether a document typed into a computer and subsequently printed out is computer output.

87 [1987] SLR(R) 443.

testimonially. For instance, many criminal prosecutions in England have succeeded through the admission of automatic number plate recognition ("ANPR") evidence, to show vehicular location, movement and time.[88] ANPR works by having specially adopted closed-circuit television cameras that are fitted with infra-red sensors that can capture the number plates of vehicles, even at night. The images are then fed into ML systems that "read" the number plates and that information is sent to the Police National Computer to find a match for the vehicle and its owner.[89] English courts appear relatively sanguine in admitting ANPR evidence, with no noted hearsay challenges raised.[90] Nonetheless, the reason for the absence of challenges is that such evidence is considered real evidence or "evidence produced purely mechanically without human intervention" and is outside the hearsay rule.[91]

34 But while real evidence from these automatic systems does not amount to "assertions" that are caught by the hearsay rule, this does not mean that such evidence is reliable or accurate. Challenges to the reliability and accuracy of such evidence will be by way of authentication, which is addressed in the next section of the article. And the absence of challenges to ANPR evidence in the English courts could be attributed to the fact that for the large part, the defendants or the parties have admitted to the accuracy of such evidence and so no real dispute arises.[92] Even so, when a discrepancy arises in relation to ANPR evidence, as in the case of *Re A (death of a baby)*,[93] there was other evidence to corroborate the drivers' testimony as to their movements and contradict the ANPR evidence. In other words, the independent verifiability of the vehicular

88 See, *eg*, *Jackson v R* [2011] EWCA Crim 1870; *Attorney General's Reference No 114–115 of 2009* [2010] EWCA Crim 1459, Annexure A; *Najib v R* [2013] EWCA Crim 86 at [9]; *Khan v R* [2013] EWCA Crim 2230; and *Welsh v R* [2014] EWCA Crim 1027 at [2].

89 See Primo Reg Plates, "Your Guide to Automatic Number Plate Recognition" <https://www.primoregistrations.co.uk/article/view/your-guide-to-automatic-number-plate-recognition> (accessed 15 July 2020).

90 See, *eg*, *R v Doyle* [2017] EWCA Crim 340, where hearsay challenges were raised on appeal in relation to the covert listening devices placed on the defendants' cars, but not in relation to the automatic number plate recognition evidence as part of the evidence of the movement of those cars. Likewise, hearsay challenges were raised but not in relation to the automatic number plate recognition evidence in *R v Brown* [2019] 1 WLR 6721; [2019] EWCA Crim 1143.

91 *Sapporo Maru v Statue of Liberty; The Statue of Liberty* [1968] 1 WLR 739; [1968] 2 All ER 195 (admitting as real evidence radar set recordings of nautical traffic).

92 See, *eg*, *Cavendish Square Holding BV v Talal El Makdessi (Rev 3)* [2016] AC 1172; [2015] UKSC 67 at [125]; *D (a child) (fact-finding appeal)* [2019] EWCA Civ 2302 at [32].

93 [2011] EWHC 2754 (Fam) at [66] and [99].

movements enabled the court to exercise its discretion and choose to draw no conclusions from the ANPR evidence.[94]

35 A large majority of AI evidence, however, will be evidence produced by Category 3 devices. In this category, the device output will comprise a mix of human-supplied input and data processed output which operates without human intervention. As supervised ML systems are trained on human-labelled data so that they can operate autonomously, evidence from ML systems will invariably fall into this category.

36 The line between evidence produced by Category 2 and Category 3 devices can be hard to draw: the difference really is one of degree that represents the relative significance of the level of contribution of human-supplied input and pre-programmed autonomous processes to the eventual output. In *Public Prosecutor v Ang Soon Huat*[95] ("*Ang Soon Huat*"), for instance, the High Pressure Liquid Chromatograph and Gas Chromatography Mass Spectrometer outputs which were adduced to prove the weight of the trafficked drug was admitted as real evidence: by supporting such automated output with the oral testimony of the technicians who calibrated and operated the machines, the automated processes were characterised as recording, processing and calculating the information fed into them without human intervention. The Singapore High Court cited in support the English case of *R v Wood*[96] ("*Wood*"), where evidence of the computer analysis of the chemical tests was held to be real evidence and admissible. As in *Ang Soon Huat*, the Crown in *Wood* secured oral evidence from both the chemists and the programmer of the program used by the chemists to derive their test results.[97] The English Court of Appeal held that the test results were real evidence. It said:[98]

> This computer was rightly described as a tool. It did not contribute its own knowledge [but] merely did a sophisticated calculation which could have been done manually by the chemist and was in fact done by the chemists using the computer programmed by the [programmer] whom the Crown called as a witness. The fact that the efficiency of a device is dependent on more than one person does not make any difference in kind. Virtually every device will involve the persons who made it, the persons who calibrated, programmed or set it up … and the person who uses or observes the device. In each particular case how many of these people it is appropriate to call must depend on the facts of, and the issues raised and the concessions made in that case. [emphasis added]

94 *Re A (death of a baby)* [2011] EWHC 2754 (Fam) at [158].
95 [1990] 2 SLR(R) 246.
96 (1983) 76 Cr App R 23.
97 *R v Wood* (1983) 76 Cr App R 23 at 26 ("[the computer's] programming and its use were both covered by oral evidence").
98 *R v Wood* (1983) 76 Cr App R 23 at 27.

37 Yet there is a limit as to how far this analogy that a sophisticated device such as an AI system is nothing more than a calculator that did no more than perform calculations based on the various persons who set it up and so on can be stretched. In *Ang Soon Huat*, the Prosecution called the technicians to testify. In *Wood*, the Prosecution also included the programmer. But supervised ML-based AI systems function are based on the "knowledge" of not just their operators and programmers, but also the collective knowledge of patterns from training and test data drawn from diverse sources, labelled by other parties and validated by even more diverse parties. For instance, ML systems that recognise handwritten numbers have been trained on the Modified National Institute of Standards and Technology dataset of handwritten digits.[99] Yet as the datasets represent how the authors[100] who provided the samples write numbers, and not how numbers are handwritten, some researchers have examined these ML systems for number recognition and found them to be "brittle".[101] Likewise, experimental results have shown that matching accuracies of both commercial and non-trainable facial recognition algorithms consistently have lower matching accuracies on females, people with black skin and those in the age group of 18 to 30 because they are heavily dependent on the datasets against which they are trained, and these ill-matching groups are poorly represented in the datasets.[102] These are but two examples of the additional complexities associated with investigating AI systems for their hearsay issues and their reliability.

