COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 10.09.2021 PC-CP (2021) 9 Rev

COUNCIL FOR PENOLOGICAL CO-OPERATION
(PC-CP)

Ethical, Strategic and Operational Guidance on the Use of Artificial Intelligence in Prison and Probation Services and the Private Companies acting on their Behalf

Written by Mike Nellis for the PC-CP
September 2021

1. Introduction: the purpose and context of the Recommendation

1.1 This Recommendation seeks to provide guidance on ethical, strategic and operational responses to the emerging and near-future use of artificial intelligence (AI) (and robots) in prison, probation and youth justice services, and also in the private companies that develop, provide or deliver services relevant to their work. To avoid both repetition and undue divergence from the emerging themes in European - and other international - debate on the ethics of AI, the proposals made here are deliberately grounded in more general work that has already been undertaken on this subject in the European Commission and in the Council of Europe. These are supplemented, however, with more critical arguments about the implications of AI and robots drawn from other standpoints and sources, which rarely figure in the literature produced by the Commission and the Council, but which are nonetheless important for criminal justice professionals concerned with, and steeped in, the traditions of *social justice* that have - or ought to have - shaped practices in probation and prison services.

1.2 For brevity and convenience of expression, the term "prisons and probation" (or "prisons and probation services") will be used throughout this document to denote by implication *all four agents* with which this Recommendation is concerned. Specific reference will be made to "youth justice" when required, when specific AI issues pertain to it, and distinct issues pertaining to "private companies" will be addressed in a separate section towards the end of this document. All the points and arguments made in the preliminary part of this Recommendation about the character and social implications of AI are intended to be of equal relevance to deliberation on all four agents, although the term "prisons and probation" is used - for convenience - to encompass them all. On a technical point, while there will be occasions in what follows when "robots" (which may or may not utilise AI)

are given particular attention, the generic comments made about "AI" invariably encompasses them, *unless otherwise specified.*

1.3 This Recommendation relies particularly on two existing European documents. Firstly, *Ethics Guidelines on Trustworthy AI* (2019) produced by the High Level Expert Group (HLEG) on Artificial Intelligence for the European Commission. Secondly, *Unboxing Artificial Intelligence - 10 Steps to Protect Human Rights (2019),* a Recommendation produced by the Council of Europe's Commissioner for Human Rights. Between them, they indicate

a) what a preliminary grounding of AI in the Fundamental Principles of the Treaties and Charter of the European Union entails;

b) how relevant ethical principles might be derived from them and

c) how these can then inform strategic and operational practices relating to AI.

The two documents overlap in some respects, and can be seen as complementary. *Ethics Guidelines on Trustworthy* AI concentrates primarily on delineating an ethical position but also offers some guidance on strategic and operational questions. *Unboxing Artificial Intelligence* is more cursory in its approach to ethics but offers a **clearer, ten-point framework** for addressing the strategic and operational issues raised by the deployment of AI, which are relevant to any organisation which attempts to do this.

1.4 The Fundamental Principles referred to here are respect for human dignity; freedom of the individual; respect for democracy, justice and the rule of law; equality, non-discrimination and solidarity; and citizen's rights (which in the EU includes "third country nationals"). As the HLEG says, these principles are far too abstract in themselves to provide a framework for appraising the moral implications of AI. It concedes that more precise ethical principles must be drawn out of them (which it then proceeds to elaborate, and which this Recommendation will address below). The HLEG further concedes that its guidelines are not tied specifically to any particular organisational or professional context, and admits "the necessity of an additional sectorial approach" to complement this (p6). This PC-CP Recommendation is one such sectorial approach, focussed on - to name the sector in full - "prisons, probation and youth justice services and the private companies who work for and with them".

1.5 The last of the Commissioner for Human Rights' ten points in *Unboxing Artificial Intelligence* is all important: "promote AI literacy". The development of, and commitment to, an ethical approach to the deployment of AI depends significantly on greater professional and (potential) stakeholder understanding of its character and implications, including the prevailing terminology and frequently shifting vocabularies used to describe it. Even *simple understanding* of AI and its full political, economic and professional requires greater awareness and literacy than is currently prevalent in European criminal justice agencies. By dint of the ground it covers and the arguments it makes, this PC-CP Recommendation will itself be a contribution to increasing AI literacy in the prison and probation sector - it seeks to make the AI-related issues that these agencies are likely to face in the near future clearer and more manageable. Moreover, as the HLEG (2019:9) say, complementing the Commissioner's point, however sophisticated and comprehensive an abstract statement of principles or a concrete code of ethics for a particular sector may turn out to be, it "can never function as a substitute for ethical reasoning itself, which must always remain sensitive to contextual details that cannot be captured in general guidelines". What AI literacy requires - in any

sector - is the creation of "an ethical culture and mind-set through public debate, education and practical learning" (HLEG 2019:9).

1.6 The confidence with which the literature on AI from major European institutions extols its transformative potential and the viability - as well as necessity - of framing its use in terms of human rights, democracy and the rule of law sometimes obscures an issue that is central to most other commentary on AI, namely the potential impact of AI on human work and patterns of employment, including professional employment (Ford 2015; Susskind D). One CEO in the AI industry, typical of many, spells out the optimistic view of what "the infusion of artificial intelligence into traditional industries" will entail:

> One defining area of AI infusion is in the automation of repetitive tasks, using technologies such as RPA (robotic process automation). RPA will see widespread application in the work that is performed by functions such as accounts payable, back-office processing and various forms of data management. The routine tasks associated with a large number of jobs will now lend themselves to automation, *freeing up people's time to focus on more complex endeavours*. RPA is currently creating some of the most advanced companies in the world. (Lee 2020:14 emphasis added).

The notion that one of AI-based automation's most important achievements is, or will be, the shift of employees energies away from "routine tasks" towards more important, "non- routine" tasks is commonplace in much of the literature that straightforwardly champions AI, including that from the European Institutions. It is a rather dubious argument. Much depends on what is defined as "a routine task". Richard and Daniel Susskind (2015), the former an established British authority on AI and work, were arguing six years ago that mainstream human tasks - core expertise, not just back-office routines - in eight different professions, including law, architecture and journalism could be replaced by AI, (a trend that was being driven in law by commercial legal firms). It is useful for AI's champions to promote AI as a benign and limited measure that will merely automate dull, routine, back office tasks but leave the recognisably core tasks of a profession, the human expertise which give it its identity, intact. But that may not be so: fully professional expertise is *already* within AI's purview. Much will depend on the economic and political value which is attached to these traditionally human/professional tasks. What might happen to probation services in such a context?

1.7 This Recommendation recognises that the rapid pace of scientific and technological change in the AI and robotics field, and its evolving implications for prisons and probation, mean that a date for a review of its own topicality and relevance. **We propose 2032**. This may seem, to some, to be too near/too soon a cut-off point. Arguably, however, much will already have changed in the decade between then and now, technologically and socially. Furthermore, arrangements made for AI in the next decade will lay the foundations for what comes after, for better or worse. Some "impacts", at least, which are now merely anticipated and imagined will have come to fruition by 2032, and their actual and likely consequences for good or ill made that much clearer. Will we (the PC-CP), a decade from now, stay on the same trajectory or will we want to change course? Reviewing this Recommendation sooner, in 2032, may enable a desirable change of direction, an opportunity which would be lost if re-appraisal of the issues were left until later. At the very least, re-appraisal would enable greater clarity about AI issues that in 2021/2 necessarily remain tentative and uncertain. It would also allow for as yet unforeseen shifts in opinion about AI's costs, capabilities and consequences to be taken account of. Government policies and penal priorities may shift between now and 2032. Public attitudes towards AI may alter, for better or worse, or both, creating

a more volatile milieu in which to make professional decisions about it. Amidst all the hyperbole (exaggeration and fantasy) which surrounds debate on AI and robotics, there is inevitably room for argument about what near-future developments *will actually occur* in respect of prisons and probation. There are probabilities, but few certainties. A revision of this Recommendation will surely become essential. 2032 is a good point to do that.

1.8 It is, nonetheless, timely to engage ethically with AI and robotics in the penal field now, because their use is already more prevalent than many realise - and already impinging on criminal justice and penal systems in Europe, where technological trajectories may already be being set. There are developments in other parts of the world, actively promoted by commercial manufacturers with global reach, which may well register with European institutions in the next decade.

