

# Northumbria Research Link

Citation: Oswald, Marion (2020) Technologies in the twilight zone: Early lie detectors, machine learning and reformist legal realism. *International Review of Law, Computers and Technology*, 34 (2). pp. 214-231. ISSN 1360-0869

Published by: Taylor & Francis

URL: <https://doi.org/10.1080/13600869.2020.1733758>  
<<https://doi.org/10.1080/13600869.2020.1733758>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/id/eprint/41944/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



**Northumbria  
University**  
NEWCASTLE



**UniversityLibrary**



## Technologies in the twilight zone: early lie detectors, machine learning and reformist legal realism

Marion Oswald

To cite this article: Marion Oswald (2020) Technologies in the twilight zone: early lie detectors, machine learning and reformist legal realism, International Review of Law, Computers & Technology, 34:2, 214-231, DOI: [10.1080/13600869.2020.1733758](https://doi.org/10.1080/13600869.2020.1733758)

To link to this article: <https://doi.org/10.1080/13600869.2020.1733758>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 04 Mar 2020.



Submit your article to this journal [↗](#)



Article views: 612



View related articles [↗](#)



View Crossmark data [↗](#)

# Technologies in the twilight zone: early lie detectors, machine learning and reformist legal realism

Marion Oswald

Northumbria Law School, University of Northumbria, Newcastle upon Tyne, UK

## ABSTRACT



Contemporary discussions and disagreements about the deployment of machine learning, especially in criminal justice contexts, have no foreseeable end. Developers, practitioners and regulators could however usefully look back one hundred years to the similar arguments made when polygraph machines were first introduced in the United States. While polygraph devices and machine learning operate in distinctly different ways, at their heart, they both attempt to predict something about a person based on how others have behaved. This paper, through an historical perspective, examines the development of the polygraph within the justice system – both in courts and during criminal investigations – and draws parallels to today's discussion. It can be argued that the promotion of lie detectors supported a reforming legal realist approach, something that continues today in the debates over the deployment of machine learning where 'public good' aims are in play, and raises questions around how key principles of the rule of law can best be upheld. Finally, this paper will propose a number of regulatory solutions informed by the early lie detector experience.

## KEYWORDS

Polygraph; machine learning; legal realism

## Introduction

If machine learning and artificial intelligence were people, they would be teenagers, and young teenagers at that. Systems powered by AI rely on several methods and techniques (The European Commission's High Level Expert Group on Artificial Intelligence 2018), many of which are in a state of flux (Hao 2019). The science behind the technology is not yet mature, established or universally accepted (Christin 2017; Syll 2019; Richardson, Schultz, and Crawford 2019; Leetaru 2019; McStay 2018), with Google's Ali Rahimi recently describing machine learning as 'alchemy' (Hutson 2018). Machine learning is already embedded into products and applications – Internet search, behavioural advertising and online shopping recommendations – where performance can be just good enough without (most of the time) serious consequence. Despite its immaturity however, machine learning is increasingly promoted and used in high-stakes decision-making

**CONTACT** Marion Oswald  marion.oswald@northumbria.ac.uk  Northumbria Law School, University of Northumbria, Newcastle upon Tyne, UK

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

environments, such as criminal justice, health, immigration, staff performance monitoring and fraud detection. In addition, ‘tech-fantasy’ has become embedded in media culture (Naughton 2019), leading to distorted dreams of what the technology can actually do (Shneiderman 2019). As McQuillan comments, ‘machine learning is nothing like the emergent general intelligence that characterizes cultural representations of AI and is instead a set of mathematical methods that can perform amazing yet utterly thoughtless feats of classification’ (McQuillan 2018).

This is not the first time that a technology with an element of the fantastical about it has become part of popular cultural life and, in parallel, promoted as a solution to serious societal issues. The first part of the twentieth century witnessed considerable enthusiasm in the United States for the potential utility of the polygraph device, or ‘lie detector’. Its use in criminal cases marked a rejection of ‘born criminal’ theories (Bunn 2012), and was supposed to herald ‘a new era, free from the crude third degree methods’ (Fioch 1949–50). It was said to be more scientific than potentially unreliable eye witness testimony (*People v. Kenny* 167 Misc. 51, 52 (N.Y. Misc. 1938)). Claims were made for high degrees of certainty, even that polygraph methods were 100% efficient and accurate. Bunn argues that the ‘legend’ of the lie detector was to a considerable extent an invention of the detective pulp fiction of the time, ‘a byproduct of the whodunit’ (Bunn 2012). Yet the device’s image of magic infallibility was contested from the start. The science was disputed, transparency demanded, legal and ethical issues raised. Even one of the inventors of the modern polygraph, Leonarde Keeler, argued that there was, in fact, no such thing as a ‘lie detector’ (Keeler 1934).

The historical approach taken in this paper follows that advocated by Pound: the use of history can be used to

illustrate ... how legal precepts – rules, principles, conceptions and standards – have met concrete situations of fact in organising human society in the past and enabling or helping us to judge how we may deal with such situations with some assurance in the present. (R. Pound quoted in Cahillane 2016, 69)

This paper also uses elements of *legal archaeology* by considering case facts in their historical context, other evidence and wider historical and social settings ‘outside the law library’ (Simpson 1995), in order to explore the ‘micro-history’ of key cases and legislation. This approach can help to understand underlying social, policy or political goals, (Novkov 2011) particularly relevant to the landmark case of *Frye* discussed below (*Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923)). The author is aware of the need to avoid the dangers of using history simply to explain the present or viewing history as linear progress (Cahillane 2016). A degree of caution is required regarding generalisations based upon past legal treatment in one jurisdiction of one particular technology deployed within a social context very different from today’s. This paper argues however that sufficient parallels can be drawn between the two technologies (and the aims of those who would promote their use) to justify learning from the early polygraph experience. The link to the polygraph ‘hangs over’ modern machine learning-based emotion detection technology (McStay 2018), and the polygraph itself has gained recent acceptance in the US and UK for post-offence monitoring of sex and other violent offenders. Bunn describes the polygraph as a ‘disciplinary technique in the arsenal of “technologies of the self” held by those authorities whose responsibilities include classification, regulation and

normalization’ (Bunn 2012). This description could apply equally to the use of reactive machine learning systems (such as fraud detection), pre-emptive predictive tools (such as those designed to predict the risk of an individual committing a serious criminal offence) (Oswald et al. 2018), and systems that administer a sanction or decision based on analysis of an individual’s profile or predicted future behaviour (Yeung 2018).

Furthermore, the early arguments played out in US courts over the use of the polygraph suggest a wider struggle between proponents of a ‘reformist’ version of legal realism, and those who favoured traditional rules and concepts. This arguably continues today in the debates over the deployment of machine learning where ‘public good/benefit’ aims are in play, and raises questions around how key principles of the rule of law are to be upheld. The resolution (or lack of it) of such struggles as it relates to the polygraph reveal issues of law and policy, justice and legitimacy that must be tackled as a matter of urgency if we are not to find machine learning tools filling gaps that should not be filled, and therefore all AI and machine learning tools becoming tarred with the same brush. This paper contributes to this important debate by concluding with regulatory proposals for machine learning informed by the early lie detector experience.