38 Therefore, there is much to be said for treating AI systems as "the witness"[103] in proceedings. Just as a human witness will be subject to an examination as to his or her experience and qualifications, subjecting AI output to the scrutiny of the hearsay rule helps to tease out the embedded human assertions from the results sought to be admitted in evidence – be it the ML code or its data. For instance, this classification may necessitate

99 Patrick J Grother, *NIST Special Database 19 – Handprinted Forms and Characters Database*, (National Institute of Standards and Technology, 16 March 1995) <https://www.nist.gov/system/files/documents/srd/nistsd19.pdf> (accessed 15 July 2020).

100 These are US Census Bureau employees and high-school students. See Yaan LeCun, Corinna Cortes & Christopher J C Burges, *The MNIST Database of Handwritten Digits* <http://yann.lecun.com/exdb/mnist/> (accessed 15 July 2020).

101 Alexander K Seewald, "On the Brittleness of Handwritten Digit Recognition Models" *International Scholarly Research Notices* (30 November 2011) <https://www.hindawi.com/journals/isrn/2012/834127/> (accessed 15 July 2020).

102 Brendan F Klare *et al*, "Face Recognition Performance: Role of Demographic Information" (2012) 7(6) *IEEE Transactions on Information Forensics and Security* 1789.

103 See Stephen Mason, "Software Code As the Witness" in *Electronic Evidence* (4th Ed) ch 5, although see the opinion of Judge Curtis E A Karnow in "The Opinion of Machines" (2017) 19 Colum Sci & Tech L Rev 136.

differences in treatment between AI systems that depend on supervised ML *versus* unsupervised ML, with reinforcement ML as some form of intermediate hybrid between the two.[104] If there is no opportunity for the human assertions to be tested – for instance, if the automatically-produced analysis is to be relied on but the programmer who wrote the software that generated the analysis is not called to testify – the analysis becomes hearsay.[105] Given that the product of AI systems will inevitably be based on a multiplicity of, and interplay between, direct and indirect human assertions, not all of which have been validated, let alone completely assessed for their accuracy and correctness,[106] it will be near impossible to call all contributors of these assertions to give evidence in legal proceedings. Therefore, considering that these models embed various human assertions and even biases, it is more apt to proceed with caution and subject AI evidence to closer scrutiny for the "human input". Of course, this closer scrutiny can be further assisted with a robust approach to authentication of such evidence and to a more effective stance regarding disclosure.

## VI. Authentication

39 There are two qualities to trustworthy evidence: its reliability and its authenticity. The rule of hearsay assesses the reliability of the evidence by determining if the record is capable of representing the facts to which it attests.[107] Authenticity of the evidence on the other hand means demonstrating that the evidence is genuine – that it is what it claims to be,[108] and that its condition is substantially unchanged.[109] It follows that the authenticity of evidence is a condition precedent to its admissibility.[110] This finds statutory form in s 9 of the Singapore Evidence Act, which reads:

104 The human input in each is different. In supervised ML, this will primarily be in the form of the human-labelled data used to train the system. In unsupervised and reinforcement ML, this will largely be in the choice of the parameters used to set up the algorithms and to establish the goals of the "objective function". See Wikipedia, "Loss Function" <https://en.wikipedia.org/wiki/Loss_function> (accessed 15 July 2020).

105 See *Mehesz v Redman* (1979) 21 SASR 569; and *Holt v Auckland City Council* [1980] 2 NZLR 124.

106 See *Electronic Evidence* (4th Ed) at para 4.44.

107 *Electronic Evidence* (4th Ed) at para 7.1.

108 *Electronic Evidence* (4th Ed) at para 7.1.

109 *McCormick on Evidence* (West Publishing Co, 3rd Ed, 1984) at p 686; John Henry Wigmore, *Evidence in Trials at Common Law* (vol 7, §2129) (revised by James H Chadbourn) (Little, Brown, 1978) at p 709.

110 Daniel Seng, "Computer Output as Evidence" [1997] Sing JLS 130 at 161–163.

> **9.** Facts necessary to explain or introduce a fact in issue or relevant fact, or which support or rebut an inference suggested by a fact in issue or relevant fact, or which establish the identity of any thing or person whose identity is relevant, or fix the time or place at which any fact in issue or relevant fact happened or which show the relation of parties by whom any such fact was transacted, are relevant in so far as they are necessary for that purpose.
>
> *Illustrations*
>
> …
>
> (*g*) *A* seeks to adduce evidence against *B* in the form of an electronic record. The method and manner in which the electronic record was (properly or improperly) generated, communicated, received or stored (by *A* or *B*), the reliability of the devices and the circumstances in which the devices were (properly or improperly) used or operated to generate, communicate, receive or store the electronic record, may be relevant facts (if the contents are relevant) as authenticating the electronic record and therefore as explaining or introducing the electronic record, or identifying it as the relevant electronic record to support a finding that the record is, or is not, what its proponent *A* claims.

40 Authentication is also explicitly provided for in s 31.1 of the Canada Evidence Act,[111] which reads:

> Any person seeking to admit an electronic document as evidence has the burden of proving its authenticity by evidence capable of supporting a finding that the electronic document is that which it is purported to be.