2. The Mechanisms of Artificial Intelligence

2.1 While many definitions of AI - of variable scope - have been proffered, there is no consensus as to which is the best - most accurate, most reliable - one. This is not because a general technical description of what AI (broadly understood) can do is impossible - *it is not* - but because the scientists and technologists who design and build AIs, and the governments and businesses which commission, market and use them, have honest disagreements about their potential, their limits and their social implications. In Europe, both the HLEG and the Commissioner for Human Rights offer highly normative descriptions of AI, grounded in prevailing political - predominantly liberal - ideals (which will be addressed in Section 3). But both, in their appendices, also contain more-or-less identical, short, technical definitions of AI, sufficient to anchor their ethical and strategic work in new technological developments, but incapable, in themselves, of specifying the many uses and purposes to which "intelligent machines" might be put. To decide on those purposes, to "apply" and "make use" of AI technology, political and professional decisions must be made which ought not in themselves be "determined" by the availability of technology. *Professional bodies should never do things with technology <u>just because they can</u>: the application of new technologies to old tasks (like probation and imprisonment) requires an overarching justification.* This is easy to say, but harder to do when technological innovation develops and gathers momentum outside the profession itself, when a "paradigm shift" occurs which normalises the use of a technology across all sectors of society. This, to a greater or lesser degree, destabilises the existing social order. To use a favoured term in the global tech industry, its "disrupts" existing institutions and practices

2.2 This is the HLEG's narrow *technical* definition of AI, as both a "system" and a "scientific discipline":

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.
>
> As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement

> learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems) (High Level Expert Group 2019: 36)

2.3 More elaborate technical definitions make a distinction between task-specific **Narrow AI** and more versatile **General AI**. General AI denotes intelligent machines which can replicate or surpass a broad range of human intellectual capacities, rather than simply mastering a single (if still vastly complex task). This is "the holy grail" of many futurists, and only General AI is seen as truly capable of rivalling human beings (Susskind 2020:65). It is also how all-powerful AI's are most often presented in science fiction. At present General AIs do not exist, but they are scientifically feasible, and DeepMind (Silver et al 2021), a Google owned subsidiary with a strong track record in AI innovation, has announced that the existing technique known as "reinforcement learning" could soon be used to build one. All practical progress to date has been made in Narrow AI which can replicate or surpass a limited, but still sophisticated, range of hitherto exclusively human capabilities - mastering chess or Go, car driving, medical diagnosis or language translation. But Narrow AI can still have socially transformational consequences, so its impact in the near future, on particular institutions and professions, should not be underestimated. An AI need not *surpass* actual human capacities to be cost-efficient - in order to justify automating human tasks, removing humans in whole or in part from the process, it can be enough that they perform *the same basic task*, but faster, consistently and ceaselessly (without breaks, sleep, illness or resentment) - more cheaply and without interest in employee rights and unionisation.

2.4 The Commissioner for Human Rights (2019:24) uses the following definition of an algorithm: "a finite suite of formal rules/commands, usually in the form of a mathematical logic, that allows for a result to be obtained from input elements". AI algorithms can be of two kinds, top down and bottom up. Top-down algorithms control a machine with a pre-determined programme, making its behaviour highly predictable. Bottom-up algorithms - also called "stochastic algorithms" - allow a machine to learn from past experience and alter the algorithms with which it was originally programmed in the light of that. This is "machine learning" referred to above. To facilitate this learning, AI's are said to be "trained" on vast datasets which enable them to engage in ever more complex and rigorous pattern recognition processes, producing higher degrees of accuracy. These data sets are often complied and itemised by low paid human labour, a consideration which rarely figures in reflections on the "social costs" and ethics of AI (Crawford 2021). Bottom-up algorithms enable machines to function with some degree of autonomy from the humans who originally wrote their programmes, and do not require human intervention to improve their performance - this is how certain Narrow AIs learned to surpass and defeat all human masters of chess and Go. Their operations - the ways in which they arrive at recommendations, decisions and/or predictions - can become opaque to the people who programmed them. Applied beyond the confines of ultra-complex board games - say, to judicial sentencing - this opacity surely raises intriguing questions about accountability for decision-making, quite apart from other considerations that would arise using AI in a court setting.

Robots - a Special Type of AI?

2.5 The term "robot" was coined in the early 20th century by Czech writer Karel Capek (1920) in a fictional context to describe artificially made, human-looking machines that were, in effect, slave labourers in a factory. Programmed machines of variable size and function (and invariably *non-*

human appearance) undertook "manual labour", and replaced the workers who had previously done it, have existed in industrial settings since the mid-20th century, and have been colloquially called "robots". These machines, although automated, were not artificial intelligences. More recently, robots have been made which are, in varying degrees, artificially intelligent and are capable of replicating (indeed, on some levels, surpassing) complex forms of "cognitive labour" previously undertaken by humans. No easy distinction can be made between AI and robots, except that the latter are *usually* understood to be capable, within limits, of moving from place to place - how well, and how far, depends on how they are designed, and for what purpose. UNESCO's (2017: 4) definition of a robot begins with "mobility", while acknowledging that it is, in fact, an optional feature. UNESCO's four characteristics are:

i) mobility (for operating in indoor or outdoor environments),

ii) interactivity (using sensors and actuators to gather information necessary to function in a given environment;

iii) communication (using computer interfaces, voice and speech synthesisers and voice recognition software) and

iv) autonomy (an ability to act upon their environment, to recognise what action is required in a given situation, without immediate or direct external direction)

2.6 Robots in this advanced sense, of varying size and appearance, mobile or static, continue to be used in industry, but also operate in military settings (as autonomous mobile weapons systems, or drones). In different parts of the world, prototypes of mobile robots exist in health care, performing menial tasks in hospitals; in crime prevention, patrolling neighbourhoods and listening to and filming whatever goes on; and in "smart" homes which can adjust automatically to the requirements and convenience of the people living there - or most basically, in an ordinary home, rely on an automated vacuum cleaner). Autonomous vehicles for public, private and/or industrial use are sensor-rich robots designed for complex mobility that are expected to have a dramatic impact on Western transportation systems. The physical appearance of all the robots mentioned above reflect the functions they are designed to fulfil: none are human-looking.

2.7 Chatbots - currently non-mobile robots designed to interact verbally (or textually) with people (for example as customer service agents accessed by phone or internet, or personal assistants like Siri and Alexa may have only minimum physical embodiment to the person with whom they are interacting - a small box on a desk, enabled by wi-fi - but the same cloud-based AI processes which enable them can in principle be built in to human-resembling robots. There is a strong expectation in the tech industry that voice - rather than text, and tapping keyboards, although without fully replacing this - will become the main means of accessing and engaging with digital devices. One CEO working in this field speaks confidently about "the development of speech recognition and natural language processing in customer service, telemarketing and telesales" adding

> New advances in these technologies will allow 80% of queries to a call centre to be dealt with through automated processes, while still achieving higher customer satisfaction (Lee 2020:14-15)

This has greater implication for the future of prisons and probation than is commonly realised.

2.8 Contemporary robots which mimic human size and morphology - head, trunk, arms and legs, hitherto most common in fictional representations of robots - are largely deemed necessary by their designers to reassure the humans they will interact with. Prototypes of walking, talking and lifting care robots can assist elderly and infirm people overcome some of the routine problems of living, like bathing, but currently still look like the metal-and-plastic surfaced machines that they are. The most human-like of AI robots, down to the smallest physical details (although they cannot yet replicate walking), are sexbots, deliberately engineered to embody popular ideas of (mostly female) sexual desirability, or seated reception bots who can greet guests and visitors in hotels and corporate offices. In the near future, the acceptability or otherwise of chatbots, variants of reception bots and even sexbots (as putatively therapeutic devices) may all be things with which probation and prison services have to contend (see, for general arguments about the ethics of building and using sexbots, Devlin 2018, Kleeman 2020).

2.9 Ultimately a hard and fast distinction between AIs in general and robots in particular cannot be made on the grounds of the latter's mobility and (optional) humanoid appearance. "Robots" remains a colloquial term for all "intelligent machines" (and thus for all AIs), mobile or not. The abbreviated terms "bot" (or "bots") is widely used to refer to automated accounts on online platforms (like Twitter) which appear to be human users, participating in online conversations or searches, but which are actually not. Singly or in orchestrated swarms, many bots are malign, intended to distort or thwart unfolding conversations, or to offer mass fake support for political positions. Attempts are continually made by platform managers to detect and remove these, but advertisers also use bots, less controversially, to "like" or inform customers comments and choices in respect of their products and brands. Online bots are one of the main ways in which human-machine identities are continually blurred: both the bots and bot swarms themselves, and the software which monitors them, are AI's.

2.10 Used in connection with AIs and Robots, words like "intelligence", "reasoning", "autonomy" - names given to human characteristics - are merely metaphors. AI systems can accomplish tasks that were once solely the preserve of humans, but they do not "see" or "think" like humans. They identify patterns, correlations and formal relationships from a matrix of symbols (which constitute data - visual, textual and/or aural) in ways that enable problem solving at unprecedented speed, in ways that are useful to humans. They do not "understand" what the symbols mean to humans, or even what "meaning" means. When people are interacting with them (visually and verbally) they can - if they have been trained to do so - recognise signs (facial movements, tone of voice, choice of words) which signify the emotions that the individual is probably experiencing, and can be programmed to respond appropriately, as a sensitive human might do. Thus, in the form of a chatbot, an AI can provide a focussed counselling service. But an AI cannot "feel" in the way that a sentient biological creature can. It can *simulate* empathy, based on data it has acquired and processed, but it cannot *experience* empathy. Some AIs are deliberately manufactured to mimic human responses as closely as possible - companion robots - precisely to assure the people who interact with them that they are cared for and being helped (and there are obvious implications for prisons and probation in this). AI's are not self-conscious: they can learn and change, within set parameters, but they have no "self". AIs do not - cannot - possess or exercise the full reflexive, rational autonomy that human beings are (potentially) capable of. They are *non-human.* While it may be sensible and even legitimate, on the grounds of utility and efficiency, to replace or augment some human labour with an AI, there are some occupations where the qualities of a) emotional sensitivity; b) wisdom, memory and experience and/or c) sophisticated moral judgement are both integral to the performance of the work (at its best) and valued by the recipients of such endeavours. Notwithstanding claims that AI can, or will in time, replicate all these qualities,

the intrinsic value of "the human" must surely be given due consideration in all contexts where plans to dispense with it are under consideration. The Recommendation will return to this, but must first turn to a broad overview of the social implication of AI, as described in key European documents.