Taking an evaluative approach of the way competing values have played out in practice, this paper will look back to early-mid twentieth century case-law and journal commentary, focused upon the US, that considered whether polygraphs had a place in the criminal law investigative and trial process. It will highlight a number of rules and proposed principles and standards that emerged, and propose how these could help guide the integration of machine learning into present and future decision-making. It focuses upon the US and UK legal frameworks and in particular upon machine learning that relates to decisions affecting individuals within their policing and criminal justice systems. It is beyond the scope of this article to consider the reasons why polygraph investigative programmes took off in the US, Russia, Japan and many Eastern European states, but less so in the UK. Rules of evidence focused on reliability and relevance, and a strong data protection and human rights framework will have been influential but cannot present the full picture, as an initial (non-comprehensive) survey of European Court of Human Rights case law does not suggest any red-lines regarding the use of polygraph tests per se.<sup>1</sup>

This paper is structured thus. A concise history of the use of the early polygraph in the United States follows this introduction. The paper then assesses the reasoning behind key decisions that influenced the admissibility (or otherwise) of lie detector testimony in early criminal cases, and considers the use of lie detectors in other justice contexts. A discussion of analogies between lie detectors and machine learning identifies parallels between legal issues that emerge. The influence of reformist legal realism in respect of both technologies is then discussed. Finally, it is proposed that the operation of law with respect to polygraphs, and the systematic flaws and gaps in regulation that have been exposed as the technology has developed, suggests areas of focus for those engaged in policy considerations around present day use of machine learning.

## **A concise history of the early polygraph**

The invention of the polygraph had no eureka moment. Machines for measuring emotions (and later, lies) via bodily changes, together with their associated tests and

procedures, began in the 1880s in Italy with the aim of rendering ‘visible the criminal’s dangerousness’ (Bunn 2012). Increasing interest in criminology spawned a wide variety of measurement devices with disquieting names – for instance, the goniometer for measuring the angle of the face and the tachyanthropometer, a chair fitted with callipers designed to extract eleven measurements from the body. Measurement and systematic data collection were crucial to building criminology’s scientific authority (Bunn 2012). By the early 1900s, accounts of the ‘electric psychometer’ were featuring in the American press and in magazine stories (Bunn 2012). In a 1914 short story, novelist GK Chesterton had his fictional detective Father Brown and his friend Flambeau discuss the ‘new psychometric method’, Brown commenting that he found the method as valueless as the medieval idea that blood would flow from a body if its murderer touched it. Rather presciently, Father Brown added that ‘no machine can lie ... nor can it tell the truth’ (Chesterton 1914).

Hugo Munsterberg, a psychologist based at Harvard University, has been described as the American pioneer of modern lie detection (Grubin and Madsen 2005). His ‘truth-compelling machines’ included the ‘pneumograph’ which recorded variations in breathing and the ‘sphygmograph’ which recorded the heart (Bunn 2012). Highlighting problems with eye witness testimony, Munsterberg hoped to see his psychological tests used to understand the veracity of witnesses, bringing him into open conflict with John Henry Wigmore, a leading expert on evidence (Weiss, Watson, and Xuan 2014). William Marston, Harvard psychologist, lawyer and former student of Munsterberg, believed that the changes in systolic blood pressure (blood pressure when the heart beats) reflected whether or not people were answering questions deceptively. To carry out his deception test, he used a blood pressure cuff and measured the heart rate using a stethoscope after each question, stating each reading as it occurred, a time-consuming process (Lepore 2014–15). A charismatic proponent of the use of lie detector testimony in court, Marston’s evidence based on his test was put forward in the *Frye* case. Lepore notes that the lawyers for the accused were former Legal Psychology students of Marston’s, and argues that ‘The point wasn’t really to defend Frye; the point was to bring into a court of law a new science of evidence’ (Lepore 2014–15). Although Marston failed in his objective in the *Frye* case, his legacy was twofold: first, ‘transmuting early applied psychology into a criminological meme’, (Weiss, Watson, and Xuan 2014) and secondly, as his alter-ego Charles Moulton, creator of the character ‘Wonder Woman’, whose lasso of truth shares the lie detector’s characteristic of benign coercion.

PhD physiology student and part time Berkeley police employee, John Augustus Larson built on the work of Marston. He studied interrogations taking place in the police department and came up with a model for a device that took note of the reactions of the interviewee indicated by changes in blood pressure, pulse rates and respiration. In 1921, he created a method for chronicling this information on a rolling drum of paper so that a permanent record could be made, the first modern polygraph instrument (Grubin and Madsen 2005). Leonarde Keeler, a protégé of Larson’s, developed the first portable polygraph instrument (Grubin and Madsen 2005). Keeler also made changes to Larson’s invention by way of recording galvanic skin response due to sweat, based on the work of Fordham University psychologist Reverend Walter G. Summers. Supported by progressive police chief August Vollmer, Keeler’s work

within various police departments established the polygraph as an interrogative device in criminal investigations (Grubin and Madsen 2005).

In 1952, the operation of Keeler's portable detector was described as follows:

Four methods of detection are used ... A blood pressure cuff attached to the subject's upper right arm controls a pen which records changes in pulse and blood pressure. A harness is paced around the upper part of the subject's chest; this is connected with a pen which records changes in the rate of breathing. Two metal plates adjusted to the left wrist pick up electrical changes in the skin which are recorded by a third pen. The three pens make their records simultaneously on a moving roll of paper about eight inches wide. The most important recent development ... came as a result of research conducted in 1945 by John E. Reid. Reid found that a subject's blood pressure could be changed by various forms of unobserved muscular activity ... The instrument devised to record this muscular activity not only minimizes possible error in the other methods but also acts as a detector itself. ('The Lie Detector' 1952)

There was no one inventor of the lie detector, despite the use of this construct in contemporary reporting. While Larson and Keeler certainly built upon the work of Marston, they disagreed on many aspects, including whether a continuous or discontinuous blood pressure technique should be favoured, whether lowering of blood pressure could indicate deception and the relative importance of respiration and word association time in indicating deception (Bunn 2012). (The questioning process can often be overlooked but could be said to be as important as the selection of input data is to a machine learning model (Synnott, Dietzel, and Ioannou 2015).) Bunn argues it was exactly because the lie detector was 'old technologies applied to a new end', that its depiction as an invention was necessary to give it scientific credibility. Furthermore, the term 'lie detector' was 'a form of linguistic "black boxing": the simplification of scientific complexity and human agency' (Bunn 2012). Fundamentally though, all recognised polygraph methodologies share the same premise: that 'certain psychological processes result in physiological cues that can be measured and interpreted with the polygraph for the purpose of aiding in the detection of deception' with such measurements remaining largely unchanged from Keeler's original models (Synnott, Dietzel, and Ioannou 2015).