41 Authenticity of electronic evidence is always an important issue in its admissibility because of the inherently mutable nature of electronic evidence. At the same time, authenticating electronic evidence raises an intertwined complex of legal and technical questions about, among others, its integrity and security. Authentication can, among others, be the basis for raising issues as to the identity of the electronic record (whether this was the correct record for evidential purposes), its integrity and immutability (whether it was altered, manipulated or damaged between the time it was created and the time it was tendered in evidence), its authorship (who created it or caused it to be created), its recipient (to whom it was targeted), its identity (what systems and processes went into producing it), its chronology, locality and sequence (in what order was it produced and when and where was it produced), its accuracy (were its contents correct and reliable, and whether the system or device was properly used and operated) and its security (who had or could have obtained access to it, and its underlying systems and processes, for legal or for illicit purposes).[112]

111 RSC, 1985, c C-5.

112 Daniel Seng, "Computer Output as Evidence" [1997] Sing JLS 130 at 161–163.

42 Where challenges as to the authenticity of electronic evidence are mounted, many judges have resiled from this evaluation. Instead, they have made erroneous assumptions about the evidence before them that are not warranted, even when the authentication challenge goes to the root of the dispute between the parties. There is an implicit willingness to assume that organisations carry on their business competently,[113] even where the challenge goes to undermining that business competency itself. This may be illustrated by the several Norwegian cases where banks have successfully sued card holders for transactions on their cards, assisted in part by the courts' conclusion that the banks' systems were secure, despite the card holders' claims that these transactions were unauthorised and that they had not been negligent in securing their card PINs. The banks had to recant their claims and apologise to the holders when evidence subsequently showed that there were security breaches with the banks' systems.[114]

43 To enable and facilitate a careful consideration of the authentication issues, a clear procedure should be set up to prescribe how matters regarding the authenticity of electronic evidence should be raised.[115] In criminal proceedings, the Defence should be required to provide advance warning to the trial judge that the authenticity of identified aspects of the evidence will be questioned, and to set out the grounds upon which the challenge is made.[116] If this first hurdle is overcome, it will be for the trial judge to decide whether a trial within a trial is necessary, and if so, to set out the scope and parameters of the hearing, including the standard of proof, for which a ruling is required. In civil proceedings, parties may, at the disclosure or discovery stage, presume the authenticity of the evidence. And it will be for the party challenging the authenticity of identified aspects of the evidence to notify the opposing party and the court in advance of the proceedings.[117]

---

113 Note the unrealistic interpretation of the evidence by Stanley Burnton LJ in *O'Shea v R* [2010] EWCA Crim 2879 at [56]: "It is also surprising in the extreme that if the supposed fraudulent webmaster was able to debit the appellant's credit card account, he did so for such limited amounts and on relatively few occasions. This is, however, a minor point."

114 See, *eg*, *Bernt Petter Jørgensen v DnB NOR Bank ASA* (Journal number 04-016794TVI-TRON, Trondheim District Court, 24 September 2004). For a translation into English, see (2012) 9 *Digital Evidence and Electronic Signature Law Review* 117; also see Maryke Silalahi Nuth, "Unauthorized Use of Bank Cards With Or Without the PIN: A Lost Case for the Customer?" (2012) 9 *Digital Evidence and Electronic Signature Law Review* 95.

115 As suggested in *Electronic Evidence* (4th Ed) at paras 6.168–6.171.

116 *Electronic Evidence* (4th Ed) at paras 6.168–6.171. To a certain extent this might be already happening, for which see Oriola Sallavaci, "Streamlined Reporting of Forensic Evidence in England and Wales: Is It the Way Forward?" (2016) 20(3) E & P 235.

117 *Electronic Evidence* (4th Ed) at paras 6.168–6.171.

## VII. Digitally-manipulated data

44 The problem of digitally-manipulated electronic records has always existed since the advent of the computer. In fact, it is with a view to addressing issues about forged electronic records that digital signature technologies were developed to provide assurances as to the integrity of such records. However, of late, another type of digitally-manipulated electronic records has become of significant concern. Popularly known as "deep fakes", these involve the alteration of existing images or videos in which a person's image or likeness in that image or video is replaced with someone else's likeness.[118] These modifications, made possible by using ML and AI techniques such as autoencoders[119] and generative adversarial networks,[120] have a high potential to deceive because of their success, including altering the very subtle gestures and movements of a person, particularly of the mouth. Some techniques can even replace the audio streams of a video with synthetic speech that replicates the speech patterns of a well-known person.[121] This development is serious, because it enables the misrepresentation of elected leaders and representatives, thereby fomenting mistrust and compromise national security.[122] It has been used to feature pornography of innocent people.[123] It has already occurred in the biomedical sector wherein scientific images and 3D imagery have been falsified through digital manipulation to support fraudulent papers.[124]

---

118 Wikipedia, "Deepfake" <https://en.wikipedia.org/wiki/Deepfake> (accessed 15 July 2020).

119 Wikipedia, "Autoencoder" <https://en.wikipedia.org/wiki/Autoencoder> (accessed 15 July 2020).

120 Wikipedia, "Generative Adversarial Network" <https://en.wikipedia.org/wiki/Generative_adversarial_network> (accessed 15 July 2020). The method uses two competing neural networks, known as the generator and the discriminator. The generator produces a convincing image to replace the original image, and the discriminator network compares the false footage with the real videos to find flaws. Each item of software learns from the others until the flaws are removed.

121 In 2019, criminals used software to impersonate the voice of a chief executive to authorise a fraudulent transfer of €220,000: Catherine Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case" *The Wall Street Journal* (30 August 2019).

122 For a high-level introduction, with a brief discussion of the techniques that can be used, see Britts Paris & Joan Donovan, *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence* (Data & Society, 2019) <https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1.pdf> (accessed 15 July 2020).

123 See, *eg*, Janko Roettgers, "Porn Producers Offer to Help Hollywood Take Down Deepfake Videos" *Variety* (21 February 2018) <https://variety.com/2018/digital/news/deepfakes-porn-adult-industry-1202705749/> (accessed 15 July 2020).

124 By way of example, see Douglas W Cromey, "Digital Images Are Data: And Should Be Treated as Such" (2013) 931 Methods Mol Biol 1; and Yisroel Mirsky *et al*, "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning", *(cont'd on the next page)*

45 Various initiatives have been started to counter the problem of the manipulation of images and voices.[125] However, the technology supporting deep fakes is also developing at a fast pace.[126] As these software programs are designed to improve as they work, this means there may come a time when it will not be possible to detect the manipulation,[127] although there are attempts being made to counter this problem.[128]

46 Based on current technologies, detecting a manipulated image takes time, expertise and appropriate tools. Should the authenticity of a digital image be doubted,[129] a digital evidence professional will conduct an investigation based on a reverse image search,[130] an analysis of the image metadata about the camera and image to detect any inconsistencies in

28th USENIX Security Symposium (USENIX Security, 2019) <https://www.usenix.org/system/files/sec19-mirsky_0.pdf> (accessed 15 July 2020).