3. What Are the Social Implications of Artificial Intelligence (AI)?

3.1 The disagreements among scientists and technologists about AIs potential and limits may derive from philosophical beliefs, ideological preferences and/or the constraints of the commercial and political cultures in which they work. Add to this the voices of other powerful stakeholders in AI's possible futures - corporations and governments - and it clear that all *public definitions* of AI (beyond the narrowly technical) will inevitably be highly normative, never merely descriptive. They reflect the social purposes to which commercial and political authorities will put AI. They are invariably infused with views on scientific realism (what is possible?), good governance (what is politically desirable?) and even geopolitical assumptions (what is economically prudent or deemed necessary to compete internationally?).

3.2 The High Level Expert Group align their hopes for AI with earlier European documentation on it, promoting a broad optimistic vision of the social - liberal/democratic - transformation that AI will bring about. The HLEG recognise that the uses and purposes of AI need to be *shaped*, not left to themselves (or "the market"), and they are confident that such "shaping" is politically feasible - they see themselves as part of that enterprise. They position themselves thus:

> We believe that AI has the potential to significantly transform society. AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation. In particular, AI systems can help to facilitate the achievement of the UN's Sustainable Development Goals, such as promoting gender balance and tackling climate change, rationalising our use of natural resources, enhancing our health, mobility and production processes, and supporting how we monitor progress against sustainability and social cohesion indicators (High Level Expert Group 2019:4).

3.3 The legitimacy of this technologically-driven transformation will depend, the HLEG believe, on whether AI is "Trustworthy". They make this the cornerstone of their ethical argument . For AI to be "Trustworthy" it must be

a) legally regulated
b) ethically defensible
c) technically robust and fit for purpose.

3.4 The three elements are entwined, but the HLEG is concerned primarily with ethics: *legal* decisions about AI will be made at a higher level, derived from pre-existing European law, treaties and international agreements, while *robustness* is largely a socio-technical question about the expertise and integrity of engineers and programmers (although ethical principles can be coded and should be designed into machines). For HLEG, thinking through an ethical position on AI is sufficient. They begin - somewhat abstractly - with the claim that the purpose of "Trustworthy AI" should be "serving humanity":

> AI systems need to be **human-centric**, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom. While offering great opportunities, AI systems also give rise to certain risks that must be handled appropriately and proportionately. We now have an important window of opportunity to shape their development. We want to ensure that we can trust the socio-technical environments in which they are embedded. We also want producers of AI systems to get a competitive advantage by embedding Trustworthy AI in their products and services. This entails seeking to **maximise the benefits of AI systems** while at the same time **preventing and minimising their risks**. (High Level Expert Group 2019:4 emphasis in original)

3.5 All European documentation on AI recognises - alongside its manifest benefits - the technical risks that deploying it in administrative infrastructures will almost certainly pose. The HLEG has a somewhat selective understanding of these risks. It sees the three main ones as:

a) threats posed by criminal/organised crime uses of AI inside and outside European territories;

b) the vulnerability of AI-managed infrastructures and institutions to "cyber-attack" from within and outside Europe;

c) the risks to viable democratic elections and social cohesion by partisan interests using AI-driven voter manipulation and deepfake technology.

The HLEG sees three more risks: d) the creation and use of autonomous weapons; e) security service's engaging in "bulk surveillance" of citizens' use of the internet; and f) the advent of "predictive policing", but takes a more ambivalent view of them. These are all AI-related activities that criminal justice and security agencies within Europe might, or already do, undertake and a confident assumption is made by HLEG that they can and will be ethically regulated. There are ethicists who take a more critical view of all of them, and "predictive policing" in particular - the use of data and algorithms to say where and when *imminent* reoffending (pre-crime?) may occur, so that pre-emptive action can be taken - raises issue which will almost certainly spill-over into the prison and probation sector (Arrigo and Sellers 2021).

3.6 The HLEG's "human-centric" approach does not extend to the level of AI's specific impact on the cognitive, interactive and relational work done by *human services* organisations, which probation and imprisonment (notwithstanding the locks, bolts and bars which feature in the latter) can be characterised as. The HLEG do not mention the possibility of "technological unemployment". In his own overview of the risks and side-effects of AI, John Tasioulas, Director of the Institute of Ethics of AI at the University of Oxford, acknowledges all those noted by HLEG, but adds three more which HLEG seems to see as acceptable collateral consequences of AI rather than risks to be pre-empted, although all are extremely pertinent to managers and employees in probation and prison services.

a) unemployment - the prospect of intelligent machines taking over core professional tasks - including cognitive and affective tasks - from human workers (as well as back-office functions).

b) the atrophying of certain human skills when AI replaces or augments human workers - the withering away of certain occupational practices and "embodied knowledge" when machines can do this in lieu of people. Most of us, I expect, would accept that with the advent of instant access to vast searchable databases we have less need to remember things "in our heads". Our attention spans may have shrunk. Mental and institutional expectations of how fast tasks can be completed and how quickly behaviour can change may have risen, because technology has raised the threshold of what can be done at speed.

c) the instrumentalising or degrading of our relationships with human beings if, instead of dealing with them on a genuinely interpersonal basis, we more and more mediate contact with them via machines, which collect and codify data on them in the course of every encounter (or even constantly, if they are monitored with tracking devices). In a professional, bureaucratic or supervisory context this data gathering is inherently asymmetrical, "we" can gather data on "them", "they" cannot gather data on "us". They become "computable bodies". The picture that authorities have of them as "data doubles" may seem more real - and more readily actionable - than the sense we have of them as real human beings.

3.7 There is a further potential social consequence of AI for work-intensified workplace surveillance - which even Tasioulsas does not mention. The European documentation used here to inform the PC-CP project pays only minimal attention to it but, as ever, there is a belief that it can or will be effectively and ethically regulated and therefore little anxiety about it. The monitoring of employee's performance and productivity in factories and offices, by means of periodic human inspections, and usually set against some specified metrics is nothing new. This managerial function can now be massively augmented if sensors (wearable and/or embedded in buildings and equipment) and software systems - not necessarily full AIs - are used to gather, analyse and compare data with an unprecedented degree of granularity. Workers themselves become "computable bodies", mirroring what some workers then do to the clients they supervise. Whatever the long term impact of AI on the nature of work in probation and prison services, it is certain, in the interim, while humans are still employed, that because of AI's capabilities they will be subject more surveillance. This complicates - some may even say *compromises* - the idea of "human centric" applications of AI. It is something that prison and probation services - and the PC-CP Recommendation - need to address.

AI, Automation and the Unemployment Question

3.8 There has been a long and complicated debate about the implications of automation and "robots" for human employment. In the early part of the twentieth century new kinds of machinery - and eventually automated machinery - progressively replaced thousands of manual jobs. Certain trades died away: new technologies obviated them. Fears about mass unemployment have accompanied each wave of automation, but in the past these have largely been confounded: new technologies increased economic productivity and created new occupations for which those rendered workless could retrain. This may or may not have consoled people who lost hitherto secure and well-paid jobs in industry and found that the jobs in the service sector (or "the gig economy") for which they have retrained were precarious, less secure and less well paid.

In respect of the likely impact of AI on cognitive jobs in the professions there are basically two views (Ford 2015, Susskind 2020). Firstly, that this time it will be harrowingly different and that there will be *no replacement jobs* for those displaced by technological innovation. Certain professions will simply vanish over time, or be reconfigured beyond all recognition. Certain industries will require

neither front line workers, nor middle managers - only highly paid directors and technicians. It is in the context of this "structural technological unemployment" (as economists call it) that serious debates have begun about Universal Basic Income and four day working weeks, more seriously in some countries than others. Secondly, the new wave of AI-based automation will be no different this time than in the past: while there will be collateral damage, with some professions losing out to AI, economic growth will once again create new ways of making a living that are still over the horizon, as yet unimaginable. But who decides who will lose out, and on what terms?

3.9 Neither of the above scenarios are particularly congenial to Probation Services. As state-based, or state-funded professions Probation Services are not intrinsically immune to technological trends. Indeed, the opposite may be true, given the ease with which "crime control" and "law enforcement" can be politicised. At the very least, Probation Services are certain to become much more data-driven organisations. To this end, they will need to train frontline workers differently and, quite likely, to employ different cadres of staff. The emphasis on gathering data (the more the better) may drive closer ties with other criminal justice agencies, including the police, jeopardising organisational boundaries that were once unambiguously protected by confidentially requirements. More data-sharing, to create larger data sets, will become inevitable if "predictivity" (of reoffending) becomes an overt, and widely-shared goal in corrections. Smaller countries may seek international collaborations to boost the size of the data sets available to them.

What will happen to the *human* supervision of offenders and the relational work that is often claimed to be the essence of probation? Even a decade ago, the cognitive and affective skills considered the hallmark of a good probation officer would have been thought immune to automation. Not so now - at least in principle. While there are clear political choices to be made about how far AI-based automation in probation goes, it seems unlikely - even without the Covid pandemic's impact on the uptake of remote communication - that probation can or will continue unaugmented by new forms of technological intervention. Smartphones may become an important mediator of communication in the officer -service user relationship - as well as being the generator of hitherto unavailable data about their owner's lifestyle, thoughts and (obviously) locations, by using smartphones as a proxy tracking device. Chatbots may replace some human communication with offenders, as they have already done in customer service in many businesses. Supervision programmes may be undertaken online.

