

# Some Implications of Artificial Intelligence for Probation and Imprisonment in Europe: guidance for a PC-CP Recommendation

Mike Nellis

Emeritus Professor of Criminal and Community Justice,

University of Strathclyde

Co-adviser to 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.**

## Fundamental EU Principles

- respect for human dignity
  - freedom of the individual
  - respect for democracy
  - justice and the rule of law
  - Equality, non-discrimination and solidarity
  - citizen's rights
- All have roots in the pre-machine world, before connectivity. Mediating **relationships** through digital tech subtly alters all the above.

## Key documents for this project

- ***Ethics Guidelines on Trustworthy AI*** (2019) by the High Level Expert Group (HLEG) on Artificial Intelligence for the European Commission.
- ***Unboxing Artificial Intelligence – 10 Steps to Protect Human Rights*** (2019), by the Council of Europe's Commissioner for Human Rights.
- Note: **AI Literacy** is vital if the upcoming debate is to be serious, inclusive and effective. **The PC-CP should actively promote AI literacy.**

# Defining Artificial Intelligence is difficult

- Straightforward **technical** definitions of AI are *relatively* easy – the application of **machine learning**, driven by **algorithms**, to identify patterns, correlations, risks & anomalies in contrived datasets relevant to the achievement of desired goals.
- AI can assess complex information on a scale and at a speed that humans cannot match: it is **computation**, not “intelligence” in the human sense.
- Conceptual models: **Narrow AI**. General AI “Superintelligence”
- **“Algorithmic governance”** = the monitoring & control of systems and processes, and the people enmeshed in them, using data, sometimes in real-time. This will transform **management** in organisations, businesses and states.
- Not all technologies used in probation and prisons constitute or require AI – Electronic Monitoring (EM), for example – but **user interaction** with them generates **data**, and the analysis of that at scale does require AI.
- An organizational commitment to **datification** – to becoming data-driven - is probably the precursor of using AI – but it may work the other way round – cost efficient, cheap AI will drive and normalize datification.

# Three Frameworks for Understanding the “Social Life” of AI

- **“The Fourth Industrial Revolution” (Davos – Klaus Schwab)**
  - Three new technologies – AI, biotech & new materials - are revolutionizing global (especially Western) society, economy and governance. This is Progress.
  - Capitalism/commerce will, as ever, be the driver of innovation, prosperity and security for all (albeit with significant transformations of work). Governments will welcome this.
  - Law and Ethics will adequately constrain the risks and dangers of “disruptive innovations”
  - This is largely the framework underpinning debate on AI in European institutions
- **“Surveillance Capitalism” (Shoshanna Zuboff)**
  - Business is driven by relentless extraction and analysis of data from everyone and everything with a view to **predicting** next moves by consumers (individual and aggregate), competitors, governments. AI = commercial edge. This is “invasive technification” (Gernot Bohme)
  - In this context, dignity and freedom do not mean what they used to: they are diminished.
  - Capitalism need not be like this – but “tech giants” escaped the constraints of ethics & law.
- **Marxist:** AI will consolidate social inequality and create “Inhuman power”

# The Putative Benefits of AI

## (versus dystopian scenarios)

### AI in Medicine, some examples

- Screening for breast cancer
- Predicting mutation in viruses
- Tracking patients vital signs, via wearables & ingestibles
- Designing new drugs, faster.
- (In US) identifying people at risk of opioid addiction, to advise doctors when not to prescribe opioid-based painkillers. (Note: it over-predicted cases).