125 This article mentions some of the projects, past and present: Eliza Strickland, "Facebook AI Launches Its Deepfake Detection Challenge at the NeurIPS Conference" *IEEE Spectrum* (11 December 2019) <https://spectrum.ieee.org/tech-talk/robotics/artificial-intelligence/facebook-ai-launches-its-deepfake-detection-challenge> (accessed 15 July 2020); Yue Wu, Wael AbdAlmageed & Premkumar Natarajan, "ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries With Anomalous Features" (2019) Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 9543 <https://openaccess.thecvf.com/content_CVPR_2019/html/Wu_ManTra-Net_Manipulation_Tracing_Network_for_Detection_and_Localization_of_Image_CVPR_2019_paper.html> (accessed 15 July 2020).

126 For an overview, see Henry Ajder *et al*, *The State of Deepfakes Landscape, Threats, and Impact* (Deeptrace, September 2019) <https://regmedia.co.uk/2019/10/08/deepfake_report.pdf> (accessed 15 July 2020).

127 Tero Karras, Samuli Laine & Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks" (2019) Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 4401 <https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html> (accessed 15 July 2020).

128 For which see the US Defense Advanced Research Projects Agency, which has initiated a new project called Semantic Forensics (SemaFor) <https://www.darpa.mil/program/semantic-forensics> (accessed 15 July 2020); also see Yuezun Li, Ming-Ching Chang & Siwei Lyu, "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking" 2018 IEEE International Workshop on Information Forensics and Security <http://www.cs.albany.edu/~lsw/papers/wifs18.pdf> (accessed 15 July 2020); and Scott McCloskey & Michael Albright, "Detecting GAN-generated Imagery Using Color Cues" (2019) <https://arxiv.org/pdf/1812.08247.pdf> (accessed 15 July 2020).

129 Hany Farid, *Fake Photos* (The MIT Press, 2019). The following process is described herein.

130 Although whether this is possible will depend on the sensitivity of the image under consideration and whether the image can be located on the Internet. See Andreas Rössler *et al*, "FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces" (2018) <https://arxiv.org/pdf/1803.09179.pdf> (accessed 15 July 2020).

the metadata and the image presented,[131] computing the image exposure including the adjustments that software and cameras will make to the image,[132] a determination of whether flash was used and analysing the light patterns, shadows and reflections,[133] and an analysis of vanishing lines, geometry of shadows, reflection analysis and lens flare. Other techniques that require intermediate technical skills include analysing the compression schemes for 2D images,[134] reconstructing faces and scenes in 3D, resampling,[135] and examining images for noise.[136]

47 Of course, one should not dispel the possibility that an image that appears to be incredible may in fact be a credible image.[137]

48 While software tools are readily available to allow an end user to test various hypotheses in the analysis of image manipulation, it remains, for the time being, the domain of the expert to interpret the test results and form a conclusion. One day, technologies might be available to perform a set of tests which can be weighted[138] and used to draw "informed conclusions" about the image being manipulated. However, given advancements in ML technologies, it is not expected that such systems would be able to determine if the content of an image is 100% real or 100% computer generated, since AI systems can likewise be programmed to generate "natural" artefacts that would fool an automated analysis system, let alone the human eye.

---

131 Metadata would include data about the camera (make, model), shutter speed, aperture size, focal length, image format, compression, compatibility, geo-location information, date, time and location tags. The metadata can be used to match an image to a particular device. In addition, when an image is saved and manipulated, the metadata might be modified, augmented or removed.

132 This explores the quantitative relationship between the camera settings and the properties of the image: exposure, depth of field, motion blur, ISO (International Organization for Standardization) settings.

133 For deep fakes, this includes reviewing light patterns on the surfaces of eyes and ears.

134 This involves looking for artefacts from the JPEG compression scheme such as JPEG signatures and ghosts. See Shruti Agarwal & Hany Farid, "A JPEG Corner Artifact from Directed Rounding of DCT Coefficients" (TR2018-838) *Dartmouth Digital Commons* (1 February 2018) <https://www.cs.dartmouth.edu/~trdata/reports/abstracts/TR2018-838/> (accessed 15 July 2020); Shruti Agarwal & Hany Farid, "Photo Forensics from JPEG Dimples" <https://farid.berkeley.edu/downloads/publications/wifs17.pdf> (accessed 15 July 2020); and Hany Farid, "Image Forensics" (2019) 5 *Annual Review of Vision Science* 549.

135 This involves comparing pixel patterns in the image that result from how the image was recorded and manipulated.

136 This involves assessing imperfections introduced by the software in a camera.

137 Hany Farid illustrates this in *Fake Photos* (The MIT Press, 2019) at pp 38–45 and points out that it can be important to establish whether a gruesome image of a beheading was plausible.

138 The authors owe this observation to Andrew Sheldon MSc, with thanks.

49 In a sense, therefore, the evidential treatment of the issue of manipulated digital data is no different from that of other electronic evidence that needs authentication. In summary, issues as to authentication, including issues with digital data that is manipulated, requires the courts to develop a set of clear procedures for managing authentication issues, a healthy appreciation of the limits of the presumption of reliability, as discussed above, and a robust approach towards disclosure or discovery, which will be discussed below. Only then can these issues be effectively elucidated.

## VIII. Disclosure or discovery

50 The primary objective of discovery[139] and disclosure[140] is to enable a party to acquire information which he does not have concerning the issues in the proceeding so that he can effectively prepare and present his case for adjudication.[141] Without effective discovery, a party may not have a sufficient opportunity to challenge the opposing party's evidence, including calling witnesses to rebut the evidence and raising new arguments in relation to the facts.[142]

51 Discovery is a matter for the inherent jurisdiction of the courts,[143] and is obtained only by way of court order.[144] In evaluating a request for discovery, the court has to consider:[145]

> … a balancing of the interests of the parties, how impelling is the need for the information and how expensive and intrusive will be the exercise of making discovery [with account] to be taken of alternative means for providing the information or otherwise satisfying the reasonable requirements of the applicant for discovery.