3.10 How will prison services fare? There is nothing about AI in itself which makes the use of prison less likely - it may be presented as a "transformational technology" but precisely *what* it transforms and what it leaves untouched, or simply reconfigured, is a political and commercial decision. The scale on which prisons are used will continue irrespective of AI. Their demographics, the preponderance of poor and disadvantaged people in them, will not be altered by AI. Prisons may well become a special kind of "smart building", and the delivery of some - perhaps many - services to prisoners may be automated. Their day to day management, drawing on coordinated, real-time data streams about all aspects of behaviour in the institution, may be made more cost-efficient. AI is as likely - perhaps more likely - to support aspects of prison security (not necessarily in the form of robot guards) as the delivery of rehabilitative services, although there is notionally a connection between the two. If automation means fewer human staff need to be involved in maintaining security, does that free them up to be more involved in personal rehabilitation - or does that get automated too. Like many aspects of implementing AI in managerial settings, its actual impact - whatever the intentions - is an empirical question. Right now, It's too early to say.

3.11. The unemployment question is often deflected by the argument, plausible in itself, that AI will not best be used to *replace* humans, but to *assist* them. The essence of such "human in the loop" arguments are that the tasks and purposes of professionals (knowledge workers) will remain much the same, but with insights from AI they will be done more efficiently. The professionals will require some new skills - and support from technical assistants - but the degree of organisational and cultural change will not be unduly dramatic. Susskind and Susskind (2015) - a British lawyer and a social policy analyst respectively - are actually sceptical about the assistive argument. They believe that assistive uses of AI to inform human-decision making will be merely temporary, a stop-gap en route to the more or less full replacement of much human knowledge work by more efficient AI's. There may well be genuine value in human-AI collaboration in decision-making, but emphasising its *assistive uses* may simply be a way of offering false reassurance about AI's longer term implications, and a devious way of marketing it.

3.12 The probable limitations of the assistive model of AI use has indeed been revealed by existing experience in workplace recruitment, where in corporate settings, automated systems now widely dominate this field. They have already - and quickly - gone beyond the merely assistive, to the replacement stage. Human eyes never see the bulk of job-seeker's applications, which are initially scanned by algorithms trained to make judgements about which qualities will make a good employee. Only the final shortlist gets human attention. And then the ethical issue arises: even if we assume that algorithms are just as a good as, or even better than, a human recruiter, is it legitimate - from the standpoint of the applicant - to exclude them from real human consideration? Are they not - literally - being appraised by a machine on what for them is arguably a very fateful moment - getting or not getting a job, with all its implications for income, security and fulfilment? How, in a world where algorithmic governance becomes pervasive and normal, will people experience this? Will they feel disrespected? Will they feel that their dignity as person has been affronted? Will they recoil and complain? Or will they simply grow used to it, and accept it? What will be the effect on personal identity of being managed by algorithms, in varying degrees, in different social settings? Will there be racial, ethnic and gender differences? We do not yet know, but it is a matter of some urgency - if AI is to expand on the scale envisaged in Europe - that we make the effort to find out.

3.13 There is another aspect to the "human in the loop" as the supposed ideal arrangement for people and AI, highlighted by Kate Crawford (2021), which questions whether, from the workers' standpoint, such arrangements can ever be freely chosen:

> The common refrain for the expansion of AI systems is that we living in a time of beneficial AI-human collaboration. But this collaboration is not fairly negotiated. The terms are based on a significant power asymmetry - is there ever a choice *not* to collaborate with algorithmic systems? When a company introduces a new AI platform workers are rarely allowed to opt out. This is less a collaboration than a forced engagement, where workers are expected to re-skill, keep up and unquestioningly accept each new technical development (Crawford 2021:58)

4. What is "the Human" in the age of AI

4.1 In making fair-minded decisions about what human activity might be replaced by (or even augmented) by AI/robot it is important to be clear about the qualities that a human person brings to the activity *and* what the human recipients of that task think or believe about it. This is a complicated question in a number of ways, but seems particularly pertinent to human service organisations. It cannot be settled directly by reference to human rights or terms like "human centric" - there is, as yet, no human right to NOT be replaced by a machine. Answers to the question usually involve reference to qualities like care, responsibility, empathy and solidarity which people can feel and show towards each other, in a way that a machine cannot - although it can mimic them in ways which a human recipient may find satisfying. As John Tasioulas puts it

> There is a valuable human solidarity and reciprocity - human beings recognising each other as fellow human beings and forming their attitudes and decisions about each other on that basis - that is lost in the context of dehumanised, fully automated decision-making (Tasioulas 2019)

4.2 One difficulty with "human" is its extreme variability. Humans can indeed by caring, responsible and empathetic , as well as skilful, experienced, astute, moral and brave. But they can also be unkind, insensitive and indifferent as well as unskilful, lazy, thoughtless, tired and burnt-out. When considering whether a machine should replace a human for a particular task we have to assume that the human we are speaking of is good, reliable character and trained to be the best they can be. Human (professional) training may be expensive - a machine may well be cheaper in the long run, but we cannot accept that financial considerations alone should determine when a machine replaces a human. If the training is proven to be effective, investment in training should continue.

4.3 Some occupations that humans can be trained to do to a high standard - bomb disposal work, clearing hazardous waste - are nonetheless so dangerous as to be routinely life threatening. Transferring such work exclusively to machines is easily justified. But what of robot carers for disabled and elderly people? Should human carers be relieved of what is often emotionally distressing - and very occasionally abusive - work? Who decides? What do the recipients of the robot's service think? Is it dehumanising for an elderly person to be bathed or toileted by a competent machine, or does the welcome impersonality of this arrangement spare them the embarrassment that the presence of a fellow human carer might cause? These are questions of which people - elderly and infirm people, relatives, the owners of care homes and the general public - may over time change their mind about, depending on their familiarity with the quality of human care and the capabilities/sensitivities of future robots.

4.4 Driving, medical diagnosis and sentencing criminals, where matters of life, liberty or justice are at stake, are regularly cited as the most challenging areas for replacing a trained and experienced human with an AI. All three represent significant changes, and may generate varying degrees of social anxiety among a range of concerned parties. They are not in fact of the same order. Autonomous vehicles are certain to come, dependent on a vast network of AI-controlled sensors - and some workers at least, who drive for a living, will lose their jobs. It has been repeatedly proven that AIs can outperform even highly trained doctors in some areas of medical diagnosis - they can collate, read and process information on symptoms far faster, gaining precious time for the patient. It too is certain to become a regular feature of medical practice. Sentencing convicted criminals is different. Even Susskind and Susskind (2015) concede that although this could be done by a machine, trained to follow methodical legal reasoning and able to draw at speed on

an array of precedents, they recognise the engrained reluctance in society at large to take the solemn practice of judgement away from a human being. The model of sentencing being used here is that which requires "treating like cases alike", based primarily on the nature of the offence. It values above all else consistency across time and territory: offences of similar character and seriousness should be dealt with in similar ways. This is what (in one view) fairness means. In principle a trained AI, drawing on a vast database of sentencing decisions, could be good at judging in this way because it would require very little attention to the individual character and circumstances of the convicted person. But there are other, humanly valuable and morally defensible approaches to sentencing where an AI would probably fare less well. In one such alternative approach

> …. there is value in a merciful judge being able to express their values and their character by, for example, choosing a more lenient sentence from a range of eligible options. There is value in a criminal justice system that offers offenders the possibility of discretionary mercy. The granting of mercy here is a kind of gift-giving, by one person to another, perhaps reflecting the hope of the former that the latter has genuinely repented of their past wrong doing. In the case of the automated sentencing system, this value would be severely curtailed or eliminated. The robot is not an individual, with values and a character if its own, who can respond to the offender's plea for mercy as one human being to another, choosing a more lenient sentence when harsher one is also rationally open (Tasioulas 2019)

4.5 Decisions on life and liberty, whose outcomes may cause suffering even if they simultaneously serve justice, lack a certain moral weight if they are made by an agent which cannot take proper responsibility for them. A defendant in court may feel it important to know that they are being appraised and judged by a human being who can at least try to understand and empathise and reach a decision based on a personal assessment of how we should be treated. In principle, a human judge could give an explanation of how they had reasoned, even if in practice they do not always do so. A sophisticated AI, trained via machine learning, may not be able to make the process by which it arrived at a decision transparent and explicable, other than by saying that it derived logically from the contents of a sentencing database. How satisfying that explanation might be to a defendant is moot.

4.6 The sentencing example in the AI-human debate is instructive, because it can provide preliminary guidance on how we might appraise prospective uses of AI in probation (and prison) services, *because they execute sentences of the court.* The same logic that seeks to preserve the human element in sentencing might also be used to make the case for preserving the human element in probation practice. It requires further defence, but this is a good place to start, when so many aspects of what we nowadays understand as probation could feasibly be subject to automation. There is no iron law of nature or society which says Probation Services must forever exist in the forms with which modern Europeans are familiar. The culture and methods of the older Services have already changed several times over the decades and adapted to new technologies - eg computerised record keeping, electronic monitoring - and there is no reason why they could not adapt again. But when does adaptation become so transformational that the existing ethos and character of an organisation dissipates beyond recognition and repair, no matter how much the retention of familiar nomenclature - calling it a probation service - is used to imply a spurious continuity?

