Nordic Journal of Studies in Policing

UNIVERSITETSFORLAGET

RESEARCH PUBLICATION

Volume 8, No. 2-2021, p. 1–19
ISSN online: 2703-7045
DOI: https://doi.org/10.18261/issn.2703-7045-2021-02-02

AI, Trustworthiness, and the Digital Dirty Harry Problem

Jens Erik Paulsen
Professor, The Norwegian Police University College
Jens.Erik.Paulsen@phs.no
https://orcid.org/0000-0002-1988-3902

Abstract
Artificial Intelligence (AI) has been a game changer on many fronts. For the police it offers new ways of carrying out investigation, surveillance, crime prevention and order maintenance. Questions have been raised about the trustworthiness of some innovative AI-driven applications. Under which circumstances and to what extent should the police be permitted to use emergent technology, i.e. use 'dirty' means in order to reach good ends? In this article, this problem is illustrated by a discussion of two emergent technologies, and possible criteria and test regimes for establishing trustworthiness are suggested towards the end of the article.

Keywords
Artificial intelligence, emergent technology, ethics, facial recognition, phenotyping, social robustness

1. Introduction
During the past decade, big data, fast processors, ingenious programming, and extensive networking capabilities have paved the way for major advances within artificial intelligence (AI). Some people fear the prospects of a 'brave new world' unleashed by AI, whereas others welcome its ability to perform complex and tedious tasks, mitigate socially induced bias, and produce useful knowledge through big data analyses.

In the context of policing, some AI-driven systems have been hotly debated, particularly within surveillance and crime prediction (Kaufmann, 2018; Richardson et al., 2019; Zuboff, 2019, pp. 387–388). In addition, there are emergent systems and techniques that generate high hopes and great fears long before they have gained operational status. Approval processes take time, as it is often difficult to establish reliability, legitimacy and trustworthiness in the context of AI. Recently, several guidelines have been proposed to assure the production of trustworthy AI systems (HLEG-AI, 2019; NENT, 2019; New Zealand Government, 2020), but, increasingly, advanced tech innovation takes place in the less regulated private sector (Gerstein, 2019, p. 56). Neither governments nor the police dictate technological innovation within the *Industry 4.0* paradigm (Rainnie & Dean, 2020). As governmental approval processes tend to be slow, emerging technologies may be applied for recreational or criminal purposes before they are even considered for law-enforcement.

As cities and societies become 'smarter', cybercrime increasingly[1] poses a threat to crucial public services – as well as to businesses and individuals. Crime prevention and criminal investigations are likely to fail if the police lack understanding of, and the capabilities to counter, novel forms of cybercrime. More intensive old-school policing cannot bridge this gap. Smart city policing demands new technological measures and methods.

As professionals, the police are supposed to use approved equipment in accordance with legal and moral standards – even if novel technologies that could potentially help to solve or prevent crime exist. Still, exceptions to the rules are imaginable. If responsible police officers are pursuing a noble cause, and the standard operating procedures prove inadequate, why not try alternative methods? This is what Hartmann (2018, p. 162) refers to as 'grey zone creativity'. After all, as is the case in most professions, there is a gap between ideal performance and how policing is actually done (Hartmann, 2018, p. 161). The temptation to venture into the grey zone may become even stronger if innovative applications are both accessible and affordable to *individuals* on both sides of the law. This sets the scene for a new heroic villain – the *Digital* Dirty Harry, asking the same question as the original Dirty Harry: *are dirty means appropriate for reaching good ends* (Klockars, 2005, p. 582)?[2]

In this article, I revisit the Dirty Harry problem in some high-tech scenarios. The aim is to analyse and clarify technological and moral issues within the grey zone. Are there circumstances in which the police can be permitted to short-cut the slower processes of technological implementation assessment and approval?[3] Before reaching the Digital Dirty Harry problem proper, the concept of artificial intelligence is explored, as well as the concepts of accountability, reliability, and trustworthiness.

2. Artificial intelligence

Artificial intelligence has been a hot topic since the advent of electronic computers, but its development has hit several standstills, or 'winters' (Broussard, 2018, kindle loc 1747). With today's dramatically improved processor speeds, storage capacity, and connectivity (Marcus & Davis, 2019, p. 10), AI applications benefit every smart phone user, for instance through optical character recognition, computer vision, and language translation. In some areas, AI has also surpassed the performance of human specialists, for instance in interpreting radiographic images (Coccia, 2020). However, AI sometimes leads one astray (Nguyen et al., 2015), or it 'hallucinates' (Marcus & Davis, 2019). Algorithms fatally overrode pilot input in the Boeing 737 Max (Mongan & Kohli, 2020). The Twitter robot Tay quickly turned racist (Hannon, 2018). Self-driving cars have been crashing (Marcus & Davis, 2019, p. 19), and US congress representatives have been identified as criminals (Levin, 2018), to mention just a few examples. Morally troublesome military AI applications exist, too, like large and small-scale kill drones where human operators, from a distance, simply consent to or reject robotic tactical solutions (Allinson, 2015). People presumably in the know, like Stephen Hawking

1. See for instance https://www.interpol.int/en/News-and-Events/News/2020/INTERPOL-report-shows-alarming-rate-of-cyberattacks-during-COVID-19.
2. In the 1971 movie, Inspector 'Dirty' Harry Callahan tortured kidnapper Scorpio as a last-ditch effort to save a young girl. In the modern setting, we must imagine Harry's descendants pursuing morally good ends with morally dirty, digital means, or, bluntly refusing to acknowledge the moral aspect of technology, subscribing to what Drengson (1982) labelled technological anarchism.
3. The sting of the question may be softened by current reinterpretation of 'privacy', 'freedom' and 'control' (Hoofnagle et al., 2019).

and Elon Musk (Schneier, 2018, p. 86) have expressed concerns over artificial intelligence and the prospect of AI getting out of hand – reaching 'singularity' and turning itself against humanity. What is AI?

The concept 'artificial intelligence' is fuzzy, as it refers to different simulated cognitive abilities, to strategies, and to systems. It has been said to mimic natural intelligence (Campolo & Crawford, 2019, p. 3), it has been characterised as 'super intelligent' (Bostrom, 2014), and as a *different* type of intelligence, a kind of 'idiot savant' (Marcus & Davis, 2019, p. 64). A High-Level Expert Group on Artificial Intelligence appointed by the European Commission states that AI 'refers to *systems* that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals' (HLEG-AI, 2019, p. 1). Put differently, AI systems *register* some state of affairs and *do* something on this basis. In its simplest form, AI operates like a thermostat that senses temperature change and compensates by switching on or off a heater. The Expert Group, however, has more advanced devices in mind:

> AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications) (HLEG-AI, 2019, p.1).

The AI systems mentioned here are 'narrow' in the sense that they perform specific tasks in more or less known environments (Marcus & Davis, 2019, p. 13), typically providing assistance for planning and reasoning purposes, or for performing searches and knowledge representation (Marcus & Davis, 2019, p. 45).[4] Simple AI systems often contain hand-coded algorithms. In principle, an automatic traffic control system performs scripts like: *'IF measured speed <80-90 km/h> THEN BEGIN RecogniseNumberPlate; RetrieveOwnerName from DMV-database; SendFine1000ToOwner; END'*. More advanced systems have *machine learning* capability, that is, they make 'educated guesses based on data' (Broussard, 2018, kindle loc 1791), improving system performance with experience.