52 Discovery in both civil and criminal proceedings is profoundly important and complex in relation to electronic evidence, especially since much of our information today is in digital form.[146] While

139 Rules of Court (Cap 322, R 5, 2014 Rev Ed) O 24.

140 Criminal Procedure Code (Cap 68, 2012 Rev Ed) Pt IX, Division 2, ss 160–163 *ff*; Pt X, Division 5, ss 212–215 *ff*; Civil Procedure Rules 1998 (UK) r 31.

141 Jeffrey Pinsler, *Civil Practice in Singapore and Malaysia* (Butterworths Asia, looseleaf, 1998) ch XXIIA, "Character and Function of Discovery Process", at para 1.

142 Jeffrey Pinsler, *Civil Practice in Singapore and Malaysia* (Butterworths Asia, looseleaf, 1998) ch XXIIA, "Character and Function of Discovery Process", at para 1.

143 Supreme Court of Judicature Act (Cap 322, 2007 Rev Ed) First Schedule, para 12.

144 Rules of Court (Cap 322, R 5, 2014 Rev Ed) O 24 r 1 (abolishing automatic mutual discovery).

145 *Baldock v Addison* [1995] 1 WLR 158 at 163.

146 For the US, see Matt Tusing, "Machine-Generated Evidence" (2016) 43(1) *The Reporter* 13; Sonia K Katyal, "The Paradox of Source Code Secrecy" (2019)

*(cont'd on the next page)*

discovery in relation to documentary evidence is routine, increasingly more discovered evidence is in electronic form. Although discovery rules will operate in largely the same way with electronic evidence (subject to rules about electronic discovery which will be discussed later), in many circumstances, it is contended that effective discovery has to go beyond a document in digital form or an electronic record to also encompass the technology – be it the system, data or software code – that generates or produces the electronic record.

53 This arises when questions are raised about, for instance, the reliability, authorship or the authenticity of the electronic evidence and discovery is sought by the proponent for additional evidence to investigate these issues. If the opponent denies this request, the proponent will rarely be in a position to offer evidence to refute the opponent's denial, especially when the proponent's discovery request is general and non-specific. This is because only the party in possession of the electronic evidence – the opponent – has the resources, access rights or knowledge to fully understand the system from which the evidence was extracted.[147] Yet the proponent has to overcome the threshold requirement to demonstrate that discovery is needed before being able to access that evidence. In criminal proceedings, if the proponent is the defendant, this has the unfair effect of undermining the presumption of innocence. This "catch-22" situation enables opponents to seize upon this lack of specificity on the part of the proponent's discovery request and challenge it as being overly broad and non-specific.

### *A. Electronic discovery or "predictive coding"*

54 The opposite situation to discovery may also occur: the popularity of the use of electronically stored information is such that one party may flood the other with such large quantities of data in electronic form such that it becomes impractical, infeasible and too costly for the party to discover the material required for his case.[148]

104 Cornell L Rev 1183; and Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System" (2018) 70 Stan L Rev 1343, in which the author considers the history of the trade secret privilege, uncovering an interesting development where it was demonstrated that Wigmore was initially hostile to the privilege (at 1383), but his opinion later changed. He admitted in an aside that his brother had suffered loss relating to intellectual piracy (at 1385).

147 *Electronic Evidence* (4th Ed) at paras 6.194 and 6.224.

148 Yeong Zee Kin, "Recent Developments in Electronic Discovery – Discovering Electronic Documents and Discovering Documents Electronically" (2007) 19 SAcLJ 101.

55 The introduction of electronic discovery seeks to address that imbalance. Basically, electronic discovery is the use of any search method that combines technology and automated processes to help search for, preserve, retrieve and produce relevant information.[149] The ease or difficulty of conducting such searches will depend on the form, location and accessibility of the electronic records that make them conducive to identification. Well-organised information with rich metadata can be rapidly identified through a combination of search terms and metatag searches. But loosely grouped documents may stymie typical searches on broad allegations and loose terms.

56 The solution is to use AI-driven searches. Also known as "predictive coding" or, more accurately, technology-assisted review ("TAR"), it has received judicial endorsement in the US,[150] England and Wales[151] and Australia.[152] While Singapore courts have yet to weigh in on this matter, the Singapore Supreme Court Practice Directions have made provision for the discovery and inspection of electronically stored documents[153] by way of encouraging parties to collaborate on mutually-agreed electronic discovery plans[154] and allowing for reasonable searches for electronically stored documents.[155] The Practice Directions are technology neutral in relation to the methodology deployed, and make no mention of predictive coding. In fact, in many legal contexts, searches conducted using search terms may yield perfectly adequate and more cost-effective results than predictive coding.[156]

57 Predictive coding works by substituting an ML system trained to search for specific information in specified documents in place of a lawyer. A sample of documents to be searched (known as a "seed set") is first manually reviewed by lawyers, who code each document as "responsive" or "not responsive" to the case. This information is fed into the ML system, which then builds an internal statistical model for

149 "The Sedona Principles, Third Edition, Best Practices, Recommendations & Principles for Addressing Electronic Document Production" (2018) 19 Sedona Conf J 1 at 165.

150 *Da Silva Moore v Publicis Groupe* 287 FRD 183 (SDNY, 2012).

151 *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch).

152 *McConnell Dowell Constructors (Aust) Pty Ltd v Santam Ltd (No 1)* [2016] VSC 734.

153 Supreme Court Practice Directions, Pt V: "Discovery and Inspection of Electronically Stored Documents".

154 Supreme Court Practice Directions, Pt V: "Discovery and Inspection of Electronically Stored Documents" at para 45.

155 Supreme Court Practice Directions, Pt V: "Discovery and Inspection of Electronically Stored Documents" at para 47.

156 "The Sedona Principles, Third Edition, Best Practices, Recommendations & Principles for Addressing Electronic Document Production" (2018) 19 Sedona Conf J 1 at 165.

predicting the "responsiveness" of documents. This model is then tested on another sample of documents. This training process is repeated until the accuracy of the ML system reaches the requisite level of relevancy, whereupon it can be used on the document set of the case.[157] The use of predictive coding to help search for relevant information can yield cost and time savings, and is often a practical necessity when the party has to deal with voluminous electronic documents.[158]

58 Predictive coding is not a substitute for review by skilled lawyers. It merely helps to narrow down the scope of documents sought to be reviewed so that a more thorough, manual review can be conducted with fewer mistakes. The effectiveness of this process also requires observance of predictive coding best practices. To work properly, the ML system also has to be trained properly, typically by the most experienced lawyers with the requisite domain knowledge of the subject matter.