4.7 Contemporary probation supervision, in its essence, is based on iterative processes of assessment, implementation and review, all undertaken by variously trained personnel, within a management structure. All stages of the process require some degree of conversation and dialogue. Risk/need assessment is becoming progressively more automated - and "assisted decision support systems" are coming on the market to take this further. The judgements made by a human probation officer about a client's risk level can already be informed by a machine. But what about the encounter itself? Let us consider a thought experiment in respect of conversation and dialogue, the once face-to-face "heart" of probation supervision, sometimes mediated by telephone. The more structured, focused and formulaic these conversations are required to be (as in customer service) the more easily a chatbot could routinely undertake them. It could counsel a client individually (with cognitive behavioural techniques), or lead them though a scripted offending behaviour programme, or a scripted restorative justice encounter (if the victim consented). It could award points and rewards to incentivise compliance. With machine learning, chatbots would improve over time, passing a client on to a human superviser only when a certain threshold of concern had been passed, eg a breach decision. Would a client care one way or the other whether the voice that was advising or questioning him/her was a machine or a human? Should the client always be told? Why? AI's have already been trained to write short pieces of journalism - football reports, for example, culling data from internet. How long before an AI is required to write a pre-sentence report - or several dozen - simultaneously?

In a recent article, two veteran probation commentators - one English, one Dutch (Pitts and Tigges 2021) - speculate bravely on what probation practice may - or could - look like in 2030. They base their defence of what it could - and in their view *should* - be on the wealth of international empirical evidence of "what works" to rehabilitate, reintegrate and encourage desistance. Promoting "what works", and researching what else may work - is indeed a vital task if probation as it has been traditionally understood is to be preserved and extended. Framed another way, the currently empirical literature on "what works" describes what is humanly possible for properly trained probation staff. Whether this *sufficient* to preserve probation is moot, as Pitts and Tigges concede - but it is *necessary* to say it. They recognise that there will probably be national variations in the political fortunes of European probation services, and fear that "populist punitiveness" in some may derail best practice in probation. They do not directly address the implications of AI for probation, which poses a different kind of challenge than the one posed by populism. The current "what works" literature is premised on a humanist paradigm in which trained professionals - people - are the primary resource. If that paradigm shifts to something more accommodating of AI, the worth of the "what works" literature - premised on old ways of doing things - is diminished, and loses its force as a basis for preserving recognisable forms of probation.

5. AI in Prisons

> At the latest Annual Conference of Directors of Prison and Probation services [in Europe] we saw an interesting contribution from Singapore, a fully automated prison. That prison does not have guards: AI and robots do everything. I don't think that's the model that everyone would want but artificial intelligence is already used by some prison administrations to allocate cells to prisoners and to decide who should share a cell with whom, who would be compatible, who would pose a risk. Jan Kleijssen, (2021) Director of the information Society and Action Against Crime Directorate, Council of Europe

5.1 Historically, prisons in the West - from the 18th century on - were literally designed and built to maximise - at least to optimise - the observation of prisoners. The immediate benefit of constantly watching was (it was assumed) compliant behaviour by prisoners, but it had the added advantage of generating information about them (via observation and dialogue) that prison authorities used to differentiate and classify them as particular "criminal types" and (sometimes) to predict recidivism. Observation and classification never precluded the use of locks, bolts and bars, and easily co-existed with a deterrent philosophy. They did however contribute to the growth of a scientific approach to the rehabilitation of offenders, premised on understanding their dispositions, motivation and behaviour. The advent of digital technologies in prisons and probation has already created new ways of generating information - data - about offenders, and data is what AI systems thrive on. While AI is as yet far from ubiquitous in European prisons - or even elsewhere - digitisation is manifestly a growing trend, and the "cell allocation AI" to which Jan Kleijssen refers above would require a comprehensive prisoner database and trained algorithms to perform that hitherto human function efficiently and effectively.

5.2 Champions of "smart prisons" - as prisons with digital capabilities are coming to be called - have readily conceded that digitisation assists with security - technological innovation has always contributed to that - but more usually present them as an advance in rehabilitation and prisoner empowerment. At the very least they enable serving prisoners to access and use the internet in classrooms - something which traditional prisons cut them off from - in order that they can retain, update or acquire employability skills that they will require after release. Increasingly "smart prisons" are more ambitious than this. Pia Puolakka (2021:51) describes one such example in Finland. The roots of it lay in legislation in 2015 enabling "prisoners to gain digital access to social educational and healthcare services and to use video calls to contact family and friends". This guaranteed all prisoners "access to the same civil services enjoyed by other citizens, in line with the principles of normality and equality"

5.3 In the late 20th century, new remote surveillance technologies - CCTV, together with automated access controls - obviated the need for a physical prison architecture that maximised direct observation of inmates by the authorities. In the 21st century, yet newer surveillance technologies have increased the reach and intensity of watching in unprecedented ways, automating the gathering of observational and behavioural "data" on wings, in workshops , even in cells - often in "real-time", sometimes managed by AI - with a view to managing individuals and maintaining collective order. These still largely experimental technologies - which, if taken to extremes, raise the prospect of a "fully automated penal institution" - have profound implications for the way future (and indeed present) prisons are managed, even if they are not, in fact, developed to their fullest capability. Penal Reform International notes that in 2020:

> In places of detention, including prisons, most of the primary functions of AI-led systems are related to security. New systems are being used to alert staff to behaviour or activity by people in prison that the system registers as 'abnormal' or 'suspicious'. (Penal Reform International 2021: 47).

5.4 They give three experimental, and one established, example of AI-based monitoring. In one Hong Kong prison, prisoners are required to wear a wrist band which monitors their heart rate, from which aspects of their behaviour can be inferred. In one Chinese prison hidden cameras and sensors in cells generate daily reports about each inmate. In one UK prison AI-equipped cameras monitor people entering it to detect contraband, drugs and weapons, by matching their movements and behaviours to a notion of "suspiciousness" embedded into algorithms. In the US, several states use

AI to monitor inmate phone calls, using "speech recognition, semantic analytics, and machine learning software to build databases of searchable words, and patterns to detect illegal activity". (Penal Reform International 2021:47). In her overview of digitisation (including some AI programmes) in seventy Indian prisons Ashna Devaprasad (2021) highlights the vertiginous contrast between the deployment of sophisticated surveillance technologies developed by elite Indian start-ups, and the poverty and illiteracy of the many prisoners who find themselves subject to them. The material contrast may not be so extreme in some western countries as in India, but the asymmetry of power - and education - may still be great.

5.5 The deployment of robot prison guards - mobile machines which can patrol prison premises - are also on the agenda of some companies and some governments, admittedly (as far as one can tell), not in Europe. Hong Kong has been piloting a robot equipped with a camera and microphone that can send images back to human staff in a control room and enable two-way communication with people with whom it is in proximity. South Korea introduced "robotic guards" in 2012 to reduce human correctional officers' workloads.

5.6 The Changi Prison Complex in Singapore, mentioned by Jan Kleijssen, is seen in some commercial quarters as the pinnacle of advanced technological management in corrections. Indeed, it is, but more interesting, perhaps, than the prison itself, is the state-corporate milieu in which this development arose. According to its website, HTX is a commercial body that describes itself as a Science and Technology Agency which aims to "transform the Homeland Security landscape and keep Singapore safe". Its publicity continually emphasises the value of safety, security and stability to the citizens and businesspeople of Singapore. The breadth of the technologies HTX makes use of is enormous: "biometrics, chemical, biological, radiological, nuclear and explosives threats, cybersecurity, artificial intelligence, forensics, robotics, automation and unmanned systems, and surveillance". The HTX Sense-Making and Surveillance Centre has been applying safer cities technologies in the city-state and working with the police and Immigration and Checkpoints Authority, developing "different surveillance systems leveraging on state-of-the-art Artificial Intelligence technologies. Integrating our technologies with the front line creates a force multiplier that enables the front line to safeguard out homeland and border security more effectively". In addition, the Centre says of itself:

> Our work can also be extended to improve the safety and efficiency of prison cells. We are collaborating with the Singapore Prison Service to develop a Human Behaviour Detection System (HBEDS) as part of the SPS 'Prison Without Guards' initiative. The HBEDS is able to pick up patterns of abnormal behaviour in prison cells for early intervention, keeping the inmates safe with the aid of technology.

5.7 Does all this seem a step too far for Europe, or will we choose - or be pushed - into having a serious debate about doing it here? Changi seems several steps beyond the development of "smart prisons" in Europe, the product of an "obviously" different penal - and political - culture. "Smart prisons" in Europe are currently understood to be serving rights-respecting, rehabilitative purposes, but will this be honoured everywhere, and can this line be held?

5.8 Arguments about the limits of automation in European prisons may first come to a head in respect of "special units" - prisons within prisons - for the most dangerous offenders. These do not currently preclude human encounters, but they may be kept to a minimum for the sake of staff safely. Educational and therapeutic interventions may be provided face to face, but under

conditions of tight security. The availability of in-cell communication with prison managers and online learning facilities, including therapeutic programmes, plus video contact with families, may well have some advantages - more enriching experiences - for such secluded prisoners, but possibly at the expense of a further decrease in human contact. That may shift the regimes of special units closer to the conditions prevailing in solitary confinement, without making them identical. "Meaningful human contact" - or the lack thereof - is an established concept in the human rights instruments used in debates on the legitimacy or otherwise of solitary confinement. It is not inconceivable that "meaningful human contact" will need to be debated in the context of highly digitised prisons - particularly special units - which, whilst permitting technologically mediated sociability to occur, still places limits on actual human presence.