### **The early lie detector in, and out, of court**

The *Frye* decision is often cited as the case that made polygraph evidence inadmissible in court in the US. As mentioned above, the indictment of James Alphonso Frye for the murder of one Doctor Brown in Washington D.C. offered William Marston with an opportunity to present his lie detector evidence in court and so publicise his research. The extended argument between defence counsel and Justice McCoy over the submission of Marston's evidence makes entertaining reading (reproduced in Lepore 2014–15). The trial judge expressed the following main reasons for his decision to exclude Marston's evidence: (a) the test had been performed on Frye over a month previously and was therefore irrelevant to the truth or falsity of his statements during the hearing; (b) it was the jury's task to decide upon the question of whether anyone was telling the truth, making 'use of that thing which God Almighty has implanted in us, the power of observation'; (c) it was too late to carry out the test at the hearing itself; (d) due to the many variables, Marston's research was based on probabilities; (e) the machine itself was not 'infallible'.



Although it is of course impossible to be certain, a sense of impatience, and even some defensiveness, might be detected in the comments of the self-confessed ‘old’ judge ‘inured to certain general principles’. He was openly dismissive of Marston’s research, stating that it had taken all of five minutes to take in. The appeal by Frye’s lawyers to the D.C. Circuit Court of Appeals was based solely upon the exclusion of Marston’s testimony. They were again given short shrift. (Lepore notes that Marston was arrested for fraud just after the appeal was lodged, and the charges not dropped until after the appeal hearing, a situation that cannot have helped their cause (Lepore 2014–15)). Associate Justice Van Orsdel highlighted in his short judgment the difficulty of defining when a scientific principle or discovery crosses the line between ‘experimental and demonstrable’ (in the sense of being logically proved); it was this line which determined whether evidence should be admissible. Van Orsdel concluded that ‘*somewhere in this twilight zone*’, the evidential force of the principle must be recognised (*Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923)). However, ‘*the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs*’. This reasoning established the Frye test as the baseline for admitting new scientific evidence in the majority of US courts for the next 70 years, until the ‘general acceptance’ test was challenged in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993) and substituted by a test of reliability and relevance.

On the face of it, *Frye* seemed determinative as regards admission of lie detector testimony in criminal cases. Study of early case-law and journal commentary reveals a situation that was not so clear-cut. US appellate courts in the 1930s and 1940s appear to have been fairly consistent in their rejection of lie detector evidence based on the *Frye* standard (even if such evidence was put forward by the defendant to demonstrate innocence). Appellate courts recognised the potential ‘utility’ of the polygraph (*State v. Bohner* 210 Wis. 651, 246 N.W. 314 (1933); *State v. Lowry* 163 Kan. 622, 185 P. 2d 147 (1947)) in particular for the investigation of crime, (*State v. Cole* 354 Mo. 181, 188 S.W. 2d 43 (1945)) but rejected its place in the courtroom. For instance, the majority of the Nebraska Supreme Court considered that, if such evidence were admitted, the function of cross-examination would be impaired, and while the examiner could be cross-examined as to his qualifications and procedures used, ‘the machine itself... escapes all cross-examination’ (*Boeche v. State* 151 Neb. 368, 37 N.W. 2d 593 (1949)). Goldblatt comments that, despite apparent ‘scientific advancements and improvements’ in the twenty years since the *Frye* case, the ‘overwhelming majority of courts have closed their eyes and ears to the admissibility of lie detector evidence’ (Goldblatt 1950).

The same consistency of approach cannot be seen in early trial court decisions however. John E. Reid, polygraph developer and co-author of a book on lie detection and criminal investigation, points to ‘a great number of unappealed trial court cases ... in which lie-detector test results have been admitted as evidence over the objection of the opposing counsel’ (Reid 1954). These include the New York case of *People v. Kenny* in 1938. Raymond Kenny was charged with the hold-up of a delicatessen store and as a second offender, was facing a sentence of between thirty to sixty years if convicted (Legal Chatter 1938). The trial revolved around conflicting eye witness testimony, and the results of a lie detector test were offered to demonstrate Kenny’s innocence.

The scientific expert that was put forward to testify as to the results was Father Walter Summers, a Jesuit and Head of the Department of Psychology of the Graduate School of



Fordham University. During initial examination, he outlined his confidence that the device was '100 per cent efficient and accurate in the detection of deception' (*People v. Kenny* 167 Misc. 51, 52 (N.Y. Misc. 1938)). As Goldblatt notes, 'a better qualified expert could not have been found' (Goldblatt 1950). Not only was he the inventor of a new form of lie detector, the psychogalvanometer which measured electrical currents rather than heartbeat, he had two doctorates and an impressive history of study in 'Europe'.

The judge was convinced, remarking,

The lie detector is a decided step forward in legal procedure, which is merely an ascertainment of truth ... It seems to me that this pathometer and the technique by which it is used indicate a new and more scientific approach to the ascertainment of truth in legal investigations. (*People v. Kenny* 167 Misc. 51, 52 (N.Y. Misc. 1938), Colden J.)

It was reported that one of the jurors claimed that 'the machine was the deciding factor ... a wonderful thing' (Legal Chatter 1938). The juror went on to predict that in time, the lie detector would simplify the work of a jury; in this case however, 'it lengthened the work because we wasted so much time talking about the machine' (Legal Chatter 1938).

The *Kenny* case was the exception rather than the rule, although it was not the only exception. It was claimed by journalist Alva Johnston in 1944 that the lie detector had been used in 60,000 cases, establishing 'its uncanny power of penetrating guilty secrets' (Johnston 1944). The 1930s and 1940s reportedly saw an approach in the Chicago courts whereby lie detector evidence was used in probation hearings, (Johnston 1944) in the solution of 'bastardy cases', and admitted where agreed to by both parties (an approach strongly advocated for by Reid, a lie detector expert (Reid 1954)). Such agreement and stipulation involved waiver of rights to object to the admissibility of the evidence, and in one trial included the requirement that the court 'be required to further instruct the jury that they should not accept the test results or the examiner's opinion as conclusive on the issue before them' (Reid 1954).

A 1952 law review reports upon a Michigan case in which both the defendant and plaintiff agreed to take a lie detector test. The trial judge found for the defendant, commenting that the test was 'a definite aid ... in supporting what appeared to be the preponderance of evidence'. The appellate court concluded that there was no prejudicial error as the evidence favoured the defendant in any event (50 N.W. 2d 172 (Mich.1951) reported in 'The Lie Detector' 1952).

During this period, accounts can be found of use of lie detector testimony both in the criminal investigatory stage, and also in semi-judicial determinations outside of court. Kiraeofe reports in 1948 that use of lie detector testimony had largely been confined to pre-trial investigations as means of obtaining confessions and suggesting clues, and expresses the fear that, if admitted as evidence, 'complete credence would be placed in them' (Kiraeofe 1948). In 1949, Fioch, a clinical psychologist working with offenders noted the potentially 'dubious' findings of the lie detector in relation to particular individuals, commenting that it was therefore 'surprising' to find 'at least one state parole board ... making use of it as an aid in the determination of innocence or guilt for commutation recommendations' (Fioch 1949–50).