There are three types of machine learning (Russell & Norvig, 2016). First, the learning may be *supervised* by a human teacher, who decides whether the computer system identifies the 'example inputs' correctly or not: 'Yes, this is a photo of a dog!' or 'No, this is not a dog! (but a cat).' The supervisor teaches the machine a general rule that maps inputs (here: a dog image) to the correct output (the label 'dog'). In *unsupervised* machine learning no labels (e.g. 'dog') are provided for the learning algorithm. The algorithm simply judges whether an image resembles another image through identifying common features or 'edges' in the input photos. Unsupervised learning may thus reveal patterns unimagined by the supervisor. Large training databases are usually needed in order to identify the relevant features reliably.

Reinforcement machine learning is favoured when computers interact with dynamic environments or operate in environments containing sparse feedback. Such AI systems may navigate their 'problem space' towards their goal without clearly defined 'domain selectors', i.e. what inputs to expect and which feedback to receive.

4. *General* AI (GAI) implies understanding, flexibility, and common sense and is beyond the scope of this article. GAI would require 'an immense amount of foundational progress—not just more of the same sort of thing that's been accomplished in the last few years, but […] something entirely different' (Marcus & Davis, 2019, p. 4).

If the task is to learn chess, the problem space is well-defined and predictable. The different 'states' of the game can be fully described at any moment, and the number of legal actions that will either improve or worsen the state are limited. If the task is to drive a car in heavy traffic and bad weather, the problem space is dynamic and less well-defined. In both cases, reinforcement learning takes place as learning-by-doing, gaining experience by erring or breaking things (preferably in simulations). More formally, reinforcement learning consists in exploring a vast number of sequences of state/action pairs, linking these to probable rewards, until a best course of action is found. Since this procedure can be repeated continuously at blistering speeds without fatigue, AI systems can surpass human intelligence in some contexts. Alpha zero can beat human grand masters at chess most of the time, while self-driving cars generally drive more safely, although they can still make amateurish mistakes. Massive artificial experience can trump human understanding.

Deep learning AI utilises 'neural' networks consisting of 'layers' representing the different features located between the input and the output of the system. The layers are interconnected through nodes or so-called *neurons* (Marcus & Davis, 2019, pp. 48–49), hence the name. The connections between the nodes are referred to as *weights*, i.e. numbers describing the degree of correlation between the nodes. The depth of deep learning is judged by its number of layers. Deep learning neural networks are particularly useful in reinforcement learning when both the possible states and proper actions are hard to define in advance.

Through extensive training on samples from the state/action space, neural networks learn to map states to values. The network not only considers immediate gains, but also rewards to be had later in the simulation sequence. The neural network thus approximates a value function, resembling crude utilitarian calculus. Through fine-tuning correlations between multitudes of features over time, the system arrives at the optimal response to various inputs. This is how it learns to navigate seemingly chaotic environments.

Deep learning can replace work-intensive human computer programming and outperform older machine-learning techniques in complex tasks such as speech recognition or earthquake aftershocks prediction (Marcus & Davis, 2019, p. 53). However, given its constant back and forth (i.e. learning), its corrections of millions of biases and weights in order to find the optimal paths ('local minima'), it becomes humanly impossible to account for all the details of how such systems arrive at their solutions. Their domain selectors might not even make much sense from a human point of view. Still, deep learning is basically a matter of identifying features (edges and simple figures) and establishing connections between them. These are the building bricks of the magic of facial recognition, gait recognition, self-driving cars or advanced predictive policing. But how can we perform responsible technology implementation assessments when it comes to deep learning systems? Is deep understanding of deep learning processes required in order to *trust* such innovative technology?

3. Accountability and trustworthiness

When AI systems work as advertised, it seems irrational to return to old ways, e.g. radioing DMV registers instead of performing automatic number-plate recognition (ANPR). Likewise, toiling through numerous manual SQL queries seems absurd if Palantir can perform the same task instantaneously. Some AI systems simply enhance workplace efficiency. Other systems make bold promises about brand new possibilities. The latter are sometimes referred to as 'emergent' technologies.

According to Rotolo, Hicks, and Martin (2015), emergent technology can be identified by five key attributes: (i) its radical novelty, (ii) a relatively fast growth, (iii) coherence, (iv)

prominent impact, and (v) uncertainty and ambiguity. Several new deep learning applications qualify as emergent technology, both because it is difficult to account for their inner workings, and because their capabilities are not fully explored.[5] Granted, the (technical) internal reliability of the AI may develop swiftly through training but establishing *simulation* reliability is not enough to make a product *operationally* trustworthy. NASA's technology readiness level (TRL) classification system is often referred to in this context. It assesses the level of maturity, running from stage 1: 'technology research' to stage 10: 'proven operation' (Straub, 2015). For academic researchers or private sector innovators the critical phase for tech innovation occurs in the stage 4–6 range, comprising the steps (4) 'Technology demonstration', via (5) 'Conceptual design and prototype demonstration' to (6) 'Preliminary design and prototype validation'. As this is where most developing projects come to a halt, this area is sometimes referred to as the 'Valley of Death'. Before reaching stage 7 the technology is typically dependent on external entrepreneur seed funding or so-called 'angel investors' (Murphy & Edwards, 2003). In the absence of operational testing, some form of *hype* might be required to push the product through the Valley of Death, towards the stage where it is seriously considered for further testing and implementation.

Implementation assessment matrixes typically take reliability, utility, and cost/benefit into account. Today, social and ethical concerns are included in such assessments (Wienroth, 2020b), but, in what manner? As mentioned, the European Commission issued its *Ethics Guidelines for Trustworthy AI* in 2019, focusing particularly on professional/public interaction systems (European Commission, 2019). Guidelines are urgent, it is maintained, because AI technology is likely to become crucial to solving future challenges (European Commission, 2019, p. 33). These include useful services that AI makes possible (European Commission, 2019, p. 9), but unfortunately, also threats like digital ID theft, covert tracking of individuals, citizen scoring, lethal autonomous weapon systems, and 'potential longer-term concerns' (European Commission, 2019, pp. 33–35).

The guidelines emphasise that only AI that is lawful, ethical, as well as technically and socially robust, can be characterised as 'trustworthy' (European Commission, 2019, pp. 2, 5), and the salient moral concerns that underpin trustworthiness are human rights, privacy, transparency, fairness, bias, and accountability (European Commission, 2019, p. 2). So, we must assume that machine-learning technologies need testing regimes that can assess the presence or absence of these values[6] through studying the biases of learning databases, algorithmic bias, knowledge presentation, etc. Establishing trustworthiness in this comprehensive sense clearly involves more than believing the hype and initial anecdotal impressions of products below TRL 7. Indeed, studying moral and social issues (and their potential impact) is a time-consuming and costly process and may be a factor that contributes to confining a product to the Valley of Death.

Today, internet crowd funding and social media hype can push products through the Valley. Mass distribution of half-finished products, perhaps for a small fee, can help establish technologies as emerging, possibly disruptive. 'Recreational' products may also be considered useful in the context of policing, perhaps as open-source intelligence (OSCINT) measures, or as investigative tools. Such 'function creep' (Dahl & Sætnan, 2009) took place

5. Technologies that actually have a game-changing impact are often referred to as 'disruptive' (Danneels, 2004).
6. These are standard tenets within science ethics, i.e. the promotion of respect for human dignity, privacy, duty to inform, consent/inform, confidentiality, avoiding harm, impartiality (avoid bias/fairness), and accountability (NESH, 2016).

in police utilisation of GEDmatch (Kennett, 2019), triggering questions of reliability and ethics. As is the case in many matters concerning police work, there will be competing views, values and interests (Wienroth, 2020b, p. 539). This clash of interest is probably exacerbated in the context of emergent technology.