6. AI in Probation Settings

6.1 The Anglo-American - and Dutch - roots of probation and parole lie firmly within a rehabilitative and reintegrative tradition, whose traces remain - in greater or lesser degree in different jurisdictions - even as probation and parole have come to be defined as an arm of law enforcement. In the late 20th century, its professional knowledge base became less exclusively grounded in social work, and drew more from psychology, sociology, criminology, law and moral philosophy. Even as agencies became more bureaucratised and managerial, reducing the discretion of individual officers, its mode of working remained interpersonal and relational, premised on the ideal of face-to-face contact. Technology, whether for control or rehabilitation, figured little in probation's repertoire until the advent, in the 1980s and 90s, of **computerised risk assessments** - to predict risk of reoffending and degree of harm - and **electronic monitoring**. Both of these are becoming, or could become, integrated with AI.

6.2 Electronic Monitoring (EM) was arguably probation's greatest technological challenge, first radio frequency to enforce curfews and house arrest, later GPS to enable tracking and exclusion zones, most recently (on a more limited scale) remote alcohol monitoring. It was initially resisted by probation services, especially but not only in jurisdictions whose governments controversially portrayed remote surveillance as technically (or ideologically) superior to personal supervision, but eventually it was accepted as a useful tool. Extensive academic and policy debates took place on EM's merits as a standalone or integrated measure. Empirical research in particular countries played some part in settling ideas about best practice and effectiveness, pushing against the idea of EM as punishment and pressing the idea that it was better used as a form of control which, in moderation, could support rehabilitative endeavours. The CEP played a significant role in debating and spreading good practice across Europe, learning from developments in the US and engaging with the commercial manufacturers (some of whom sponsored CEP EM events). The Council of Europe PC-CP made an ethical recommendation on EM in 2014. There are some small similarities in the PC-CP's previous work on EM with its current work on AI - not least is respect of regulating the private sector - but the key differences are vast: European institutions are driving elements of AI governance across all public services, and the consequences of such transformation, if that is what it proves to be, are way in excess of EM's limited impact in criminal justice.

6.3 Penal Reform International (2021) summarises the increasing use of new technologies in probation and community corrections (including parole) settings, citing EM itself, as well as telephone check-ins and biometric check-in kiosks for automating probation reporting conditions. These too had roots in the 1990s, but never had the cachet or public profile of EM and to a degree remain under the radar. These are not (yet) universal and in themselves they can easily be portrayed as pragmatic and innocuous technologies. But they could grow, and as sources of

behavioural data they could contribute to - and be managed by - AI systems in community corrections, if such things emerge. Even in their present form, PRI is sanguine:

> Most of these technologies tend to place excessive emphasis on control and security, rather than on rehabilitation, with increased and enhanced controlling of the probationer's strict compliance with conditions of a measure or sanction and less human contact and support beyond those rules (Penal Reform International 2021).

PRI's observation that technologies used in, and being developed for, community corrections are largely about control and security does not preclude the use of technologies for rehabilitative purposes - it just identifies the dominant trend.

6.4 An ostensibly more helpful technological development in probation - a new form of EM, though it may not be what it seems - concerns smartphones, both for use as tracking devices (a mainly US development) and as the basis of "probation with apps" (developing in both Europe and the US). These were briefly two separate developments, which, while still *potentially* separate, can, as US experience is showing, converge very quickly. Smartphone EM arose as a commercial proposition in the established US "EM industry" and among start-ups on the fringes of it, simply because smartphones were trackable, more discreet than wearable ankle bracelets, and many (though not all) offenders already possessed them, and were comfortable doing so. "Probation with apps" arose in the context of ever increasing digitisation across correctional and welfare services, an increasingly mobile, tech-savvy workforce and digital governance more generally: they have been as much a modernising demand from *within* services themselves as expressions of "outside" commercial interests (although they were ever present, always in dialogue, awaiting cues). This App movement prioritises the communicative capabilities of smartphones to promote offender engagement and the educational uses of apps to support rehabilitation and desistance, and, initially, was not discursively framed as a form of "electronic monitoring" (which was semantically associated with location monitoring and surveillance). Equivalent developments with apps have been taking place in social work, mental health and drug treatment.

6.5 Smartphone EM - monitoring all aspects of an offender's phone use, plus their response to probation apps, as well as tracking them - extracts much more data from offenders than earlier forms of location monitoring, and therein lies a potential link to AI. In the USA, the National Institute for Justice (NIJ) in May 2019 sought bids from commerce and academia for experimental projects which applied "advances in artificial intelligence (AI) to promote the successful re-entry of offenders under community supervision in the United States", ideally resulting in sustainable operational programmes (NIJ 2019a:4). It presented this both managerially, as a means of increasing efficiency in community supervision agencies struggling with high caseloads, and also substantively, as a qualitative improvement in real-time responsivity to offenders - *something new, hitherto unachievable by human supervisors*. AI, it said, would enable "the degree of supervision *and immediacy* that may be required to help guide an individual's use of services and programs to assist their successful re-entry into their communities" (p5 *emphasis added*).

Three types of project were envisaged. Firstly, situationally dependent, real-time updates to an offender's risk-need-responsivity (RNR) assessment. Secondly, mobile service delivery to offenders (via smartphones) of personalised rehabilitation resources and reinforcement of compliance outside of treatment and education sessions. Thirdly, intelligent offender tracking, building on data from existing GPS monitoring systems. It then outlined what AI-enhanced real-time responsivity

might look like, as if its normality and desirability were already obvious, indicating how smartphones could alter the character of communication with an offender quite fundamentally:

> Using the geospatial and temporal data from [GPS] devices, coupled with an understanding of how the attributes of the places an offender visits interact with their recidivism risks and how that changes with the time of the visit, AI can detect (and possibly predict) potentially risky behaviour. Based on the nature of the event, an AI could autonomously take a number of different actions to address the risk. Those might include, alerting the supervising officer or a mentor, *or initiating a chat bot system through an offender's mobile device that is trained to de-escalate situations.* AI-initiated actions may also include notifying the offender through their mobile device to suggest a cooling-off period in a safe space, or to promote behaviour modification techniques (2019 p5-6 *emphasis added*).

6.6 In September 2019 Purdue University and Tippecanoe County, Indiana were awarded $2million by the NIJ to create, manage and evaluate an "AI-based support and monitoring system" (AI-SMS) for 250 high risk adult offenders over four and half years: "Offenders will be provided with a smartphone and a tracking-health-related wearable device or bracelet", Purdue wrote, while "officers and practitioners (e.g., clinicians) and caseworkers ….. will be provided with smartphones/tablets with specific dashboards (user interfaces) that would be used to communicate with the offender" (NIJ 2019b). Ostensibly, the project took the known evidence-base on desistance and re-entry, and then insinuated - without compelling evidence - that such humanly desirable goals might *not even be feasible* without digital augmentation.

6.7 This US project is ongoing, and doubtless there will be international interest in its evaluation. Recidivism prediction already seems to have become a more prominent concern. It may, at present, be no more than an AI-augmented extension of the existing capabilities of Risk Assessment Instruments, but latent within what is being proposed is a process whose endpoint, given the commercial interests in play, may lie well beyond the NIJ's own near future expectations of what the impact will actually be. Will such developments begin to occur in Europe?

Virtual reality: a new probation intervention?

6.8 There is a danger, when talking about the implications of AI, to concentrate on it as a "thing in itself", and to concentrate less on changes to professional practice that it - running invisibly in the background - makes possible. The use of immersive Virtual Reality (VR) environments is one instance of this. They are already used for training in commercial, sporting, educational and military settings, quite apart for its use as a form of entertainment. VR can simulate, via a wearable headset, a psychological sense of presence and involvement (for one or more people simultaneously) in a seeming three-dimensional space, in which one can, while remaining in one place physically, have the sensation of "moving around". Users can represent themselves as avatars in the virtual space and interact, on the headset screen, with other users who do not share the same physical locations as themselves. Hand tracking and gesture control in VR platforms enable users to manipulate virtual objects and mimic actual bodily movements, including facial expressions, within the virtual environment. Back in 2014, Mark Zuckerberg, CEO of Facebook, conceived of VR as "the next major computer platform" which would come to affect the way people in general connect, work and socialise (quoted in Rose 2016). Hardware remains expensive, and has not yet become

that, but there has been immense commercial investment in it - especially, as expected, from advertisers, keen to further influence consumer behaviour - and new applications abound.