Goldblatt recounts that, where lie detectors were used to obtain admissions and confessions, the courts have rejected any objection to the use of the device as a means of inducing such admissions and confessions (Goldblatt 1950). Inbau highlights common mistakes made in lie detector test procedures during such investigatory stages: a

person might be unfit for the test due to extensive interrogation, or even physical abuse, beforehand; the examiner might be unqualified or unfit and in particular may be one who 'will feel impelled to make a definite diagnosis in practically every case' (Inbau 1949–50). Inbau reports on the risk that examiner reports are in many instances accepted at face value and upon the assumption that the technique 'produced results approximately perfection' (Inbau 1949–50). As a former director of Chicago Police Scientific Crime Laboratory, expert in interrogation techniques, advocate of the use of lies and deceit in police interrogation, and author of a book about lie detection and criminal investigations, there may have been some self-interest in Professor Inbau's concerns regarding unqualified and unsuitable examiners. However, his concern that students should be made to realise that lie detector technique was subject to limitations and should not be represented as infallible (Inbau 1949–50) has much resonance today.

## **Selected analogies between the early lie detector and machine learning**

### ***Prediction and categorisation of human behaviour***

Early lie detectors and modern machine learning tools are of course different things. They operate in different ways, analyse data differently and interact with human operators and subjects in different ways. When considering the criminal justice context however, fundamentally both technologies attempt to predict or categorise human behaviour based on assumptions about, and comparison with, the behaviour of others. Machine learning systems do this by predicting 'future action or behavior based on algorithmic identification of unexpected correlations within massive data sets that would not be detectable by human cognition' (Yeung 2018). Supplied with input data and a target function, machine learning generates a mathematical function, a pattern, that maps inputs to that target, with the aim that this function will also predict the target for new data. Yet when applied to human behaviour, the overarching concern remains that

the adaptive nature of complex social phenomena remains elusive even when a system is trained on unprecedented volumes of data. This means that the fundamental assumption that underlies any ML system i.e. that reality is governed by mathematical functions, does not necessarily hold for human society. (Hildebrandt 2018)

The polygraph is based upon the theory that psychological states associated with deception generally affect physiological responses in a consistent way which can then be measured, but there remains little basis for claims of extremely high accuracy: 'Although psychological states often associated with deception ... do tend to affect the physiological responses that the polygraph measures, these same states can arise in the absence of deception' (National Research Council 2003). Even in the over-optimistic early years, this was appreciated. Fioch highlights 'significant deviations in reactions to the test' which would make results somewhat dubious: from individuals who had no appreciation of the significance of falsification, 'antisocial types', 'pathological liars', those with 'circumscribed amnesia' who would deny the crime without demonstrating physiological change (Fioch 1949–50). Further concerns are raised by use of polygraph testing for preemployment screening as 'it involves inferences about further behavior on the basis of information about past behaviors that may be quite different' (National Research Council 2003).

## **Societal aims**

Overcoming perceived human inadequacies and fixing societal issues are motivations that can be detected in the development of both technologies within the criminal justice system. Machine learning tools are advocated and deployed as prioritisation tools and as forms of risk-based regulation and actuarial justice (Yeung 2018). Risk based approaches to the prioritisation of resources around the management and triaging of offenders and those at risk of offending have grown in importance in recent decades, (Yeung 2018) with the impact of austerity and resulting reduction in police officer numbers playing a significant role in the perceived need to work differently and prioritise effectively (Babuta and Oswald 2020).

This crime prevention policing role is linked to crime risk factors (such as hot spot policing (Farrington et al. 2003)) and actuarial methods to assess the 'future dangerousness' (Hyatt and Berk 2015) of an individual. Yeung describes the process as follows:

Individual and social phenomena are reconstructed as risk objects so that the focus of analysis is no longer the biographical individual but their risk profile, created by reconstructing fragments of individual identity by combining variables associated with different categories and levels of risk (Yeung 2018).

Despite their objective portrayal, machine learning tools are consistently *subjective*. They can process only the data provided to them, historical data which will contain elements of subjectivity. In the criminal justice context, tools making predictions based on historical offender data will be affected by past arrest history, force targeting decisions, social trends and prioritisation of certain offences (Oswald et al. 2018).

Polygraphs, too, reconstruct individuals based on fragments of data – physiological responses – and place them along a line between truthful and deceptive. Early machines were said to be more reliable than human witnesses, a more scientific way of ascertaining the 'truth' and even as a 'cure for crime' (William Marston quoted in Bunn 2012). The development of the lie detector was closely associated, as machine learning in policing is today, with a 'social work' and crime prevention role for the police, and a scientific approach applied towards 'the investigation and removal of social, economic, physical, mental and moral factors underlying crime' (August Vollmer quoted in Bunn 2012). In recent years, the lie detector has been resurrected as a 'therapeutic polygraphy' (Grubin and Madsen 2005) based on studies that concluded that the threat of the test is more likely to induce increased disclosures from monitored sex offenders because of fear of the result (Gannon et al. 2012).

Bunn argues that no lie detector examination takes place under 'objective' scientific conditions; the person is also subject to more covert scrutiny of their gestures, expressions, talkativeness and enthusiasm (Bunn 2012). Although a suspect could not be compelled to face the lie detector, 'it doesn't look good, however, for one who proclaims his innocence to refuse' (Johnston 1944). This supported, Bunn argues, the concept of the lie detector as an 'alternative legal system' – if you volunteered, you were innocent; if you didn't, you were guilty (Bunn 2012). By analogy, machine learning is now so embedded in many commercial processes that it is virtually impossible for a consumer to refuse to use it. Within policing and criminal justice, machine learning tools have been deployed throughout

the system, from investigatory and intelligence-generation, to sentencing and post-sentence monitoring. Due to the nature of policing and criminal justice, this is not generally regarded as an 'opt-in/out' process for individuals.

### ***Accuracy and the human in the loop***

A tendency can be detected in the operation of both technologies to down-play the role of the human, as operator and interpreter of the output, and as subject of the analysis. Hildebrandt points out, in relation to machine learning, that:

Machines do not experience light, temperature, sound or touch, nor do they have any understanding of natural language; they merely process data. Whereas such data may be a trace of, or a representation of light, temperature, sound, touch or text, it should not be confused with what it traces or represents. (Hildebrandt 2018)

Despite this, outputs can be represented as the truth or the whole picture, and the need for human interpretation, discretion and judgement reduced. Bunn reminds us that the lie detector is not only a machine – it is a complex array of techniques, concepts, procedures and symbols, (Bunn 2012) many of which are still subject to dispute and criticism today (Stanley 2018). Although diagnosis of deception was a subjective, interpretative human act, the use of graph paper with its peaks and troughs led people to believe that the machine 'could almost speak for itself, so obvious was the appearance of the lie' (Bunn 2012). At best, however 'the device can claim only to detect symptoms of emotions consistent with the examinee's belief in the truth of his or her answers' (Underwood 1995) and then only within a continuum of expected responses (Elton 2017). Hyperbolic assertions of accuracy rates of over 90% were common in early reporting and in scientific literature, (Bunn 2012) despite the difficulty of ascertaining the occasions on which deception was practised undetected or an innocent person was misreported guilty! Reid claims that 'the few mistakes that do occasionally occur are those in which a guilty person's lying is undetected, rather than those of an innocent person being reported guilty' but provides no evidence to back up this claim (Reid 1954).