### The Hard Questions

- AI could well be used for benign social purposes – **but will it be?**
- Are ethics and law enough to ensure **equal** benefits - or to whom?
- AI has underacknowledged downsides – immense carbon footprint, hidden labour costs (in creating databases).
- AI systems will never be perfect – glitches and vulnerabilities are endemic
- **Will AI bestow more political and commercial power on the already powerful. How can it not?**

-

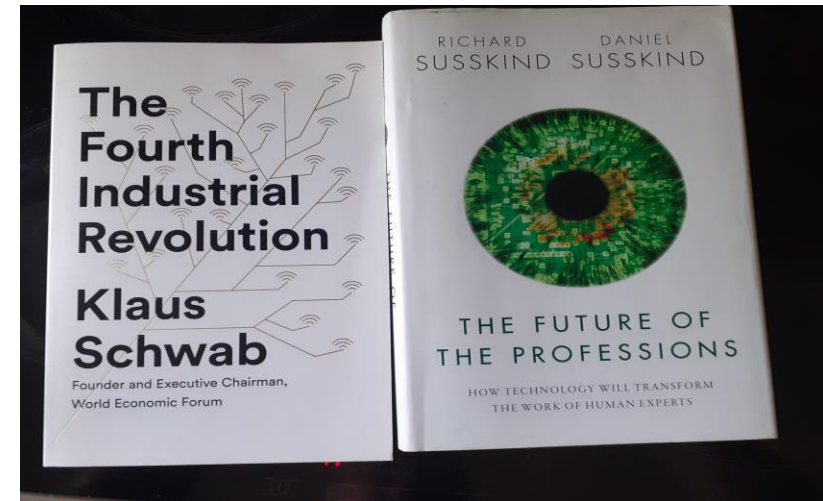
# Trustworthy and Human-centric AI (HLEG)

(in all sectors, & some Serious Practical Questions)

- Trustworthy AI is
  - legally regulated
  - ethically defensible
  - Technically robust and reliable
- Human-centricity requires of AI:
  - Respect for human autonomy
  - Prevention of intentional and unintentional harm
  - Fairness
  - Explicability (some machine learning outcomes are opaque to its creators)
- How far and how fast will AI in prisons and probation (and private sector) go? The PC-CP Recommendation needs a **time limit**, beyond which it can't reliably "see", and will need to review progress. Say 2032?
- Will the PC-CP take a **limiting approach** to AI – use only when absolutely necessary – based on its own Probation Rules and empirical evidence of what is already **humanly possible** in probation OR prepare probation services for transformational changes? OR both?
- National probation services in Europe will surely adapt unevenly to AI ?
-

# Keeping the “Human in The Loop” (HitL)

- Jargon for the optimal use of AI guidance/ recommendations/ predictions = **not replacing** humans, but **working in tandem** with them, with humans as the final decision maker.
- Davos/HLEG supports this.
- “Automated Decision Support Systems” were a step towards this
- Protocols are required for such human/AI partnership.
- HitL still requires technical upgrading of the workforce,



- “Human in the loop” is not a stable arrangement in the long run – cost-efficiency of AI will make human replacement economically inevitable in many cognitive professions. (says Prof. Susskind)
- By 2032? Debatable.
- **A moral/political choice will be needed to preserve human presence/probation in the face of economic logic and AI capabilities.**

# Automating Probation?

- Probation is a form of crime control. Governments want cost-efficient means of crime control. **The personal & relational essence of probation is not sacrosanct unless it is specifically upheld by moral/political values which support it.** If cheaper technical options can accomplish core probation tasks, **24/7**, what then?

**Dialogue/Counselling** - chatbots (voice and text) and/or avatars.

**Risk Assessment** – already automated, using limited data/variables

**Writing Court Reports** - easily automated, like some sports and news reports

**Delivering Programmes** - online, multiple platforms, including (maybe) VR

**Tracking locations (and health signs)** – use smartphones and fitbits for enriched EM

**Enforcing alcohol monitoring** - observe portable breathalyser use via  
smartphone camera (video-check-ins)

**Co-production with service users** --- they may welcome automated service delivery and reduced personal interaction!





# Probation and Virtual Reality?

- Some experiments, premised on improved user engagement, with likely cognitive & affective benefits
- “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 to recreate the privations of prison by requiring a certain number of hours in headset solitude” (Tom Gash 2020)
- VR headset (not voice, keyboard or touchscreen) is generator of vast new troves of corporeal & psychological data.
- Vital? Or “invasive technification”?