4. Two emergent police applications

At first, the police used AI for knowledge management purposes (Alzou'bi et al., 2014; Vestby & Vestby, 2019). Today, AI is crucial to many police forces in performing order maintenance, criminal investigations and crime prevention through systems like PredPol, Valcri or Palantir.[7] Well-established systems might of course involve ethical and social challenges too, but *emergent* deep learning technology is more likely to produce 'unknown unknowns' (Taleb, 2010) – particularly when the documentation for both the learning databases and the valuing systems (algorithms) is sparse. In the following, an emerging facial recognition application and 3D DNA phenotyping are studied in some detail.

4.1 Facial recognition: Clearview AI

Well before Bertillon's standardisation of mug shots in 1888, facial recognition was used to establish the identity of offenders. In *The Pickwick Papers*, Charles Dickens (1836, 2009) described how the police/wardens memorised the faces of offenders in prisons as the inmates were 'sitting for their portrait' and 'having their likeness taken'. Later, mugshots were catalogued according to different facial features, age, geographical origin, etc., before digital photos and computers revolutionised the speed and flexibility of retrieval. With the increased accessibility of automatic facial recognition from 2006 onwards,[8] automatic recognition of persons of interest in real-time became possible, and such systems are now found at any border checkpoint or sensitive area.

Even though face recognition is commonplace, and the police have access to digital, searchable databases, the success rate is far from 100 percent. A blurry tele-lens photo of a person may fail to match with a high-contrast mugshot (or 'face print') taken years ago. The algorithm comparing the two might be too weak, or the person of interest might not even exist in the police database.[9] The former problem presents a technical challenge; the latter problem could in part be mitigated by having unlimited access to social media images. Google's or Yandex's image search can prove helpful, but a recent app from Clearview AI purportedly performs both tasks far better. By uploading a photo from a surveillance camera or a smart phone, the Clearview app compares the snapshot to a database of three billion photos, 'scraped from Facebook, YouTube, Venmo and millions of other websites' (Hill, 2020b), exceeding governmental databases or those of the Silicon Valley giants by many factors. Clearview AI's code is also said to support augmented-reality glasses, thus enabling wearers to identify anyone within view, immediately revealing their names, addresses, occupations, as well as their online networks (Hill, 2020b). According to *The Guardian*, officers in more than 2200 law enforcement agencies have been 'live testing' this app by February 2020 (Taylor, 2020), and according to the provider it has also been licenced to 'a handful of

7. See Revell (2017). AI in policing is defined loosely as the 'growing use of technologies that apply algorithms to large sets of data to either assist human police work or replace it' (Joh, 2017, p. 1139).
8. Cf. the Face Recognition Grand Challenge competition (Phillips et al., 2005).
9. For live face-recognition NeoFace is used by several police forces (Klontz & Jain, 2013).

companies for security purposes' (Hill, 2020b). It was also made available for some investors at an early stage (Hill, 2020a).

The algorithms of the Clearview AI application are opaque to the customers, as is the exact scope of its database. But the app seems to provide otherwise unavailable information instantly, and has, allegedly, helped solve several difficult cases. It qualifies as *emergent* technology according to the criteria cited above: It is novel, it is growing in use, and it has had some impact. Uncertainty and ambiguity are definitely still present. It is also 'coherent' in the sense that it has moved beyond the conceptual stage and found a commercially viable form within the security market.

The Clearview AI system may qualify as 'emergent', but it is not *alien* to us. It performs trivial tasks, only in a superior way. Google has refrained from developing similar far-reaching apps – allegedly for moral reasons, and Facebook and Twitter have responded by denying Clearview's use of 'their' photos (Hill, 2020c). But, as the technology and its 'dirty' databases already exist, the system may well resurface, even if 'Clearview AI' is ushered back to the Valley of Death. One might think that its surveillance capability would diminish public willingness to document lives online, but Clearview's success would have been impossible in the first place without an altered public valuing of privacy (Hoofnagle et al., 2019; Zuboff, 2019). After all, people willingly post photos and coordinates on social media, and even their own DNA to online services like GEDmatch and AncestryDNA. In smart, online societies, privacy seems to have been offset by other concerns, particularly if understood as 'the extent to which [personal] information […] is communicated to others' (Bellaby, 2012, p. 102).

A Dirty Harry of the digital age may therefore argue that Clearview AI is socially *acceptable*, at least for purposes of policing – but is it *trustworthy*? Technically, a Clearview AI search seems similar to performing an ANPR (automatic number plate recognition) query. The app simply maps certain features in an input photo and searches for a set of similar features ('faceprints') in a database. If ANPR is considered trustworthy, Clearview must, by analogy, be trustworthy, too. However, the vehicle registration database, which ANPR relies on, consists of a set of records identified by unique letters and numbers, whereas Clearview's data set cannot be validated (labelled) in the same manner. Faces change, and a match is always a matter of degree of similarity between two faceprints. Superior matches may in principle *always* exist outside of Clearview's database. Worse, its database content is not even initially subject to any quality control. Millions of fake social media profiles exist.[10] The fact that the database is collected without permission raises moral issues as well; as does the fact that the learning strategies/algorithms of Clearview are the secret of a private company. The ANPR analogy is therefore not as striking as it may seem at first glance, even when only the technical issues are considered.

Still the Clearview AI application may be of value to the police. As time goes by, the 'operator' may become familiar with the limitations of the app, e.g. the types of errors that are likely to occur. The police officer can also visually compare the input photos and the matches. If the matches seem reasonable most of the time, the trustworthiness of the app increases. If it yields better results than the alternative systems, one might at least argue that the app is *useful* – despite its inherent legal and moral issues. Some police officers and departments are clearly willing to trust the application without proper testing or understanding of the system:

10. FaceBook acknowledges that there are at least 100 million fake FaceBook accounts (Rosen, 2018).

> Federal and state law enforcement officers said that while they had only limited knowledge of how Clearview works and who is behind it, they had used its app to help solve shoplifting, identity theft, credit card fraud, murder and child sexual exploitation cases (Hill, 2020b).

A *wish* to trust a tool based on anecdotal results is hardly a sufficient criterion for trustworthiness in the abovementioned sense. Rather, I would like to claim, it encourages Digital Dirty Harryism, i.e. a conscious utilisation of a dirty means in order to achieve some good. For instance, the system *biases* are still unknown. Although algorithms lack *human* implicit and explicit biases, machine learning processes may introduce biases stemming from historically biased training databases, or even inherent *algorithmic* bias (Kitchin, 2017) through the manner the algorithm assigns the numerous weights. Whether such biases represent a problem in the case of Clearview AI is impossible to determine without rigorous testing. The system *seems* to work well under some circumstances, but without thorough experimentation it is impossible to establish whether the AI in question is trustworthy (Marcus & Davis, 2019, pp. 57–58). Offering this system to law enforcement agencies at the present stage seems like a desperate attempt to save the project from the Valley of Death.

As a last-ditch attempt to solve a cold case, or in an otherwise exceptionally important case, Clearview AI certainly poses a temptation. The quick fix perspective is exactly what invites Digital Dirty Harryism. There are, however, good reasons for restraint in this case, perhaps insisting on the importance of privacy, even if its flavour has changed. After all, Clearview AI is developed and owned by a private company willing to steal data, and unwilling to share its business secrets.[11] Data *security* is also clearly an issue. In using the app, the police are required to upload sensitive photos to Clearview's servers. The company's ability and commitment to protecting data has not been vetted by independent agencies. Some early users have even claimed that the company monitors their input and tampers with data (Hill, 2020b). Such socio-ethical issues weaken the application's trustworthiness, and add to its 'dirtiness', not through the operator's bad intention, but through covert interests of ownership.