59 Finally, a lawyer using predictive coding ought to have some knowledge about how the ML system works, to help the lawyer use the predictive coding system correctly and to enable her to discharge her legal and ethical obligations to both her client and to the court. For instance, the requesting party may be challenged by the responding party to explain her choice of predictive coding as a search method and the processes employed, including how the seed sets are constructed,[159] and also to explain the possible limitations of predictive coding to her clients. Of course, when the dispute turns on the possible negligence of the lawyers and developers of the system concerned in the discovery process, the predictive coding tool that is used for discovery itself and its software code may be the subject matter of its own discovery application.

### *B. Difficulties of discovery of software code*

60 In practice, software codes are rarely disclosed voluntarily. The applicant usually has to apply to a judge to order the software code to be

157 "How to Make the E-Discovery Process More Efficient With Predictive Coding" *Thomson Reuters, Legal* <https://legal.thomsonreuters.com/en/insights/articles/how-predictive-coding-makes-e-discovery-more-efficient> (accessed 15 July 2020); exterro, *Basics of E-Discovery*: "Chapter 7B: Predictive Coding (Technology Assisted Review) & Artificial Intelligence" <https://www.exterro.com/basics-of-e-discovery/predictive-coding> (accessed 15 July 2020).

158 "The Sedona Principles, Third Edition, Best Practices, Recommendations & Principles for Addressing Electronic Document Production" (2018) 19 Sedona Conf J 1 at 164.

159 "The Sedona Principles, Third Edition, Best Practices, Recommendations & Principles for Addressing Electronic Document Production" (2018) 19 Sedona Conf J 1 at 166; *Rio Tinto plc v Vale SA* 2015 WL 872294 at 2 (SDNY).

provided to them. The owner or manufacturer will often object on the grounds of confidentiality and trade secrecy, or protecting intellectual property, to prevent the software from being the subject of scrutiny by an independent third party. The court may also rule that such discovery is not necessary for disposing fairly of the action or will incur unnecessary costs.[160] It is therefore relatively unusual for a judge to order the disclosure of relevant software, yet arguably it may be crucial to fully understand a report or test result generated by an automated system.[161]

61 This may be best illustrated with the recent brouhaha regarding alcohol breath test machines. For decades, printouts of breath alcohol analysis machines taken by police officers were routinely presented in court as evidence of suspect drink-drivers and accepted for what they were.[162] Defence lawyers had tried to forensically examine these machines and inspect their code, but courts in at least six US states, including New York, had repeatedly rebuffed these discovery requests.[163] Likewise the manufacturers of these devices had objected, raising intellectual property and trade secret objections to the disclosure of their source code.[164] All this changed in 2015, when defence counsel succeeded in getting discovery orders to review the software that operated the breath-testing devices. It transpired that there were troubling mistakes including calibration problems and calculation errors in the codes of these machines. As a consequence, many of these devices were generating results that were 20% to 40% too high. This has put 42,000 drink-driving convictions in Massachusetts and New Jersey at risk, with other state courts starting to recognise the lack of reliability of these machines and convictions secured using these devices.[165]

62 Another example can be drawn from the "sudden unintended acceleration" spate of cases, where vehicles suddenly and unexpectedly accelerate, often accompanied by an apparent loss of braking effectiveness.[166] While many of these reported incidents were due

160 See, *eg*, O 24 r 7 of the Rules of Court (Cap 322, R 5, 2014 Rev Ed).

161 For a discussion of the US position, see, *eg*, Christopher M Mislow, "Protecting Source Code from Disclosure during Pretrial Discovery" (1984)12 Utah Bar J 39.

162 See, *eg*, *R v McKeown; R v Jones* [1997] 1 WLR 295; and *Director of Public Prosecutions v Barber* (1998) 163 JP 457. For a discussion of the US litigation surrounding breath alcohol analysis devices, see *Electronic Evidence* (4th Ed) at paras 6.184–6.185.

163 "These Machines Can Put You in Jail. Don't Trust Them" *The New York Times* (3 November 2019).

164 "Breathalyzer Giant Accused of Fraud Won't Come Clean about Booze Tests" *Daily Beast* (3 February 2020).

165 "These Machines Can Put You in Jail. Don't Trust Them" *The New York Times* (3 November 2019).

166 Wikipedia, "Sudden Unintended Acceleration" <https://en.wikipedia.org/wiki/Sudden_unintended_acceleration> (accessed 15 July 2020).

to human mistake, many of these could not be so attributed. In the Oklahoma case of *Bookout v Toyota Motor Corp*,[167] the defendant and her passenger were in their 2005 Toyota Camry when it suddenly accelerated. The passenger made a harrowing telephone call briefly describing that the car was running away, that the defendant could not stop the car and that something was wrong with it.[168] The skid marks and the call together undermined any suggestion that the acceleration was due to a physical problem, which led the judge to order the discovery of the software code for the Camry.[169] Subsequent investigations by the experts suggested that the Toyota electronic throttle control system contained many software defects and that at least one of them was capable of causing a malfunction in the electronic throttle control module that could cause unintended acceleration. Conversely, had it not been for the disclosure of the telemetry and sensor data for the investigations surrounding the Uber autonomous vehicle, we would not have known the weaknesses in the environmental perception system and the overall sensor and vehicular control system[170] deployed by Uber. It is only with the robust operation of discovery laws to facilitate proper investigations into the causes of autonomous vehicle accidents that such technologies can be further refined and improved, with a view towards the deployment of fully-autonomous vehicles.[171]

### *C. Sufficiency of case to order discovery*

63 The procedures for discovery in civil proceedings are clear as to the sufficiency of the case to be made to order discovery: the proponent must make a case that discovery is necessary either for disposing fairly of the case or matter or for saving costs.[172] However, in criminal

167 Case No CJ-2008-7969. The trial was held in the District Court of Oklahoma County, State of Oklahoma before the Hon Patricia G Parrish, District Judge.