# AI, Work and Professional (Un)Employment

- Automation of manual work in the past “threatened jobs” – created the spectre of mass unemployment - but expanding economies created new jobs ..... (including in probation services)
- Automation of cognitive/professional work – by AI – may have more adverse qualities – no replacement jobs of similar skill, status or remuneration – hence debates on Universal Basic Income & the 4-day working week
- AI will accelerate and intensify trends in **workplace surveillance**
- AI **will transform** and **could eliminate** probation work *as we know it* (how far by 2032?).
- **Probation workers (and unions) MUST be consulted and engaged with about the coming of AI and the future shape of probation. Those consultations should start now, ahead of the PC-CP Recommendation.**

# AI and Imprisonment

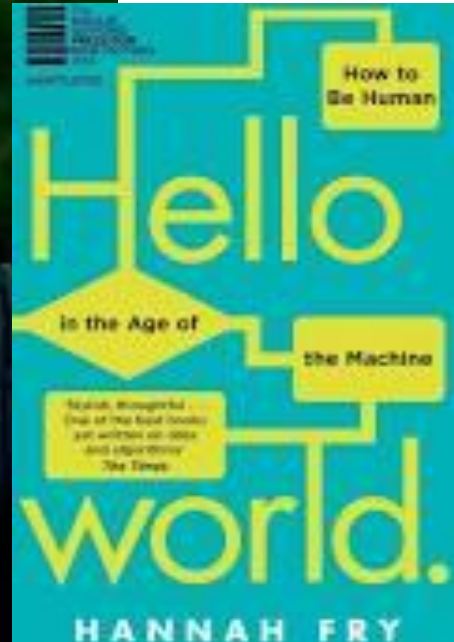
- Empowering access to digital services for serving prisons is clearly useful for rehabilitation and reintegration. Finland leads the way.
- Security technologies in prison grow ever more sophisticated. Neither these nor digital rehabilitation services necessarily **require** management by AI, but they do generate data, and **could be** managed by AI.
- Automation of routine staff tasks frees them up to do more important work! A mantra - but does this necessarily happen? Will it happen in the future?
- Prisons could become a special kind of **smart building** – gathering data on behavior, movement and life-signs from embedded sensors – **eg to prevent suicide**. *Questionable!*
- The spectre of all-automated Changi prison complex in Singapore, born in a culture of relentless tech innovation and no human rights considerations. **Legislate against this in Europe.**
- Resistance to solitary confinement is premised in the case FOR **Meaningful Human Contact**: algorithmic management (digital isolation?) of prisoners should take note of this.

# AI and Private Companies in the prisons-probation space

- Private companies are the locus of innovation in AI – even when governments are funding it - and have a vested interest in its success and expansion. They seek new markets, and sometimes deliver services.
- Governments see the value (for prosperity and security etc) and actively encourage and sponsor AI – and absorbs private sector into itself.
- Private companies (or consortia of them) devise their own ethics committees and policies – of which we should be careful
- Private companies set AI agenda in particular fields – seductive commercial visions of how AI can improve/disrupt/transform penal practices will be a permanent presence in debates about public service futures from now on.

# Dr. Hannah Fry *learns a lesson* .....

Professor of the Mathematics of Cities, University College London



- Once a naïve, true believer in benign algorithmic management – more cautious now.
- **The Berlin incident (2014?)**– very hostile audience reaction to presentation of her (London-based) research on the utility, for policing, of **urban crime prediction**.
- **What if**, said this audience, this tool fell into the hands of a future authoritarian government? What would **they** predict and suppress?

# The End

Thank You

- Do not create “**systems so perfect that nobody has to be good**”. T S Eliot, English poet.

• [mike.nellis@strath.ac.uk](mailto:mike.nellis@strath.ac.uk)