In a comment that could have been written in respect of modern machine learning, Forkosch says:

Preconceived theories are relegated to the dung heap and the mathematical probability curve is invoked as a check. General laws, or probability curves, are plotted against the background of unending tests, to the end that the truth may become more and more certain. (Forkosch 1939)

With respect to Summers' tests with his 'psychogalvanometer', where his conclusions had resulted in the prosecution of two cases and non-prosecution and release of two others, Forkosch further comments that:

These experiments upon actual cases have been predicated upon the infallibility of Father Summers' conclusions for, if fallible in the slightest degree, it would be shocking to permit a life to be gambled upon the wheel of chance. Probability has given place to certainty, as indeed it must, if a *mechanical jury* is to be substituted for our unpredictable jury of twelve human and emotional people. (Forkosch 1939)

Forkosch concluded that the machine was neither infallible nor undebatable, but an instrument to 'diagnose hypothetical subjective occurrences' (Forkosch 1939). It should

be noted that Forkosch was a member of the New York bar, and the opening provocation of his article – ‘Shall there be a re-evaluation of our historic legal traditions culminating in an impersonal and mechanical court and jury?’ – suggests a kindred spirit to Justice McCoy in the *Frye* case, as does his conclusion that ‘our legal ship best lie in its harbor of accepted rules of evidence’.

Percentage accuracy rates are common also in literature relating to machine learning models, statistical risk assessment and machine based emotion detection. McStay warns against displacing the contextual nature of emotional life by a ‘veneer of metric-based certainty about what emotions are’, especially where used in automated systems to flag deception (McStay 2018). The AUC (area under the curve) statistic favoured to demonstrate a model’s predictive validity has been argued to be misleading and uninformative in the context of offender risk management due to the high margins of error often involved (Cooke and Michie 2014). Where interventions are put in place to prevent the event or action being predicted, percentage accuracy rates are blind to such interventions. Furthermore, such tools are designed to score or categorise an individual – low risk, high risk and so on – which can be as equally unnuanced and open to misinterpretation as the early lie detector graph. Other relevant uncoded factors are not taken into account. When dealing with humans, risk prediction accuracy is likely to be subject to some fundamental limit due to the importance of extrinsic external factors relevant to the specific individual (Hofman, Sharma, and Watts 2017). Hildebrandt sets out that ‘in real life situations ... future data can always disrupt the predictive accuracy of the hypothesis target function’ (Hildebrandt 2019). Forecasts, classifications or predictions produced by many existing algorithmic tools are probabilities (that the person or situation in question has a similarity to people or situations in the past). ‘But they appear at times to be presented as something more: a prediction of reoffending becomes a “risk” of reoffending and thus the risk if, say, a person is given parole’ (Oswald 2018).

The polygraph was also presented as something more. Polygraph inventor Keeler emphasises that there are no instruments recording bodily changes that deserve the name ‘lie-detector’ – a diagnosis of deception is made from tangible symptoms by an examiner ‘using whatever mechanical aids he has at his disposal’ (Keeler 1934). The fact that there was a ‘human in the loop’, the operator, was used to support the case for admission of lie detector testimony in court proceedings. Goldberg argues that the risk that an innocent person might be convicted is minimised because of the burden of ‘showing the lie’ is placed upon the operator, who is subject to cross-examination. Furthermore, the lie detector results will not displace the jury who, it is argued, will weigh the test results and opinion of the operator like any other evidence (Goldberg 1949).

This assertion was countered somewhat by a piece of contemporary research. After the Kenny case, Forkosch was able to poll the entire jury. Although none admitted to basing their decision solely on the lie detector testimony, six jurors thought that the testimony was ‘*conclusive proof*’ of the defendant’s guilt or innocence, and five agreed that they had accepted it ‘*without question*’ (an early example of automation bias perhaps). As Forkosch remarks, ‘what has become of Judge Colden’s closing words, “The testimony will be received and the jury permitted to evaluate it” when five categorically accept it without hesitation and do not “evaluate” it in the light of the entire testimony?’ (Forkosch 1939).

## Opacity

The early lie detector was the original black box. Its ‘magic’ – its theatricality, opacity and intimidating character – benefited those who would promote its use (Lewis 1939). Bunn explains that:

Expository articles often included a photograph of the enigmatic instrument, a depiction that explained yet mystified at the same time. Here was a gadget fabricated from reassuringly complex components, all of which were encased within a scientific-looking ‘black box’. By describing the instrument thus, however, the question what exactly it was fabricated from remained unanswered. (Bunn 2012)

Much has been written about the opacity of machine learning tools, (Burrell 2016; Wachter, Mittelstadt, and Russell 2017) and the need for ‘transparency’ and ‘explanations’ in order for the individual to judge for herself the fairness of the system (although this approach has been strongly criticised (Edwards and Veale 2017)). This contrasts with the approach taken with early lie detectors where the ‘critical link’ was the operator (Underwood 1995). This man (because inevitably it was a man, dressed in a white lab coat or dark suit) was the keeper of the machine, the priest who could interpret the oracle’s musings. But there was no official regulation of lie detector practitioners or their methods. Johnston states that ‘any man, woman or child can set up as a lie-detector expert, and a crooked official would have no difficulty in finding just exactly the kind of expert that he needed’. (Johnston 1944) As a result, Keeler argued for the licensing of ‘medico-legal’ technicians (Keeler 1934).

## Reformist legal realism, the lie detector growing up and regulatory lessons for machine learning

### Reformist legal realism

It can be speculated that Justice Colden, who presided over the *Kenny* case, was a reforming legal realist. He favoured a forward-thinking version of jurisprudence in which ‘new concepts must beat down the crystalized resistance of the legally trained mind that always seeks precedent before the new is accepted into the law’ (Colden J quoting Steinbrink J in *Beuschal v. Manowitz* 151 Misc. 899 (N.Y. Misc. 1934)). To Colden, the ‘truth’ equated with the facts, and the lie detector, with its underlying behaviourist approach, was a better way of ascertaining that truth, especially as he had been convinced that the lie detector represented ‘a science from which varying inferences may not be drawn’ (*People v. Kenny* 167 Misc. 51, 52 (N.Y. Misc. 1938)). Rules of evidence and cross-examination had been used for hundreds of years but only ‘for lack of any better approach’. Perhaps he would have supported Cohen’s view that ‘it is through the union of objective legal science and a critical theory of social values that our understanding of the human significance of law will be enriched’, (Cohen 1935) and Llewellyn’s view of law as a means to social end, not an end in itself (Llewellyn 1931). This is not to say however that legal rules and principles of justice were discounted. As we have seen above, ‘celebrity’ inventors of the lie detector were themselves involved in campaigns to change those rules of evidence to admit polygraph testimony on the basis of arguments around fairness, human unreliability and scientific progress, for instance Reid’s support of the admission of lie detector evidence upon agreement and stipulation, and Marston’s

evidence in the *Frye* case itself. They published articles in legal and criminology academic journals, with press reports of the machine as ‘witness’, or returning verdicts of innocence or guilt, contributing to their cause (see examples on page 147 of Bunn 2012).