4.2 DNA phenotyping

Fingerprinting was the most prominent biometrics identification method until DNA analysis became available in the 1980s. On the basis of 13 (or sometimes 20) highly variable markers, individuals can be identified with near 100% certainty (Matheson, 2016). DNA samples may thus link a person to a crime scene if the DNA markers match with an entry in a (police) DNA database.

The DNA markers used by the police have until recently been *genotypes*, i.e. not connected to specific visible traits of the person. Other parts of the DNA contain *phenotype* information (visible features). Walsh and Keyser reliably managed to predict eye-, hair- and skin colour from other DNA markers – insisting, however, that that their method is not intended for identification of individuals. Still, examining these traits may provide a starting point in cold cases (Matheson, 2016). If the DNA from a suspected offender is determined to belong to a blue-eyed, brown-haired, fair-skinned person, this narrows down the scope for later genotype screening considerably. Interestingly, Mark Shriver and Peter Claes claim to be able to predict *face shapes* by DNA, based on 200 genes associated with facial development (Mathe-

11. The threat of 'vendor lock-in' must be considered seriously. Even with the widely used Palantir system, data has been lost when police departments have terminated their subscription to the service (Harris, 2017).

son, 2016). If facial features can be reliably modelled from DNA alone, it would present a major breakthrough: A 3D model of the face of a person can then be shaped from a crime scene DNA sample (Tremblay, 2014). Granted, aging and non-biological factors may influence the looks of a face, but the 3D-model may give a telling impression of the face, perhaps at different degrees of maturity.

This technology, based on medical science, is 'emergent' in the sense that it has a novel character, a potential for fast growth and impact, and involves uncertainty and ambiguity. 3D phenotyping can, of course, not produce faceprints to be fed into facial recognition systems, but the police can at least get a rough facial composite of an offender based on DNA, even if no observations exist. What was recently a sci-fi dream is now a possibility (Janos, 2018). In the case of the Golden State Killer, tracked down by forensic genealogy, the police also had a DNA-based 2D image resembling the perpetrator (Wickenheiser, 2019). Forensic DNA-based phenotyping conducted by Parabon NanoLabs has been useful in solving several cases and has been portrayed as a 'fully operational' method, although this assessment is contested (Wienroth, 2020a).

*

3D phenotyping clearly differs from face recognition. Whereas the latter compares a sample with a database, the former creates a visual model based on invisible DNA information. Of course, phenotyping depends on system learning by comparing a database of facial forms with DNA samples but rendering a 3D face model still requires extrapolation. Uncertainties involved in making correlations between the DNA and visible features are readily acknowledged (Marano et al., 2019), and non-DNA factors interfere, too (Janos, 2018). As the complexity of the task invites deep machine learning, further challenges to the accountability are introduced, since the deep learning weighing mechanisms are hard to account for.[12] Unless the training database is huge, reliability is bound to be quite low, and tracing system-inherent biases becomes difficult as well. In addition, there is no common-sense quality control for 3D modelling by phenotyping – unless the person is known in advance.[13] Bias is likely to occur in the training of the system, and fear of racial profiling has been an issue (Sero et al., 2019.). Groups of people have simply refused to participate in phenotyping trials,[14] thereby reducing its trustworthiness. One may also well imagine different forms of function creep in using DNA for such purposes.[15] For all these reasons, trustworthiness is hard to establish.

Presently, neither the Clearview AI app nor 3D phenotyping seem to meet the European Commission's criteria of trustworthiness. But one may ask, what if 3D phenotyping or the Clearview app could help the police solve cold cases, or identify child molesters? Why focus solely on *trustworthiness*? After all, every single day we rely on the forces of nature, other advanced systems, and persons we hardly know or understand. If everything else fails, why

12. As Marcus and Davis (2019, p. 57) claim 'there is an unsolved mystery about why neural networks work as well as they do, and a lack of clarity about the exact circumstances in which they don't'.
13. Identikit face reconstruction may seem analogous, but identification by facial composites is still based on direct observation. Offender *profiling*, i.e. creating an offender 'bio' from traces of actions on the crime scene, seems methodically closer to 3D phenotyping.
14. On the other hand, we might be even more vulnerable to bias if we remain content with natural intelligence. See also Brantingham, Valasik, and Mohler (2018).
15. Futile attempts to identify criminal features have been performed by phrenologists like Franz Joseph Gall, as well as in a more recent project (Dellinger, 2020).

should the police refrain from trying emergent technology, especially if criminals or terrorists are utilising the same measures? Sometimes legalities are set aside:

> [In Germany] they saw the deployment of forensic DNA phenotyping as legitimate even though it was not considered legal at the time, calling for a revision of the law to permit routine use in policing by ascribing significant value to them as law and order tools (Wienroth, 2020b, p. 593).

From a pragmatist point of view, this seems acceptable, and police officers are not necessarily interested in knowing *how* high-tech systems work, but rather *that* they work (Kaufmann, 2018, p. 157). Can discrete acts of Digital Dirty Harryism be considered acceptable? The EC ethical guidelines offer little guidance in the grey zone.

5. Digital responsibility, digital dirtiness

The discussion of acceptability and trustworthiness should not be confined purely to technological issues. The mode of interaction is also an important factor – as system *reliability* may well be enhanced by human involvement. If judgments and decisions are solely a human responsibility, systems are often referred to as 'human in the loop' (HITL) systems. Such AI systems (e.g. guided missile systems or some language translation tools) typically generate options on the operator's request, which the operator accepts or rejects. HITL-systems do not 'act' unless human consent is given. In other contexts, human involvement is not necessary or even desirable, e.g. in well-defined, but tedious and work-intensive tasks that may cause humans to err, or in tasks that require super-human abilities (e.g. split-second rocket launch abort systems). Such systems are referred to as human-out-of-the-loop systems or HOTL for short (Eliot, 2019). HOTL systems are supposed to be 'responsible' by design, in so far as 'responsibility' is a meaningful attribution to digital systems (Coeckelbergh, 2019). However, as we have seen, AI responsibility is clearly problematic in the discussion of *emergent* systems.

Clearview AI and 3D phenotyping do not fit the HOTL or HITL descriptions well. Rather, these tools or methods enhance or *extend* the operators' cognitive or physical capability,[16] without necessarily limiting their discretionary power. These systems may be thought of as Human-governing-the-loop (HGTL) applications. Operators are not supposed to blindly accept their suggestions, although part of the deliberation process is sourced out to an opaque artificial intelligence component. However, with augmented powers, increased responsibility ought to follow. Responsible governing of emergent system loops requires a *critical* attitude, where the operator also actively *tests* the system. Adequate understanding of different *modes of interaction* between technology and humans (HITL, HGTL and HOTL) is also needed, as these may vary. For instance, in the context of self-driving cars, the autonomy distinction is made in terms of levels 1 to 5 (Litman, 2020, p. 8). Level 5 indicates a HOTL conception, where no driver is needed. Level 4 resembles HITL (where a human driver continuously oversees the AI choices), whereas level 2 is similar to the HGTL model, where the driver is driving the car, but using 'adaptive cruise control' or 'lane holder' in order to ease the workload. At times the driver might find him- or herself on a slippery slope, going from governing the loop to just being a human in the loop, accepting every suggestion the application/car presents. As the trust in the system increases, it may take a conscious effort to

16. For more on extension theory, see Brey (2017).

remain attentive, to avoid leaving the 'loop' to itself. Part of assessing the emergent system is to decide the level of human involvement that yields trustworthiness.