6.9 Under laboratory conditions, carefully crafted virtual environments have already been used therapeutically to treat anxiety, depression and PTSD. It has similarly been experimented with in criminal justice, both in professional training and as a technique of offender rehabilitation. It has been claimed to be particularly useful as a means of engaging young people already familiar with screen-based entertainment and virtual gaming environments. An advertisement in *Justice Trends* 7 notes that pilot VR programmes are being rolled out in correctional services across Europe. Specific reference is made to the use of VR in the training of refugee offenders, prison inmates and drug user rehabilitation. The advertisement continues

> VR technology makes it possible to meet the special needs of different offender populations and overcome language barriers in a practical and inexpensive way. In training, virtual reality can help compensate for the shortage of resources in prison facilities. Virtual reality is also widely helpful and effective for prison staff training. In addition, it allows motivating interactive and gamification elements, hence contributing to engagement and leading to better training outcomes. (p53)

6.10 What is usually emphasised in the promotion of VR as a training and therapeutic tool is the cognitive and affective quality - including its enjoyability - of the learning experience for the user. Through the repetition of actions in safe virtual environments physical skills can be developed or improved, decision-making skills honed and self-reflection promoted. It is on the basis of these assumed benefits that a probation officer, say, could ask or require a client to participate in an approved VR programme. In a British context, for example, Tom Gash (2020) has suggested the feasibility of combining VR and electronic monitoring:

> There is also the opportunity to make use of virtual reality for those under house arrest or curfew to support remote learning (think plumbing courses in VR headsets), or ot recreate the privations of prison by requiring a certain number of hours in headset solitude Gash 2020:75)

6.11 What is not usually part of any sales pitch with VR is the extent - and novel form - of the data that the user generates while immersed in VR. This is either downplayed, or ignored completely. While the potential therapeutic impact of experiencing VR environments is not be to be denied, as something of possible value in itself, data extraction (and analysis by AI software) is a major feature of VR. Michael Madary, researcher at Joannes Gutenberg University who with Thomas Metzinger T co-authored a (2016) *Code of Ethics for VR* has said

> The information that current marketeers can use in order to generate targeted advertising is limited to the input devices that we use: keyboards, mouse, touch screen. ….. *VR analytics offers a way to capture much more information about the interests and habits of users, information that may reveal a great deal more about what is going on in their minds*. (quoted in Rose 2016, emphasis added)

6.12 There is already an industry devoted to VR analytics - a specialised form of data analytics. Head movements, measured by a gyroscopic sensor in the head mounted display can be used to create

"heat maps" of where users look while immersed in VR. Emotion detection is a growth areas, of interest to advertisers monitoring responses to brands. Affectiva, a company spun off from MIT, offers "emotion detection as a service" to client organisations, enabling them to mine images from publicly accessible webcams and video feeds to see how faces react to certain cues. Yotta Technologies claims to record and read microexpressions and muscle movements in users faces. Madary and Metzinger anticipate that VR will eventually be able to map a complete range of human body movements, taken from a user's prolonged immersion in VR environments, creating a "kinetic fingerprint" which could be used identify a person - say, an image on CCTV - on the basis of motion and posture.

7. AI, the Private Sector and Prisons and Probation Services.

7.1 The development of AI systems and the expertise in using them largely resides in the private sector, and this Recommendation requires commercial organisations working in this field - or aspiring to introduce technologies into this field - to abide by the ethical, strategic and operational principles described here. While the boundaries between public and private sectors have grown more blurred in a number of ways since the turn of the 21st century, notably in respect of commercial sector management techniques being taken up by the public sector, and the use of consultants, the development of AI seems to be accelerating that. AI itself can be understood as a management tool. Increasingly, state agencies recruit private sector people, as employees or consultants, to steer their internal transformation programmes.

7.2 In understanding the contribution of the private sector to the work and prospects of the prisons and probation sector, it is necessary to look both at

a) the specific contracted companies who deliver services on behalf of state agencies. The European experience of using EM could illuminate this. The collection and processing of location data by international companies based outside the European territories they operate in is one issue that has emerged from EM use, and becomes more salient if that data is processed by an AI.

b) the agenda-setting aspects of the AI commercial sector as a whole, which outlines, promotes - and invests in - future visions of what a transformed criminal justice system might look like. The Commissioner for Human Rights specifically suggests that in the interests of responsible innovation and marketing examining this agenda-setting role is a vital task.

7.3 Some of the larger companies and corporations in the AI field have the resources and authority to engage with senior levels of government above the heads of particular government agencies, and may promote AI-based visions of how those agencies could be improved before the professionals in those agencies have been consulted with. The CEO of Switzerland-based EM-provider Geosatis, for example, has signalled its interest in reconfiguring its approach to EM, by using AI, which acknowledges its likely impact on the staff profile of supervising agencies:

> a new AI-powered EM system could take historical data into account and mix it with any other relevant source of data to be correlated (other offenders, a map of criminogenic zones, crime scenes, CCTV records, car speed, etc)". [This] "set of functionalities … would enable *a predictive intervention* … that allows a superviser or probation officer to identify a potential problem before it actually occurs …….

> [Progress] will not be linear: it will certainly pose challenges with regard to data protection and bias, *not to mention the implications that such EM advanced features will have on the academic-technical profile of probation agencies' staff who will be in charge of analysing, drawing conclusions and acting on them* (Demetriou 2018, emphasis added)

7.4 The enduring controversy over "privatising" criminal justice services - and in particular the obscuring of some elements of the service on the grounds of "commercial confidentiality" - acquires new dimensions in the age of AI, as Penal Reform International notes.

> Algorithms used to direct sentencing or prison classification are usually developed by private sector companies and are often considered trade secrets. This means justice actors may not understand the complex functions and removes the opportunity for a suspect, defendant or person in prison (or their legal representation) to enquire or understand the computer-generated decision (Penal Reform International 2021:47)

7.5 John Tasioulas links the obscurity of algorithmic operations in risk assessments with the agenda-setting power of private companies to paint a rather alarming picture:

> The challenges already mentioned to the just operation of algorithms are compounded by the fact often, for commercial reasons, neither the algorithm not the data on which it is trained are made public. Moreover, in the case of bottom up algorithms, it can be opaque even to the people operating the RAI precisely what algorithm is governing its activity. *This creates the ever present danger that powerful corporations may be able to shape any resulting laws in ways favourable to their interests rather than the common good* (Tasioulas 2019 emphasis added).

7.6 This points clearly to the fundamental difficulty of developing a binding ethics of AI in *any* sector of society in the coming decade - and beyond. Irrespective of what governments want from AI, the commercial manufacturers and marketers of it cannot but remain a powerful and seductive voice in all future debates about prosperity and security. In principle the power of commerce in technological innovation can be constrained. Arguably Europe has a good record of doing this. But the continuing challenge of it should not be underestimated, not least in the probation and prisons sector, where commerce is relentless in its confidence that technological innovation can disrupt and improve practice in these fields.

8. Towards Ethical Accountability in AI in the Prisons-Probation Sector

8.1 The bulk of this Recommendation has been concerned with delineating an ethics of AI appropriate to the work of prison, probation and youth justice services, and the private companies who work in this field. Implicit in the understanding of ethics developed here is the idea that they must be more than abstract principles to which an agency "merely" gives assent. They must also be capable of being translated into strategic and operational practices which have an actual bearing on the work of the agencies in question. They can then provide mechanisms with which an agency's compliance with ethical requirements can be judged, and for which they can be held accountable. Both the HLEG and the Commissioner for Human Rights acknowledge this - neither promote ethics in isolation from strategic and operational considerations. The Commissioner is, in fact, primarily

focussed on the latter, and her ten points will form the basis of what this Recommendation requires for the use of AI in the prisons-probation sector.

8.2 The tenth point in the Commissioner's list - "promoting AI literacy" - has already informed the character and terminology of this Recommendation itself. The tenth point also relates to the second point, a requirement for "public consultations" about the deployment of AI, and it is our expectation that this Recommendation will be used as a key document in any European consultations relating to the deployment of AI in the prisons-probation sector.

The Commissioner for Human Rights ten points are:

1. Human Rights Impact Assessments (HRIAs) - regular appraisal of public AI users
2. Public Consultations - open debate about intended AI deployments
3. Obligation of Member States to Facilitate the Implementation of Human Rights Standards in the Private Sector - to affect them at the design and marketing stage
4. Information and Transparency - explaining the why and how of AI processing
5. Independent Oversight - by administrative, judicial and parliamentary bodies
6. Non-discrimination and Equality - designing-in fairness and avoidance of bias
7. Data Protection and Privacy - AIs extract & process data on an unprecedented scale
8. Freedom of Expression, Freedom of Assembly and Association and the Right to Work
9. Remedies - for harm done to people by AI decision-making
10. Promotion of AI Literacy - greater public understanding of AI realities is needed

8.3 Informed by the observations contained in this paper, and experience in their own countries, my suggestion here is that PC-CP members work out **what a sector specific expression** of these ten requirements would look like in relation to "prisons, probation, youth justice and relevant private companies sector". Some members of the committee may feel that it is difficult to devise arrangements to monitor the ethical use of AI in respect of prisons and probation in isolation from arrangements that, in any given country, are already in place, or in train for the use of AI in police, prosecution and court services. What is required in the PC-CP's proposed Recommendation is clarity for prison, probation and youth justice services - and the private companies that aspire to serve them - about what is permitted and what is prohibited, premised on an understanding that institutional arrangements for taking new, AI-related tasks forward, will necessarily look different in different places.

9. Conclusion.

9.1 I hope that the information contained in this report will enable the PC-CP to write a Recommendation that is consistent with established European documentation on the ethical and strategic implications of AI in respect of the prisons, probation, youth justice and affiliated commercial organisations. I hope it serves as a contribution to the "AI literacy" that the Council of Europe has already called for, in respect of this field. I hope too that the shortcomings of the European documentation will also be recognised, and account taken of other ethical standpoints. The humanistic traditions and evidentially-founded successes of probation, in particular, sit somewhat uncomfortably with the deployment of AI, depending of course, on how it is used. It is not - as the European documentation fully recognises - that AI could not be used to good effect in the administration of democracies and their public services: the question is whether it will be, and what it will take to ensure that. The aspiration to constrain and shape the use of AI in Europe by

human rights instruments and the rule of law is indeed an admirable and necessary one, but it is far less easy to accomplish than to desire.