Today’s machine learning could also be said to support ‘*reformist legal realism*’, an approach that aims to advance productivity, efficiency and the human condition (although often narrowly defined) by an emphasis on empirical evidence and scientific methods. The creators of a machine learning classification tool that analysed textual content from existing ECHR case law, and was said to predict the court’s decisions with 79% accuracy, argue that predictive accuracy depended on non-legal facts of the case rather than on legal arguments, thus giving some support to the theory of legal realism (Aletras et al. 2016). (Aletras et al. was critiqued for its behaviourist approach in Pasquale and Cashwell 2018) So if a legal realist approach is the one that actually exists, why not adapt the legal process to incorporate such AI judges? (Kleinberg et al. 2017).

Machine learning tools as means of prioritisation, risk-based regulation and detecting rule violations have been discussed above. To take another of many possible examples, it has been argued that an algorithm could provide a more specific and transparent way of detecting *and evidencing* unlawful discrimination, compared with attempting to analyse ambiguous human decision-making (Kleinberg et al. 2019). The authors’ proposals, followed through to their logical conclusion, would appear to have an outcome that early lie detector proponents would likely have supported, and which would transform the relevant decision-making environment itself: i.e. *it may be possible to detect discrimination statistically via an algorithm; therefore, the algorithm must therefore make the decision – and only the algorithm – resulting in significant alteration of decision-making processes and the requirement for a new legal framework to support the prevention of discrimination through the use of algorithms.*

### ***The lie detector growing up***

Legal rules of evidence and due process held back the early lie detector proponents. Yet their reformist legal realism ambitions were not totally thwarted. Limitations on use of lie detector testimony in court did not prevent use of the lie detector *outside court* forging ahead, in particular for the assessment of evidence, vetting potential employees, and in fraud investigations.

Although the results of polygraph examinations may not accompany testimony, they have been used in everyday police work [in the US] since Vollmer introduced them in Berkeley 1921. The utility of using polygraphy in police work lies in the procedure’s ability to induce truthful incriminating statements from suspects. Jurors judging the reliability of the confession would not be privy to this tactic or exposed to the prejudicial effect of learning that a machine had been used in determination of truth. Thus, in the courtroom, polygraphy is kept behind the scene. (Weiss, Watson, and Xuan 2014)

The 1960s and 1970s in the US saw widespread use of the lie detector, not only in police investigations, but for pre-employment screening in both public and private sectors, with other countries including Japan, China, Israel and Korea starting their own programmes (Grubin and Madsen 2005). Increasing concern about potential abuses following a report from the US House of Representatives Office of Technology Assessment, (US Congress 1983) led to the 1988 Employee Polygraph Protection Act: ‘*An Act to prevent the*



*denial of employment opportunities by prohibiting the use of lie detectors by employers involved in or affecting interstate commerce*’. Introduced by President Reagan to control and regulate the use of polygraph machines by employers in private practice, it had both significant limitations and exemptions for its use. While the Act sought to prohibit the use of polygraphs, its many exemptions in reality meant that the most controversial uses were still permissible. This included polygraph use by Governmental Agencies to test potential employees, any expert or consultant involved with counter intelligence and use by employers to test with respect to drug offences or drug investigations. It also had no impact on ongoing criminal justice uses.

### ***Regulatory lessons for machine learning***

The utilitarian approach to the polygraph within criminal justice, irrespective of its reliability or validity, could be said to be epitomised by this quote from Warner, a special agent of the FBI. ‘Confessions, admissions ... and additional information of investigative value gained through ... testing come about due to the *utility* of the polygraph and the determination of the examiner, irrespective of the instrument’s reliability or validity’ (Warner 2005). Confessions come about following the ‘mere suggestion’ of testing. Warner asks

Your child is abducted and investigators come to you and say, ‘We have a suspect who we will be giving a polygraph to’. Would you be so bold as to reply, ‘The polygraph technique is unreasonable, find my child another way?’. (Warner 2005)

This question demonstrates the temptation of the polygraph, and other technologies that promise important insights. Most parents would not answer in the affirmative, in case the polygraph test *might* be right. But from the perspective of the rule of law, it is the wrong question. It suggests an alternative legal system that is not accessible, clear or predictable, and one which may infringe fundamental rights without clear and valid justification. There are effectively two systems running in parallel: one in relation to criminal court proceedings, in which expert testimony is admitted only if based upon a scientifically valid foundation relevant to the issue at hand, and one which runs parallel to the court system without such constraints, based on realist and utilitarian principles.

We risk a similar situation arising in respect of modern machine learning. Indeed, in the challenge by Eric Loomis to the use of the COMPAS risk assessment tool in his sentencing, the Wisconsin Supreme Court stated that the sentencing court ‘would have imposed the exact same sentence without it. Accordingly, we determine that the circuit court’s consideration of COMPAS in this case did not violate Loomis’s due process rights’, (*State v. Loomis* 881 N.W.2d 749 (Wis. 2016)) creating a ‘troubling paradox’ (Villasenor and Foggo 2019). Why use the assessment if its relevance or necessity cannot be demonstrated? Does this statement suggest the court acting in the same way as the parent of the abducted child? *It might work and so we might be missing out if we exclude it.*

The use of machine learning, especially when backed by commercial interests, is likely to expand to fill whatever gap is available. As well as for purely commercial decisions, we already see private sector predictive tools and emotion or deception detection marketed for use in hiring decisions, fraud detection, immigration and other screening, decisions that come with high-stakes for individuals (Sanchez-Monedero 2019). In terms of

governance and regulation, focus to date has been on data protection, individual rights, privacy and ethics. Guided by the early polygraph experience, and giving consideration to the way that admission of scientific findings and expert testimony in court is assessed, this paper proposes the development and application of appropriate '*scientific validity*' and '*relevance*' standards for AI and machine learning. These would be constructed for specific criminal justice contexts (investigative, offender management, risk-assessment and so on), and to include the presentation of results (to ensure these are not presented as 'something more'), thus moving to red-line restrictions to prevent AI morphing problematically. The author has previously called for the development of a framework around the use of automated facial recognition as a trigger for intervention and in an evidential concept, bearing in mind the officer's over-arching decision-making prerogative (Kotsoglou and Oswald 2020). Such standards could augment the work of existing oversight bodies, and complement Hildebrandt's proposal for preregistration of machine learning research design (Hildebrandt 2018) and Nemitz's call for a precautionary legal framework around artificial intelligence, and the generalisation for AI of regulatory principles found in specific bodies of law, (Nemitz 2018) namely here those relating to the admission of expert evidence. Much could be learned from issues that have arisen from the use of forms of scientific evidence, such as forensics, in particular understanding of caveats, interpretations and methodologies, (McCartney 2019) measurements of reliability and validity, (Stern, Cuellar, and Kaye 2019) and communication of uncertainty (Mejia et al. 2019).