Another argument for maintaining a critical attitude relates to the preliminary user interface of emergent technology. It matters whether a match is presented like a 'jackpot' (sound and blinking lights), or '83 % similarity' displayed in Helvetica typeface. If a thermometer turning red predicts excessive vulnerability in child welfare contexts (Eubanks, 2018, p. 141), a busy social worker/operator is emotionally nudged into taking action. Campolo & Crawford (2019, p. 10) claims that the 'epistemological flattening' of complex social contexts into clean 'signals' for the purposes of prediction, has a bearing on the social applications of machine learning. Suggestive *presentation* of probable correlations is easily taken for artificial *decision-making*.

In our context, facial models based on phenotyping hardly facilitate reasoning about probabilities or critical thinking, as the expression of face models appeals intuitively. A resemblance between a 3D face model and some person known to the police (but hitherto unrelated to the case) is hard to suppress. But particularly in an experimental HGTL setting, AI should merely provide decision *support*, not clear-cut consent options. Marcus and Davis (2019, p. 192) suggest that we should ask ourselves: can we reach the same conclusions based on the same facts [as the system] in another manner? If this is impossible, being a morally responsible governor of the loop is hard. In our setting, Marcus and Davies' common sense criterion seems like a necessary, but hardly a sufficient condition. System trustworthiness depends also on the operator's *critical assessment* of the validity of the output as well as the degree of operator involvement. Adequate understanding of the output presentation is crucial. Responsible loop-governing implies both an understanding for *when* to use systems and *how* to interpret system output.

Emergent technology may also intensify the *layers of responsibilities* problem. As the system in question is a novelty, resulting errors and failures are not necessarily the fault of the operator. System designers, programmers, machine-learning supervisors, or other testers/operators share the blame if the system fails. Although the multiple sources of error should make any operator reluctant to trust the system, this fragmentation of (moral) responsibility may also increase the boldness of some operators (Campolo & Crawford, 2019, p. 12). The role of the operator of emergent systems, however, should not be that of Milgram's test subjects. As part of the reinforcement or 'post-learning' of the system, the operators should adhere to the values of research ethics which are more or less identical to the socio-ethical values pointed out by the EC guidelines on trustworthiness. Similar values are central to policing as well.[17] Thus the trustworthiness of the emergent system hinges on the trustworthiness of the user. With augmented powers comes increased responsibility.

Use of untrustworthy equipment in *irresponsible* manners is, of course, hard to defend morally. The simple solution would be to prohibit emergent technology in policing. However, early adoption or aggressive testing of less than trustworthy systems may seem reasonable under some circumstances (Gartenstein-Ross et al., 2019).

Digital Dirty Harrys may cling to ambiguities of responsibility, and pragmatic forms of integrity. They push the limits by appealing to the greater good. During the Cold War 'missile gap' crisis, the US government maintained that 'every known technique should be

17. The Norwegian Police Act, for instance, emphasises officers' duty to take into considerations concerns of human rights (Police Instruction § 3-1), privacy (in terms of 'public exposure', Police Act § 6), fairness ('proportionality', Police Act § 6, Police Instruction §5-2), impartiality (i.e. 'avoid bias', Police Act § 6), and accountability (Police Instruction § 7-6).

used and new ones developed to increase our Intelligence by high altitude photographic reconnaissance and other means', and that 'no price would be too high to pay for the knowledge to be derived therefrom' (Dulles, 1954, p. 1). More recently, The Police Foundation applauded comprehensive airborne city surveillance, covertly tested by the Baltimore Police Department, arguing that…

> …[t]he police do not always have the luxury of waiting until research yields scientific evidence about the efficacy of a particular approach. When people are dying the police must act to stop the violence – even when doing so carries a degree of political risk (…) This is the hallmark of courageous leadership and should be acknowledged (the lack of clarity regarding the implementation of the program notwithstanding) (Police Foundation, 2017, p. 21).

Other commentators held the covert use of this emergent surveillance system to be both undemocratic and immoral (Rector & Broadwater, 2016). But what do democracy and morality require? Is there an acceptable, pragmatic interpretation of trustworthiness? After all, the values that trustworthiness encompasses according to the European Commission (transparency, privacy and security), may be up for reinterpretation in a smart city, surveillance capitalist era (Zuboff, 2019, pp. 35–37). Further, the capabilities of dubious systems may be utilised by deviant groups and also by other professions, and there might be morally intense situations where the public expects the police to push the limits. For instance, even if the Norwegian police do not, at present, utilise genealogic genetics, other police forces do, and private providers offer various forms of genetic analyses. The questions of *when* and *how* emergent measures should be utilised could benefit from a broader discussion.[18] Initiatives by inventive police officers, like the submitting of crime-scene DNA to GEDmatch as one's personal DNA, the performing of private social media searches that leave electronic traces, or private testing of Clearview AI, are cyber-symptoms of a need for a more robust strategy.

Establishing a high level of trustworthiness (TRL10) for police technology is time-consuming, and police forces may easily find themselves out of synch with fast-paced technological and societal developments. The life cycle of emergent products may also be so short that by the time they are deemed 'operational' the next big thing is already out, or criminals may have found effective countermeasures (Gartenstein-Ross et al., 2019). If the police cannot keep up with the trends, they are easily outsmarted, and militarisation or *old-fashioned* Dirty Harryism may be the only responses available. But brute force seems misplaced in an ever 'smarter' society. Being at the forefront of technological innovation, on the other hand, seems to require live testing of unfinished technology, that is, to some extent *Digital* Dirty Harryism.

Now, one may agree with Waddington (1999, p. 299) that much police work is in fact 'dirty work', and that the primary function of the police is the use of coercion. The present-day focus on accountability and trustworthiness are just novel forms of 'concealment and circumlocution' (Klockars, 1988) of this brutal fact. In 'Ruthlessness in public life', the philosopher Thomas Nagel (1979) argues that people are often willing to grant government officials some moral leeway, if it serves the interest of the public. Still, Waddington, Klockars, and Nagel hardly claim that 'anything goes', or, that suboptimal solutions or practices should be promoted. Rather, the police ought to be one step ahead of the crime trends; they ought to

18. As Wienroth (2020b, p. 594) writes: 'Forensic genetics in criminal justice reflects on the social, even political nature of the ways in which socio-technical innovations are debated, legitimised and deployed.'

anticipate novel forms of crime rather than work pre-emptively (Halvorsen, 2018, p. 29). In our context, it is the lack of technical situational awareness and planning that make improvised solutions necessary. Klockars (2005, p. 583) argues that three criteria must be fulfilled in order to justify 'dirtiness' – or in our context: using less than trustworthy means.

The first criterion states that *the method* (here: emergent AI) *must actually be able to provide the promised output.* As we have seen, the hype claims that the Clearview AI application does so. Still, there are reasons for questioning its data quality. 'Scraping' social media with its millions of fake accounts for images hardly ensures data integrity. Over time, the algorithm/valuing mechanism of the system may to some extent compensate for a less than optimal learning database, and the system's reliability may also increase through repeated successful in-field use. But anecdotal success cannot establish trustworthiness in general. Proper testing demands a methodical search for the *weaknesses* of the technology. The reliability criterion can probably only be reached through actively seeking out system vulnerabilities (Popper, 1963).