168 *Electronic Evidence* (4th Ed) at para 6.226.

169 *United States of America, Central District of California, Case Protective Order In re: Toyota Motor Corp Unintended Acceleration Marketing, Sales Practices and Products Liability Litigation* (Case Number: 8:10ML2151 JVS (FMOx)) (2018) 15 *Digital Evidence and Electronic Signature Law Review* 98.

170 Youngjin Yoo, Ola Henfridsson & Kalle Lyytinen, "Research Commentary: The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research" (2010) 21(4) *Information Systems Research* 724; Jianfeng Zhao, Bodong Liang & Qiuxia Chen, "The Key Technology toward the Self-driving Car" (2018) 6(1) *International Journal of Intelligent Unmanned Systems* 2.

171 Professor Martyn Thomas, CBE, Gresham College, "Is Society Ready for Driverless Cars?", lecture at the Museum of London (24 October 2017) <https://www.gresham.ac.uk/lectures-and-events/is-society-ready-for-driverless-cars> (accessed 15 July 2020); Roger Kemp, "Autonomous Vehicles – Who Will Be Liable for Accidents?" (2018) 15 *Digital Evidence and Electronic Signature Law Review* 33.

172 Rules of Court (Cap 322, R 5, 2014 Rev Ed) O 24 r 7; for England and Wales: Civil Procedure Rules 1998 Pt 31, "Disclosure and inspection of documents".

proceedings, there has been little discussion regarding the sufficiency of the case the Defence must adduce to persuade a judge to order appropriate disclosure,[173] let alone the code or internal data of a device or system that generated an incriminating report relied on by the Prosecution.[174] And even when such a request is tendered, courts have tended to reject the Defence request as "inadequate" or amounting to "fishing expeditions".[175] However, claims that the Prosecution would have neither ownership nor access to the code or data of the device or system sought by the Defence, because the code or data is protected by third-party intellectual property rights,[176] should not be a bar to disclosure. This is because courts can make various protective orders to remedy the problem of loss of intellectual property rights by restricting access to the code or data only by approved experts, requiring proceedings to be conducted "*in camera*", mandating the redaction of court orders and judgments that refer to sensitive information and destroying all pertinent exhibits and notes.[177]

64 As an alternative to ordering the disclosure of the source code or data, Imwinkelried has proposed in its place the use of validation studies. Drawing upon the approach taken with the admission of modern DNA typing evidence, Imwinkelried proposed that prosecutors first lay a foundation for the reliability of evidence generated by devices by presenting testimony about validation studies conducted to establish the reliability of the scientific techniques used, including those incorporated in the device or system.[178] The onus will then be placed on the Defence to show that the facts of the instant case are outside the validation range of the validation studies and that this will have a material bearing on the test outcomes. If this is established, the device or system manufacturer is given the choice of resubmitting the system and its code or data to a fresh validation study or, in the alternative, disclosing its code (or data) to the

173 In Singapore, the defendant may make an objection to any issue of fact contained in any matter contained in the Case for the Prosecution. See ss 165(1)(*d*) and 217(1)(*d*) of the Criminal Procedure Code (Cap 68, 2012 Rev Ed); see also ss 225A and 225B. For England and Wales, see Pt 15 of the Criminal Rules and Practice Directions.

174 Edward J Imwinkelried, "Computer Source Code: A Source of the Growing Controversy over the Reliability of Automated Forensic Techniques" (2017) 66 DePaul L Rev 97.

175 See, *eg*, *State v Underdahl* 749 NW 2d 117 at 121 (MinnApp, 2008); *People v Robinson* 53 AD 3d 63 at 72; 860 NYS 2d 159 (2008); and *People v Cialino* 14 Misc 3d 999 at 1001; 831 NYS 2d 680 (NYCity CrimCt, 2007).

176 *People v Cialino* 14 Misc 3d 999; 831 NYS 2d 680 at 681 (NYCity CrimCt, 2007).

177 Edward J Imwinkelried, "Computer Source Code: A Source of the Growing Controversy over the Reliability of Automated Forensic Techniques" (2017) 66 DePaul L Rev 97.

178 Edward J Imwinkelried, "Computer Source Code: A Source of the Growing Controversy over the Reliability of Automated Forensic Techniques" (2017) 66 DePaul L Rev 97, fn 74.

Defence.[179] However well intentioned this proposal may be, it ought to be noted that validation studies are likely to be adequate only for processes with well-defined and scientifically testable processes operating on simple procedures such as those used in forensic chemistry. They are unlikely to be helpful for complex systems such those used in AI systems, and their use may raise additional questions such as the number of validation tests required, the assumptions made as to the number of such tests, the procedures used to conduct the tests and how these can be conducted within a practical period of time.[180] The tests are merely a proxy for a detailed examination of the system itself in the context of the instant fact situation with its specific parameters. Even if the system passes the validation test or even if the full test history of all validation tests conducted are made available, this may not be indicative as to whether the system has correctly processed the instant case.

### *D. Transparency and regulatory oversight*

65 In this regard, some form of regulatory oversight to promote the accessibility of the data that is fed into the AI system will enable greater governance over such platforms and facilitate transparency of the AI system.[181] In fact, it is a regulatory requirement in Singapore and the US for the licensed operator of an autonomous vehicle to

179 Edward J Imwinkelried, "Computer Source Code: A Source of the Growing Controversy over the Reliability of Automated Forensic Techniques" (2017) 66 DePaul L Rev 97, fn 191.

180 With thanks to Professor Martyn Thomas, CBE, e-mail communication between Stephen Mason and Professor Thomas. In August 2020, Alex Chalk MP (Parliamentary Under Secretary of State at the Ministry of Justice) invited Paul Marshall (Barrister, Cornerstone Barristers, 2-3 Gray's Inn Square, Gray's Inn) to submit a paper to the Ministry of Justice on suggestions for improving the existing approach to the proof in court proceedings of computer-derived evidence. A working group was established to consider the issue. At the time the paper was submitted to the Ministry of Justice, in November 2020, it was explained that it was intended to publish the paper. That paper, slightly edited, is now published: Paul Marshall *et al*, "Recommendations for the Probity of Computer Evidence" (2021) 18 *Digital Evidence and Electronic Signature Law Review* 18.