## Conclusion

The polygraph machine cannot 'detect lies'; it merely records bodily changes, and human interpretation of its results is required for any diagnosis of deception. Neither can a machine learning tool independently 'predict' risk or a person's future; rather real-world experience is reduced to variables and an algorithm trained to detect patterns or similarities based on probabilities. The interpretation of the output as a prediction or contribution to a 'risk' assessment should be a human one.

This paper has drawn analogies between two technologies in the 'twilight zone' of their development, utilisation and acceptance. When considering how we should respond to the increasing use of machine learning within criminal justice, and elsewhere, lessons can be learned from the early lie detector: from its deployment in a 'quest for certainty' (Hildebrandt 2020) by a society concerned about human inadequacies and a crime prevention agenda; from the attempt to individualise the machine's conclusions based on comparison with the behaviour of others; from the downplaying of the role of human interpretation. Both technologies support a reforming legal realist approach, thus raising questions for key principles of the rule of law. The lie detector experience shows us that, despite a ban on its use in some contexts, it will continue to be deployed in others due to a belief in its utility, whether or not it 'works'. If we are not to find ourselves in a similar position with respect to machine learning, we must develop an oversight framework based upon a combination of scientific validity and relevance standards, fairness principles and the role of the legitimate human decision-maker. Father Brown wisely said that 'no machine can lie ... nor can it tell the truth;' (Chesterton 1914) if we remember that machines are just that – machines – and cannot make the decisions which in our legal systems should be reserved for human judgement and discretion, then we will be half way there.

## Note

1. *FAZLIYSKI v. BULGARIA* (Application no. 40908/05), 16 April 2013: breach of article 6(1). Nothing wrong with a polygraph test being carried out by an expert but decision-making panel considered themselves bound by it and refused to scrutinise it in any way; *TYAGUNOVA v. RUSSIA* (Application no. 19433/07) 31 July 2012. Breach of Articles 3 and 8 included failure to follow up suspect's polygraph test results, which showed that the latter had 'guilty knowledge' in connection with the applicant's rape; *BRAGADIREANU v. ROMANIA* (Application no. 22088/04) 6 December 2007: applicant took polygraph test without a lawyer present. His conviction was not based solely on this evidence. Therefore, although regrettable, such a circumstance was not in itself sufficient for the finding of a violation of Article 6.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## References

- Aletras, N., D. Tsarapatsanis, D. Preotjiuc-Pietro D, and V. Lampos. 2016. "Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective." *PeerJ Computer Science* 2: e93.
- Babuta, Alexander, and Marion Oswald. 2020. "Data Analytics and Algorithms in Policing in England and Wales: Towards a New Policy Framework." *RUSI Occasional Paper*.
- Boeche v. State* 151 Neb. 368, 37 N.W. 2d 593 (1949).
- Bunn, Geoffrey C. 2012. *The Truth Machine: A Social History of the Lie Detector*. Baltimore, MD: John Hopkins University Press.
- Burrell, J. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3: 1–12.
- Cahillane, L. 2016. "The Use of History in Law: Avoiding the Pitfalls." In *Legal Research Methods: Principles and Practicalities*, edited by Laura Cahillane and Jennifer Schweppe, 57–69. Dublin: Clarus Press.
- Chesterton, G. K. 1914. *The Mistake of the Machine*.
- Christin, Angèle. 2017. "The Mistrials of Algorithmic Sentencing." *Logic*.
- Cohen, Felix. 1935. "Transcendental Nonsense and the Functional Approach." *Columbia Law Review* 35: 809–849.
- Cooke, D. J., and C. Michie. 2014. "The Generalizability of the Risk Matrix 2000: On Model Shrinkage and the Misinterpretation of the Area Under the Curve." *Journal of Threat Assessment and Management* 1 (1): 42–55.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)
- Edwards, Lilian, and Michael Veale. 2017. "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For." *Duke Law & Technology Review* 16: 18.
- Elton, Jacqueline. 2017. "The Polygraph in the English Courts: A Creeping Inevitability or a Step too Far?" *JCL* 81: 66–79.
- Employee Polygraph Protection Act. 1988. Pub. L. No. 100–347, 102 Stat. 646.
- European Commission's High Level Expert Group on Artificial Intelligence. 2018. "A Definition of AI: Main Capabilities and Scientific Disciplines".
- Farrington, David P., Doris Layton MacKenzie, Lawrence W. Sherman, and Brandon C. Welsh. 2003. *Evidence-Based Crime Prevention*. London: Routledge.
- Fioch, Maurice. 1949–50. "Limitations of the Lie Detector." *J. Crim. L. & Criminology* 40: 651.
- Forkosch, Morris D. 1939. "The Lie Detector and the Courts." *New York University Law Quarterly Review* 1 (13): 202–231.
- Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