9.2 Most readers of this report will recall, from living memory, that in the first decade of the 21st century there was considerable hype and expectation about the use of social media to deepen extend and enrich democracy, by enabling much greater public participation - from hitherto unheard voices - and much fuller exchanges of information. We know better know. We have seen how social media became as much of a mechanism for spreading fake news and disinformation, and subverting democracy. We are learning in retrospect to regulate against this, but the is damage done, and the world has changed irrevocably. Whatever ethics were supposed to constrain the use and spread of social media, their impact has been limited. AI has been - and remains - directly implicated in the misuses of social media, and has made its reach and penetration into everyday life possible, and - for some - profitable.

9.3 AI *could* be used well in public services, but the question is - will it? The greatest danger of AI is that, because of the investment needed to build, operate and upgrade it, it will make powerful global interests more powerful - the same commercial and political interests that show no inclination to alleviate social injustice or improve social wellbeing - so powerful, in fact, that can disregard merely ethical attempts to constrain them. So much of the work undertaken by probation and prison services is rooted in lives affected by social injustice. I hope that the PC-CP Recommendation, and the stance it takes on AI, will reflect this.

References (and books consulted)

Arrigo B A and Sellers B G (eds) *The Pre-Crime Society: crime, culture and control in the ultramodern age.* Bristol: Bristol University Press

Benasayag M (2021) *The Tyranny of Algorithms: freedom, democracy and the challenge of AI.* London: Europa Editions. (Originally published in Paris: Edition Textuel)

Canals J and Heukamp F (eds) (2020) *The Future of Management in an AI World.* London: Palgrave MacMillan

Capek K (1920) R.U.R - Rossum's Universal Robots.

Commissioner for Human Rights (2019) *Unboxing Artificial Intelligence - 10 Steps to Protect Human Rights.* Strasbourg: Council of Europe.

Crawford K (2021) *Atlas of AI: power politics and the planetary costs of artificial intelligence.* New Haven, USA: Yale University Press

Demetriou J(2018) From a reactive to a preventive approach: what is on the horizon of electronic monitoring technologies? *Justice Trends* 3

Devaprasad A (2021) Technically Worse: The paradox of smart prisons in India. 6th September 2021. https://thebscblog.wordpress.com

Devlin K (2018*) Turned On: Science, Sex and Robots.* London: Bloomsbury

Dubber M D, Pasquale F and Das S (eds) (2020) *The Oxford Handbook of Ethics in AI.* Oxford: Oxford University Press

Dyer-Witherford N, Kjosen A M and Steinhoff J (2019) *Inhuman Power: Artificial intelligence and the Future of Capitalism*. London: Pluto Press

European Commission (2021) *Proposal for a Regulation of Artificial Intelligence.* Brussels: European commission

Fallows J (2021) Can Humans be Replaced by Machines. *New York Times* 19th March 2021

Ford M (2015) *The Rise of the Robots: technology and the threat of mass unemployment.* London: Oneworld

Green B and Rigano C (2020) Specialised Smartphones Could keep Released Offenders On Track for Successful Re-entry. file:///Users/mike/Desktop/%20NIJ%20-%20Smartphones%20Could%20Keep%20Released%20Offenders%20on%20Track%20for%20Successful%20Reentry%20%7C%20National%20Ins.webarchive

Harasimiuk D E and Braun T (2021) *Regulating Artificial Intelligence: Binary Ethics and the Law.* London Routledge

High Level Expert Group (2019). On Artificial Intelligence. Brussels: European Commission

Kleeman J (2020) Sex Robots and Vegan Meat: Adventures on the frontier of birth, food, sex and death. London: Picador

Kleijssen J (2021) Justice and Beyond: Council for Europe working on Setting Global Benchmarks on Artificial Intelligence. Interview, in *Justice Trends* 7: 27-31

Lee K-F (2020) AI will go from Rocket Science to Mainstream. in *The Wired World in 2020.* London: Conde Nast Publications.

Madary M and Metzinger (2016) T Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology.
https://www.frontiersin.org/articles/10.3389/frobt.2016.00003/full

Mijatovic D (2019) *Speech.* Conference of Council of Europe Justice Ministers "Justice in Europe facing the challenges of digital technology" Strasbourg, 15 October 2019

National institute of Justice (2019a) *Artificial Intelligence Research and Development to Support Community Supervision: A Solicitation*. Washington, DC: US Department of Justice.

National Institute of Justice (2019b) *AI Enabled Community Supervision for Criminal Justice Services: Award Information.* Washington, DC: US Department of Justice.

Penal Reform International (2021) *Global Prison Trends 2020*. London: Penal Reform International

Pitts S and Tigges L (2021) Probation on 2030: pitfalls and possibilities. In *Advancing Corrections 11*. (Special issue on envisioning corrections in 2030: where should the evidence take us?)

Revell T (2021) The Hidden Costs of AI: interview with Kate Crawford. *New Scientist* 27th March 2021. 47-49

Roose K (2021) *FutureProof: 9 Rules for Humans in the Age of Automation*
New York: Random House.

Rose J (2016) The Dark Side of VR: virtual reality allows the most detailed, intimate digital surveillance yet. *The intercept.* December 23rd 2016.

Schuilenberg M and Peeters R (eds) *The Algorithmic Society: technology, power, knowledge*.
London: Routledge

Silver D, Singh S, Precup D, and Sutton R S (2021) *Reward is Enough.* Artificial Intelligence Vol 299, October 2021

Susskind R and Susskind D (2015) *The Future of the Professions: how technology will transform the work of human experts*. Oxford: Oxford University Press

Susskind D (2020) *A World Without Work: technology, automation and how we should respond.*
London: Allen Lane

Susskind R and Susskind D (2015) *The Future of the Professions: how technology will transform the work of human experts*. Oxford: Oxford University Press

Tasioulas J (2019) First Steps towards an Ethics of Robots and Artifical Inteligence. Journal of Practical Ethics. 7(1) 49-83

UNESCO (2017) Report of COMEST on Robotics Ethics. Paris: UNESCO

http://unesdoc.unesco.org/images/0025/002539/253952E.pdf

Zavrsnik A (ed) (2019) *Big Data, Crime and Social Control.* London: Routledge

Some Novels about AI

Ian MacEwan (2019) *Machines Like Me.* London: Jonathan Cape. A young London man buys an expensive male humanoid AI/robot (the women robots had sold out before he got to the shop) and takes him home to meet his girlfriend. The novel is a serious engagement with AI. It questions how human a machine can become and, among other things, suggests that while the robot can come to understand "justice", it will never understand "mercy". All this takes place in the UK in 1987 - an imagined UK in which computer-pioneer Alan Turing never committed suicide in 1952, and went to make Britain a world leader in AI and robotic technology. Turing appears as a minor character in the novel, which is in many ways a very respectful tribute to him, whilst mindful of the complex moral dilemmas his technology bequeathed to us.

Andromeda Romano-Lax (2018) *Plum Rains.* New York: Soho Press. Set in Japan in 2029, focussed on a 100-year old Japanese lady, her Philipino nurse-carer, and the humanoid AI Robot bought by the old lady's son to (maybe) replace the nurse. By going into the strange political back story of the old lady, the novel explores how identities develop, consolidate and change - hers and the robot's - while the nurse, brilliantly drawn, has to work through all her feelings about the prospect of being replaced by a smart machine - but that's not what happens.

Rob Reid (2017) *After On: a novel of Silicon Valley*. New York: Del Rey/Random House . Written by a Silicon Valley insider, *After On* is the story of a dubious attempt to build and train an AI from data gathered on an imagined Twitter-like social network. The Silicon Valley ambience is real enough, the story a bit slow and too long.

Louisa Hall (2015) *Speak*. London: Little Brown/Orbit. A beautiful, highly original novel about the origins, history and raw materials of a software programme that becomes the "personality" of a child-robot, Mary3. Built by a tech entrepreneur who had intended her as a children's companion, tho' all the models ever built have been scrapped by the time the story opens, and he is in jail. Told in multiple, alternating voices, including Mary3's and her creator, as well as Alan Turing, it's a deep and serious mediation on how entwined the interaction of people and machines already is, and how messy it might become.

Chris Beckett (2003) *The Holy Machine.* London: Atlantic Books. Pure, entertaining feminist science fiction, about Lucy, a realistic-looking woman robot, deployed as a prostitute, and the rather naïve young man who decides to rescue her before the authorities "reset" her unduly sentient mind. The chapters written from Lucy's own viewpoint, about her work, are funny, but astute.

Mike Ashley (ed) (2019) *Menace of the Machine: the rise of AI in classic science fiction* . London: the British Library. A collection of short sf stories about "thinking machines" from the first half of the twentieth century, some more prescient than others.

Extra references . There is an Autonomous Robot Evolution: Cradle to Grave project at the University of York which is examining ways in which robots in particular environments (particularly inhospitable ones, like the surface of Mars) might assess needs, determine requirements and reproduce improved versions themselves with software and 3D printers in order to solve new, initially unanticipated, problems.

https://www.theguardian.com/commentisfree/2021/jun/21/robots-reproduce-evolution-nature-technology

The IEEE Computational Intelligence Society has debated "the autonomous evolution of robot ecosystems" 9.10.19

A Google dermatology assist programme - and machine learning-based AI which identifies skin conditions from images has received EU approval ahead of any approval in the US, and may be in limited use by the end of 2021. Googles claim is that it is "on par" with what trained human dermatologists can do. Tom Simonite 20210 Google launces a new medical app - outside the US. Wired 23.6.21.