If it seems likely that 'dirty means' can produce the information in question, Klockars adds that *it must be the* only *way of finding that information.* In other words, if all other means and methods fail (or are obviously futile), it might be acceptable to utilise experimental systems. A 3D model based on phenotyping may perhaps help investigators to generate ideas 'outside of the box' or reduce the scope of an investigation. It *may* thus point the police in the direction of possible persons of interest, in much the same manner as unreliable informants *may* provide leads in a case. If trustworthy methods such as genotype testing can corroborate the 'findings' of phenotype 3D modelling, the latter may at least be instrumental in the production of trustworthy investigative leads. If further *reliable* steps are available, both of the above-mentioned AI applications may jumpstart trustworthy knowledge production. Still, this is dependent on a responsible human governing the loop. The temptation to trust and over-use such emergent means should be tempered by awareness of system limitations and side-effects.

In addition, according to Klockars, *it must also be likely that the output will produce a morally good end.* One might argue that the WW2 Nazi hypothermia research provided useful results, as the results indicate the upper timeframe for rescue operations in the North Sea in wintertime. Having such knowledge would be an asset to rescue crews – and potentially to those lost at sea – but the grotesque killing of the research subjects precludes the possibility of a *morally* good end to the research project.[19] In addition, the project was ill-designed, and also failed the first two criteria.

Even if the emergent technologies presented above are less grotesque, they should not be allowed to take exception from moral constraints – on the contrary. If shrouded in secrecy, the assessments of their moral goodness (if any) are limited to the parties 'in the know' and typically focus on short-term gains. Techno-optimism may serve to demonstrate the validity of Amara's law, i.e. that the short-term effect of technology is usually less than expected, whereas the long-term consequences (and side-effects) are typically greater.[20] In a longer-term perspective, the interests, concerns, and expectations of *all* parties involved ought to be included. For instance, what are the side-effects of violating private data ownership of citizens through a third-party application (as in the Clearview case)?[21]

19. Conducted in Dachau (Berger, 1990). The discussion about whether the use of the studies would serve to honour the victims of this research, or the opposite has been shelved as the study lacks both moral and scientific integrity.
20. Coccia (2020) claims that artificial intelligence itself is a paradigmatic example of said law.
21. That the police seek matches to Joe Public's photo is probably valuable information for a company considering employing Joe.

Klockars's three criteria state the *conditions* for possible utilisation of emergent technologies, and as we have seen, the operators should be highly attentive to their own role in the loop. They should, I would like to suggest, consider themselves as research staff or 'beta testers' in the development cycle.

6. Critical thinking, testing, trustworthiness

The mentioned criteria are easily justified in a crude utilitarian framework, but the classical utilitarian requirements of considering long-term utility and side-effects (duration, fruitfulness, purity) for all involved parties deserve closer scrutiny[22]. A thorough 'pre-learning' phase and beta-testing[23] do not guarantee the *legitimacy* of the system, as beta tests typically address functionality, accountability and reliability in the *technical* sense.

It is difficult to imagine AI trustworthiness, as defined by the European Commission, without a 'post-learning' *controlled test regime*. Ethical and social robustness (legitimacy) requires comprehensive testing, similar to clinical drug/therapies trial regimes. The latter take place in several stages, first as small-scale tests, then larger scale, and finally, in some form of 'post market surveillance trials' to make sure that the drug/therapy/system works as advertised in practice (Pegoraro et al., 2007, p. 160). In addition, research projects involving people must be vetted by research ethics committees (RECs) up front, and clinical trials require the *informed consent* of those exposed to the testing.

In large-scale, later stage trials (100+ research subjects) new interventions or drugs are compared to existing ones in randomised trials, providing knowledge of whether the results are due to the new intervention, or luck, coincidence, or the competence of the 'research staff'. In medicine, new drugs are seldom approved by health authorities unless they are tested at this level. Similarly, nor should emergent police technology. Thanks to the comprehensive test regime, drugs come with reliable information about dosage, limits of use, probabilities for side-effects, etc. The testing regime also helps uphold a professional value system within medical research, a cultural feature that should not be taken for granted. The situation is different in the cyber world. Most programs/apps, even the Internet itself, contain grave and notoriously undocumented security issues (Schneier, 2018). The police should not contribute to this sorry state of affairs.[24]

In the case of emergent police technologies, it is understandable that the police are reluctant to reveal the full scope of their new capabilities. Citizens may well protest to being exposed to the new, potentially intrusive, measures. Further, openness may incite the production of counter-measures before the efficacy of the emergent technology is established (Gartenstein-Ross et al., 2019). However, far from all police technologies require a shroud of secrecy. In our context, 3D phenotyping modelling and forensic genetics analyses are types of socio-technical innovation that involve so many aspects that they are actually *in need of* wide 'professional and public discourses about law and order, criminality (…) [that may guide] adoption, appropriation, and further innovation' (Wienroth, 2020b, p. 595). Otherwise, it is impossible truly to establish their legitimacy.

22. Cf. Bentham's felicific calculus, see Troyer (2003).
23. See https://www.softwaretestinghelp.com/beta-testing/ for an overview.
24. In addition, one conducts late stage or 'post market' trials of approved drugs. The main purpose of such large-scale trials is to study short/long-lasting side effects and safety, which require large groups of people to be studied over time.

Comprehensive testing regimes involve research investigators, people actually doing the testing (police officers), and test subjects exposed to the trial (the public). In some cases, it may seem far-fetched to seek informed consent from the public in order to test cutting-edge police technology. However, if reliability, moral legitimacy, and social robustness are important issues, one could certainly do worse than seek public consent. At least, the deliberative horizon cannot be confined to the policing cultures and investigative practices alone, as emergent, possibly disruptive technologies in addition typically involve and concern many parties: 'scientific cultures, legal and regulatory frameworks […]; the market (forensic service provision and technology development), subjects of technology use (e.g. victims and their families, suspect individuals or communities)' (Wienroth, 2020b, p. 594), as well as the public in general. If this is the case, informed consent or other expressions of participation ought to be *required*. In practice, citizens might willingly consent to sharing both photos and DNA if the purpose is *only* to train algorithms into being less biased or to construct faces more reliably. But this hinges upon there being a high degree of public trust in the police. As documented by the British police (Metropolitan Police Service, 2020), live testing is possible, also for surveillance measures. In contrast, secretly using rogue facial recognition applications hardly serves to increase trust.

7. Concluding remarks

Should the police make use of emergent AI technology if it is likely to produce a favourable outcome? The answer seems to be a qualified 'yes', insofar as the information sought is important and obtainable by no other means, and the risk of suffering injury – both short and long term – is *proportionate* to the good to be gained, and serves a just cause (Bellaby, 2012, pp. 114–115). As professionals, the police have a duty to be up to speed on social and technological developments – but not in a haphazard manner. As we have seen in the case of Clearview AI, the application may work, but the police officers using the application seem, more or less unknowingly, to have taken the role of beta-testers, submitting sensitive data to a private company. In addition, its database is 'stolen', and the potential citizen victims of this 'experiment' were not informed. In sum, it seems like a paradigmatic case of irresponsible use of dirty technology. Private live testing of hyped-up tools might be tempting, but is morally unacceptable. As a general rule, the research subjects should be informed and consenting to being exposed. If secrecy is absolutely necessary, at least representatives of the people (e.g. control committees) must be able to give some form of consent by proxy.