- Gannon, Theresa A., Jane Wood, Afroditi Pina, Eduardo Vasquez, and Iain Fraser. 2012. "The Evaluation of the Mandatory Polygraph Pilot." *Ministry of Justice Research Series* 14/12: 1–69.
- Goldberg, Benjamin. 1949. 'In Re Lie Detector' 2 S. C. L. Q. 266.
- Goldblatt, Samuel A. 1950. 'Evidence – Admissibility of Results of Lie Detector Tests' 1950 WASH U.L.Q. 255.
- Grubin, Don, and Lars Madsen. 2005. "Lie Detection and the Polygraph: A Historical Review." *Journal of Forensic Psychiatry & Psychology* 16 (2): 357–369.
- Hao, Karen. 2019. "We Analyzed 16,625 Paper to Figure out where AI is Headed Next." *MIT Technology Review*.
- Hildebrandt, Mireille. 2018. "Preregistration of Machine Learning Research Design." In *Being Profiled: Cogitas Ergo Sum*, edited by E. Emre Bayamlioğlu, I. Irina Baraliuc, and L. Liisa Janssens, 102–105. Amsterdam: Amsterdam University Press.
- Hildebrandt, Mireille. 2019. "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning." *Theoretical Inquiries of Law* 19 (1): 83–121.
- Hildebrandt, Mireille. 2020. "The Artificial Intelligence of European Union Law." *German Law Journal* 21 (1): 74–79.
- Hofman, J. M., A. Sharma, and D. J. Watts. 2017. "Prediction and Explanation in Social Systems." *Science* 355: 486–488.
- Hutson, Matthew. 2018. "AI Researchers Allege That Machine Learning is Alchemy." *Science* 360 (6388): 861.
- Hyatt, Jordan, and Richard Berk. 2015. "Machine Learning Forecasts of Risk to Inform Sentencing Decisions." *Federal Sentencing Reporter* 27: 222–228.
- Inbau, Fred E. 1949–50. "Some Avoidable Lie Detector Mistakes." *J. Crim. L. & Criminology* 40: 791.
- Johnston, Alva. 1944. "The Magic Lie Detector: The Saga of Leonarde Keeler." V.A., Editor and Compiler Leonard. Academy Lectures on Lie Detection (1957) Original article dated 1944.
- Keeler, Leonarde. 1934. "Debunking the Lie Detector." *Am. Inst. Crim. L. & Criminology* 25: 153.
- Kiraef, W. C. 1948. "Lie Detector Tests as Evidence." *JAG J* 5: 5–7.
- Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. 2017. "Human Decisions and Machine Predictions." NBER Working Paper no. 23180. National Bureau of Economic Research, Cambridge, MA.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein. 2019. "Discrimination in the Age of Algorithms." (February 5, 2019). Available at SSRN: <https://ssrn.com/abstract=3329669> or <http://dx.doi.org/10.2139/ssrn.3329669>.
- Kotsoglou, Kyriakos N., and Marion Oswald. 2020. "The Long arm of the Algorithm? Automated Facial Recognition as Evidence and Trigger for Police Intervention." *Forensic Science International: Synergy* 2: 86–89.
- Leetaru, Kalev. 2019. "How Data Scientists Turned Against Statistics." *Forbes*.
- Legalizing the Lie Detector. 1938. 1 *Legal Chatter* 41.
- Lepore, Jill. 2014–15. "On Evidence: Proving Frye as a Matter of Law, Science, and History." *The Yale Law Journal* 124 (4): 882–1345.
- Lewis, A. A. 1939. "Looking for an Honest Man." *Scientific Monthly* 49: 267–275.
- The Lie Detector. 1952. 1 *Drake L. Rev.* 50.
- Llewellyn, Karl N. 1931. "Some Realism About Realism." *Harvard Law Review* 44: 1222.
- McCartney, Carole. 2019. "Streamlined Forensic Reporting: Rhetoric and Reality." *Forensic Science International: Synergy* 1: 83–85.
- McQuillan, Dan. 2018. "People's Councils for Ethical Machine Learning." *Social Media & Society* 4 (2): 1–10.
- McStay, Andrew. 2018. *Emotional AI: The Rise of Empathic Media*. London: Sage.
- Mejia, Robin, Maria Cuellar, Dana Delger, and Bill Eddy. 2019. "What Does a Match Mean? A Framework for Understanding Forensic Comparisons." *Significance* 16 (2): 25–28.
- National Research Council. 2003. *The Polygraph and Lie Detection*. Washington, DC: The National Academies Press.
- Naughton, John. 2019. "Don't Believe the Hype: The Media are Unwittingly Selling us an AI Fantasy." *The Guardian*.

- Nemitz, Paul. 2018. "Constitutional Democracy and Technology in the Age of Artificial Intelligence." *Phil. Trans. R. Soc. A*.
- Novkov, Julie. 2011. "Legal Archaeology." *Political Research Quarterly* 64 (2): 348–361.
- Oswald, Marion. 2018. "Algorithmic-assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative law Rules Governing Discretionary Power." *Philosophical Transactions of the Royal Society A* 376 (2128).
- Oswald, Marion, Jamie Grace, Sheena Urwin, and Geoffrey C. Barnes. 2018. "Algorithmic Risk Assessment Policing Models: Lessons From the Durham HART Model and 'Experimental' Proportionality." *Information & Communications Technology Law* 27 (2): 223–250.
- Pasquale, Frank and G. Cashwell. 2018. "Prediction, Persuasion, and the Jurisprudence of Behaviourism." *University of Toronto Law Journal* 68: 63–81.
- People v. Kenny* 167 Misc. 51, 52 (N.Y. Misc. 1938).
- Reid, John E. 1954. "The Lie Detector in Court." *DePaul L. Rev.* 4: 31.
- Richardson, Rashida, Jason Schultz, and Kate Crawford. 2019. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *New York University Law Review Online*, Forthcoming.
- Sanchez-Monedero, Javier. 2019. "The Politics of Deceptive Borders: 'Biomarkers of Deceit' and the Case of iBorderCtrl." arXiv:1911.09156.
- Shneiderman, Ben. 2019. "What Alchemy and Astrology can Teach Artificial Intelligence Researchers." *The Conversation*.
- Simpson, A W Brian. 1995. *Leading Cases in the Common Law*. Oxford: Clarendon Press.
- Stanley, Jay. 2018. "How Lie Detectors Enable Racial Bias." *American Civil Liberties Union*.
- State v. Bohner* 210 Wis. 651, 246 N.W. 314 (1933).
- State v. Cole* 354 Mo. 181, 188 S.W. 2d 43 (1945).
- State v. Loomis* 881 N.W.2d 749 (Wis. 2016).
- State v. Lowry* 163 Kan. 622, 185 P. 2d 147 (1947).
- Stern, Hal S., Maria Cuellar, and David Kaye. 2019. "Reliability and Validity of Forensic Science Evidence." *Significance* 16 (2): 21–24.
- Syll, Lar. 2019. "Machine Learning — Getting Results that are Completely Wrong." *Real-World Economics Review Blog* <https://rwer.wordpress.com/2019/02/20/machine-learning-getting-results-that-are-completely-wrong/>.
- Synnott, John, David Dietzel, and Maria Ioannou. 2015. "A Review of the Polygraph: History, Methodology and Current Status." *Crime Psychology Review* 1 (1): 59–83.
- Underwood, Richard H. 1995. "Truth Verifiers: From the Hot Iron to the Lie Detector." *Ky. L.J.* 84: 597.
- US Congress. 1983. *Scientific Validity of Polygraph Testing: A Research Review and Evaluation – A Technical Memorandum*. Washington, D.C.: Office of Technology Assessment, OTA-TM-H-15.
- Villasenor, John, and Virginia Foggo. 2019. "Algorithms and Sentencing: What does Due Process Require?" *Brookings TechTank*.
- Wachter S, B. Mittelstadt, and C. Russell. 2017. "Counterfactual Explanations Without Opening the Black box: Automated Decisions and the GDPR." *Harvard Journal of Law & Technology* 31 (2): 841.
- Warner, William J. 2005. "Polygraph Testing: A Utilitarian Tool." *FBI L. Enforcement Bull* 74: 10.
- Weiss, Kenneth J., Clarence Watson, and Yan Xuan. 2014. "Frye's Backstory: A Tale of Murder, a Retracted Confession, and Scientific Hubris." *Journal of the American Academy of Psychiatry and the Law Online* 42: 226–233.
- Yeung, Karen. 2018. "Algorithmic Regulation: A Critical Interrogation." *Regulation & Governance* 12: 505–523.