181 See, by way of example: Riccardo Guidotti *et al*, "A Survey of Methods for Explaining Black Box Models" (2018) 51(5) *ACM Computing Surveys*; Emily Berman, "A Government of Laws and Not of Machines" (2018) 98 BU L Rev 1277; Cary Coglianese & David Lehr, "Transparency and Algorithmic Governance" (2019) 71 *Administrative Law Review* 1; Roger Clarke, "Regulatory Alternatives for AI" (2019) 35(4) *Computer Law & Security Review* 398; Melissa Hamilton, "The Biased Algorithm: Evidence of Disparate Impact on Hispanics" (2019) 56 Geo LJ 1553; Leah Wisser, "Pandora's Algorithm Black Box: The Challenges of Using Algorithmic Risk Assessment in Sentencing" (2019) 56 Geo LJ 1811; *Responsible AI A Global Policy Framework* (Charles Morgan ed) (International Technology Law Association, 2019); and Deven R Desai & Joshua S Kroll, "Trust but Verify: A Guide to Algorithms and the Law" (2017) 31 Harv JL Tech 1 for a counter argument.

ensure that it has records of the sensory data captured by the vehicle for accident-investigation purposes.[182] Because ML systems are able to execute numerous functionalities or exhibit different features in varying conditions, it may be difficult to explain how they function and how they arrive at a particular prediction or functionality in a particular context or situation.

66 This inherent opaqueness of ML systems is paradoxically what makes these systems so powerful and yet so difficult to understand.[183] The move towards explainable AI (or "XAI") engineering to develop tools that provide some level of interpretability of an AI model's operation represents the use of technical solutions to make the operation of AI systems more transparent and more explainable.[184] But this should not detract from the overall AI system design and more specifically, model training and feature selection obligations by the developer.[185]

67 Therefore, whether it is necessary, and in fact helpful, to disclose the code for an algorithm audit and, crucially, in the case of AI systems, whether this transparency requirement extends to disclosing the training and test data themselves so that the models can be reviewed will depend on the jurisdiction of the reviewing party, actual application of the

182 See, *eg*, r 17(4) of the Road Traffic (Autonomous Motor Vehicles) Rules 2017 (S 464/2017) (requiring data recorder on autonomous vehicle to be kept for at least three years from date of recording, and to record, among others, the vehicle speed, location, status, operator override history, sensor data and camera or video footage), California Title 13 Division 1, Chapter 1, Art 3.7: Testing of Autonomous Vehicles, § 288.06(a)(6).

183 Tal Zarsky, "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making" *Science, Technology, & Human Values* (14 October 2015) <https://journals.sagepub.com/doi/10.1177/0162243915605575> (accessed 15 July 2020).

184 See Personal Data Protection Commission, *Singapore Model AI Governance Framework* (2nd Ed, 21 January 2020) at para 3.27 (hereinafter "*Singapore AI Framework*"). Examples of such tools include the use of surrogate models, partial dependence plots, global variable importance/interaction, sensitivity analysis, and counterfactual explanations.

185 Other solutions include robust government-led testing, the introduction of fairness, ethical and safety standards into algorithms themselves and auditing algorithms to enhance their interpretability to detect biases and verify outputs against safety requirements. See, *eg*, Hazel Si Min Lim & Araz Taeihagh, "Algorithmic Decision-Making in AVs: Understanding Ethical and Technical Concerns for Smart Cities" (2019) 11(20) *Sustainability* 5791; Noah J Goodall, "From Trolleys to Risk: Models for Ethical Autonomous Driving" *AJPH* (April 2017) <http://dx.doi.org/10.2105/AJPH.2017.303672> (accessed 15 July 2020); and Rowan McAllister *et al*, "Concrete Problems for Autonomous Vehicle Safety: Advantages of Bayesian Deep Learning" Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia (2017) <https://www.ijcai.org/Proceedings/2017/0661.pdf> (accessed 15 July 2020).

AI system and the nature of the ensuing dispute.[186] In any event, it is in an organisation's own interest to ensure that its automated decision-making processes are explainable, transparent and fair, not only for building trust and confidence in the organisation's AI systems,[187] but also as part of the organisation's corporate risk management framework.[188]

## IX. Conclusion

68 The increasing sophistication of AI systems and their pervasive use show no signs of abating. Likewise, courts and lawyers alike will increasingly have to contend with the issues associated with the admissibility of evidence generated and produced by AI systems. Reliability, hearsay, authentication and disclosure issues will undergird many of the admissibility considerations of courts and lawyers, not just for legal proceedings, but also for non-contentious matters as well.

69 This article demonstrates that a resolution of these issues in evidence calls for a firm foundation in evidence and an awareness of the pertinent technological considerations. AI evidence should be carefully scrutinised and not readily presumed to be reliable. The possible hearsay elements in AI evidence should be teased out and information attributable as human assertions should be evaluated carefully. The veracity of AI evidence itself in what it stands for should also be subjected to the anvil of authentication. And lastly, disclosure or discovery rules should accommodate the relatively opaque and non-transparent nature of AI systems and not be too chary to deny a request for evidence that would explain how such systems work.

70 Finally, it is hoped that this article has adequately conveyed the point that electronic evidence is not a specialist area of legal practice but one that cuts across all practice groups and domains. If these issues in electronic evidence are not easily understood by both the Bench and the Bar, the legal profession as a whole should seek urgent steps to secure the requisite knowledge and expertise. And it is hoped that this call for the education of lawyers in electronic evidence will be heard and met,[189]

186 *Singapore AI Framework*, Annex B.
187 *Singapore AI Framework*, Annex B at p 15.
188 *Singapore AI Framework*, Annex B at p 16.
189 "Editorial" (2010) 7 *Digital Evidence and Electronic Signature Law Review* 5; Denise H Wong, "Educating for the Future: Teaching Evidence in the Technological Age" (2013) 10 *Digital Evidence and Electronic Signature Law Review* 16; Deveral Capps, "Fitting a Quart into a Pint Pot: The Legal Curriculum and Meeting the Requirements of Practice" (2013) 10 *Digital Evidence and Electronic Signature Law Review* 23.

because it behoves the legal profession to ensure that it remains relevant and pertinent, and is not left behind in the information era.