However, as trials are time-consuming, there will still be scenarios where it seems permissible and perhaps appropriate to push limits. In medicine, too, there are situations where 'off-label' and 'unlicensed' use of approved drugs is considered acceptable (McIntyre et al., 2000), and the recent, rushed development of Covid-19 vaccines provides a striking example (Rupali et al., 2020). Admittedly, in the context of policing, such grey zones may invite function creep and Digital Dirty Harryism. In general, it is better to be well prepared, than caught-off by exceptional circumstances. Fostering environments for police innovation and research that are capable of facilitating the testing of emerging systems – also in terms of their moral and social aspects – seems like a reasonable requirement. Undoubtedly, grey zones will still occur, but they should be made as small as possible.

Pushing moral limits may to some represent a display of strength and courage. To others it represents a foolish lack of foresight, justified by 'exceptional' circumstances that most of the time would have been anticipated by a technologically and socially vigilant police force. With the advent of smart cities and the increase in cybercrime, a scientific, forward-looking

culture seems like a requirement for policing. In any case, the police should avoid ending up, as Schneier (2018, p. 20) writes, discussing 21st century phenomena, using 20th century language, and fighting them with 19th century means.

References

Allinson, J. (2015). The necropolitics of drones. *International Political Sociology*, *9*(2), 113–127. https://doi.org/10.1111/ips.12086

Alzou'bi, Suhaib, Alshibly, H., & Al-Ma'aitah, M. (2014). Artificial intelligence in law enforcement, a review. *International Journal of Advanced Information Technology*, *4*(4). https://doi.org/10.5121/ijait.2014.4401

Bellaby, R. (2012). What's the harm? The ethics of intelligence collection. *Intelligence and National Security*, *27*(1), 93–117. https://doi.org/10.1080/02684527.2012.621600

Berger, R. L. (1990). Nazi science – The Dachau hypothermia experiments. *The New England Journal of Medicine* (322), 1435–1440. https://doi.org/10.1056/NEJM199005173222006

Bostrom, N. (2014). *Superintelligence*. Oxford University Press.

Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and Public Policy*, *5*(1), 1–6. https://doi.org/10.1080/2330443X.2018.1438940

Brey, P. (2017). Theorizing technology and its role in crime and law enforcement. In M. McGuire & T. J. Holt (Eds.), *The Routledge handbook of technology, crime and justice* (pp. 17–34). Routledge.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the worlds*. The MIT Press.

Campolo, A., & Crawford, K. (2019). Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*, *6*, 1–19. https://doi.org/10.17351/ests2020.277

Coccia, M. (2020). Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. *Technology in Society, 60*, 1–11. https://doi.org/10.1016/j.techsoc.2019.101198

Coeckelbergh, M. (2019). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-019-00146-8

Dahl, J. Y., & Sætnan, A. R. (2009). "It all happened so slowly" – on controlling function creep in forensic DNA databases. *International Law, Crime and Justice*, *37*, 83–103. https://doi.org/10.1016/j.ijlcj.2009.04.002

Danneels, E. (2004). Disruptive technology reconsidered: A critique and research agenda. *The Journal of Product Innovation Management*, *21*(4). https://doi.org/10.1111/j.0737-6782.2004.00076.x

Dellinger, A. (2020). A twisted project that tried to predict criminals from a photo has come to an end. *Mic*. https://www.mic.com/p/a-twisted-project-that-tried-to-predict-criminals-from-a-photo-has-come-to-end-27621049

Dickens, C. (1836, 2009). *The Pickwick papers*. Project Gutenberg. https://www.gutenberg.org/files/580/580-h/580-h.htm#link2HCH0040

Drengson, A. R. (1982). Four philosophies of technology. *Philosophy Today*, *26*(2), 103–117. http://alandrengson.com/wp-content/uploads/2015/08/Four-Philosophies-of-Technology.pdf

Dulles, A. W. (1954). *Reconnaissance*. Washington, DC. https://www.cia.gov/library/readingroom/docs/1954-11-24a.pdf

Eliot, L. (2019). Human in-the-loop vs. out-of-the-loop in AI systems: The case of AI self-driving cars. *Aitrends* (April 9). https://www.aitrends.com/ai-insider/human-in-the-loop-vs-out-of-the-loop-in-ai-systems-the-case-of-ai-self-driving-cars/

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* Picador.

European Commission. (2019). *Ethics guidelines for trustworthy AI.* https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Gartenstein-Ross, D., Shear, M., & Jones, D. (2019). *Virtual plotters. Drones. Weaponized AI?: Violent non-state actors as deadly early adopter.* https://valensglobal.com/virtual-plotters-drones-weaponized-ai-violent-non-state-actors-as-deadly-early-adopters/

Gerstein, D. M. (2019). *The story of technology. How we got here and what the future holds.* Prometheus Books.

Halvorsen, V. (2018). Police practices in the age of precaution: A moral typology. In N. R. Fyfe, H. I. Gundhus, & K. V. Rønn (Eds.), *Moral issues in intelligence-led policing* (pp. 25–42). Routledge.

Hannon, C. (2018). Avoiding bias in robot speech. *Interactions*, *25*(5). https://doi.org/10.1145/3236671

Harris, M. (2017). How Peter Thiel's secretive data company pushed into policing. *Wired*, 08/09/2017. https://www.wired.com/story/how-peter-thiels-secretive-data-company-pushed-into-policing/

Hartmann, M. R. K. (2018). Grey zone creativity. In N. R. Fyfe, H. I. Gundhus, & K. V. Rønn (Eds.), *Moral issues in intelligence-led policing* (pp. 161–181). Routledge.

Hill, K. (2020a, 6 March). Before Clearview became a police tool, it was a secret plaything of the rich. *The New York Times.* https://www.nytimes.com/2020/03/05/technology/clearview-investors.html

Hill, K. (2020b, 18 January). The secretive company that might end privacy as we know it. *The New York Times.* https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

Hill, K. (2020c, 22 Jan). Twitter tells facial recognition trailblazer to stop using site's photos. *The New York Times.* https://www.nytimes.com/2020/01/22/technology/clearview-ai-twitter-letter.html

HLEG-AI (High-Level Expert Group on Artificial Intelligence). (2019). *A definition of AI: Main capabilities and scientific disciplines.* European Commission. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

Hoofnagle, C. J., King, J., Li, S., & Turow, J. (2019). *How different are young adults from older adults when it comes to information privacy attitudes & policies?* http://dx.doi.org/10.2139/ssrn.1589864

Janos, A. (2018). *Phenotyping: How a DNA 'snapshot' can create the face of an unknown criminal.* https://www.aetv.com/real-crime/phenotyping-dna-face-prediction-crime-investigating

Joh, E. E. (2017). Artificial intelligence and policing: First questions. *Seattle University Law Review, 41*, 1139–1149. https://ssrn.com/abstract=3168779

Kaufmann, M. (2018). The co-construction of crime predictions. In N. R. Fyfe, H. I. Gundhus, & K. V. Rønn (Eds.), *Moral issues in intelligence-led policing* (pp. 143–160). Routledge.

Kennett, D. (2019). Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. *Forensic Science International, 301*(May). https://doi.org/10.1016/j.forsciint.2019.05.016

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society, 20*(1), 14–29. http://dx.doi.org/10.1080/1369118X.2016.1154087

Klockars, C. B. (1988). The rhetoric of community policing. In J. R. Greene & S. D. Mastrofski (Eds.), *Community policing: Rhetoric or reality* (pp. 239–258). Praeger.

Klockars, C. B. (2005). The Dirty Harry problem. In T. Newburn (Ed.), *Policing: Key readings* (pp. 581–595). Willan Publishing.

Klontz, J. C., & Jain, A. K. (2013). A case study on unconstrained facial recognition using the Boston Marathon bombings suspects. *Computer, 46*, 91–94. https://doi.org/10.1109/MC.2013.377

Levin, S. (2018, 26 July). Amazon face recognition falsely matches 28 lawmakers with mugshots, ACLU says. *The Guardian*. https://www.theguardian.com/technology/2018/jul/26/amazon-facial-rekognition-congress-mugshots-aclu#:~:

Litman, T. (2020) *Autonomous vehicle implementation predictions implications for transport planning.* Victoria Transport Policy Institute. https://www.vtpi.org/avip.pdf

Marano, L. A., Andersen, J. D., Goncalves, F. T., Garcia, A. L. O., & Fridman, C. (2019). Evaluation of HIrisplex-S system markers for eye, skin and hair color prediction in an admixed Brazilian population. *Forensic Science International: Genetics Supplement Series, 7*(1), 427–428. https://doi.org/10.1016/j.fsigss.2019.10.038

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust.* Pantheon Books.

Matheson, S. (2016). DNA phenotyping: Snapshot of a criminal. *Cell, 166*(5), 1061–1064. https://doi.org/10.1016/j.cell.2016.08.016

McIntyre, J., Conroy, S., Avery, A., Cornsa, H., & Choonaraa, I. (2000). Unlicensed and off label prescribing of drugs in general practice. *Archives of Disease in Childhood, 83*, 498–501. https://doi.org/10.1136/adc.83.6.498

Metropolitan Police Service (2020). *Live facial recognition trials.* https://www.met.police.uk/SysSiteAssets/media/downloads/central/advice/met/facial-recognition/met-evaluation-report.pdf

Mongan, J., & Kohli, M. (2020). Artificial intelligence and human life: Five lessons for radiology from the 737 MAX disasters. *Radiology: Artificial Intelligence, 2*(2). https://doi.org/10.1148/ryai.2020190111

Murphy, L. M., & Edwards, P. L. (2003). *Bridging the Valley of Death: Transitioning from public to private sector financing.* NREL. https://www.nrel.gov/docs/gen/fy03/34036.pdf

Nagel, T. (1979). Ruthlessness in public life. In *Mortal questions.* Cambridge University Press.

NENT (2019). *Forskningsetiske retningslinjer for naturvitenskap og teknologi.* De nasjonale forskningsetiske komiteer.

NESH (2016). *Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology*. De nasjonale forskningsetiske komiteer.

New Zealand Government. (2020). *Algorithm charter for Aotearoa New Zealand*. https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf.

Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *Computer Vision and Pattern Recognition*, 427–436. https://doi.org/10.1109/CVPR.2015.7298640

Pegoraro, R., Putoto, G., & Wray, E. (Eds.). (2007). *Hospital based bioethics. A European perspective.* Piccin.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Jaesik, M., & Worek, W. (2005). Overview of the face recognition grand challenge. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1*, 947–995. https://doi.org/10.1109/CVPR.2005.268

Police Foundation. (2017). *A review of the Baltimore Police Department's use of persistent surveillance.* https://docs.wixstatic.com/ugd/9845f4_f7fd26e764374fcaa45115fd32edc22a.pdf

Popper, K. R. (1963). *Conjectures and refutations. The growth of Scientific Knowledge.* Routledge & Kegan Paul.

Rainnie, A., & Dean, M. (2020). Industry 4.0 and the future of quality work in the global digital economy. *Labour and Industry: A journal of the social and economic relations of work, 30*(1), 16–33. https://doi.org/10.1080/10301763.2019.1697598

Rector, K., & Broadwater, L. (2016, 24 August). Report of secret aerial surveillance by Baltimore police prompts questions, outrage. *Baltimore Sun.* https://www.baltimoresun.com/maryland/baltimore-city/bs-md-ci-secret-surveillance-20160824-story.html

Revell, T. (2017). AI detective analyses police data to learn how to crack cases. *New Scientist* (3125). https://www.newscientist.com/article/mg23431254-000-ai-detective-analyses-police-data-to-learn-how-to-crack-cases/

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review online, 94*(15). https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson_etal-FIN.pdf

Rosen, G. (2018). *Facebook publishes enforcement numbers for the first time.* https://about.fb.com/news/2018/05/enforcement-numbers/

Rotolo, D., Hicks, D., & Martin, B. (2015). What is an emerging technology? *Research Policy, 44*(10), 1827–1843. https://doi.org/10.1016/j.respol.2015.06.006

Rupali J. L., Sauer, M., & Truelove, S.A. (2021). Politicizing public health: The powder keg of rushing COVID-19 vaccines. *Human Vaccines & Immunotherapeutics, 17*(6), 1662–1663. https://doi.org/10.1080/21645515.2020.1846400

Russell, S., & Norvig, P. (2016). *Artificial Intelligence. A modern approach* (3 ed.). Pearson.

Schneier, B. (2018). *Click here to kill everybody: Security and survival in a hyper-connected world.* W. W. Norton & Company.

Sero, D., Zaidi, A., Li, J., White, J. D., Zarzar, T. B. G., Marazita, M. L., Weinberg, S. M., Suetens, P., Vandermeulen, D., Wagner, J. K., Shriver, M. D., & Claes, P. (2019). Facial recognition from DNA using face-to-DNA classifiers. *Nature Communications, 10*(2557). https://doi.org/10.1038/s41467-019-10617-y

Straub, J. (2015). In search of technology readiness level (TRL) 10. *Aerospace Science and Technology, 46*(October–November), 312–320. https://doi.org/10.1016/j.ast.2015.07.007

Taleb, N. N. (2010). *The black swan: Second edition: The impact of the highly improbable* (2nd ed.). Random House Publishing Group. Kindle Edition.

Taylor, J. (2020, Fri 19 Jun). Victoria police distances itself from controversial facial recognition firm Clearview AI. *The Guardian.* https://www.theguardian.com/australia-news/2020/jun/19/victoria-police-distances-itself-from-controversial-facial-recognition-firm-clearview-ai

Tremblay, É. (2014). Stranger visions by Heather Dewey-Hagborg. Reinterpreting portraiture through new forensic and 3D printing techniques. *Érudit, 103*, 11–100. https://www.academia.edu/download/37390673/articleHeatherDH.pdf

Troyer, J. (Ed.) (2003). *The Classical Utilitarians: Bentham and Mill.* Hackett Publishing Company.

Vestby, A., & Vestby, J. (2019). Machine learning and the police: Asking the right questions. *Policing: A Journal of Policy and Practice, 15*(1). https://doi.org/10.1093/police/paz035

Waddington, P. A. J. (1999). *Policing citizens.* UCL Press Limited.

Wickenheiser, R. (2019). Forensic genealogy, bioethics and the Golden State Killer case. *Forensic Science International: Synergy, 1*, 114–125. https://doi.org/10.1016/j.fsisyn.2019.07.003

Wienroth, M. (2020a). Socio-technical disagreements as ethical fora: Parabon NanoLab's forensic DNA snapshot™ service at the intersection of discourses around robust science, technology validation, and commerce. *BioSocieties, 15*(1), 28–45. https://doi.org/10.1057/s41292-018-0138-8

Wienroth, M. (2020b). Value beyond scientific validity: let's RULE (Reliability, Utility, LEgitimacy). *Journal of Responsible Innovation, 7*(sup1), 92–103. https://doi.org/10.1080/23299460.2020.1835152

Zuboff, S. (2019). *The age of surveillance capitalism. The fight for a human future at the new frontier of power.* Profile Books Ltd.