

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Managing risks, passing over harms? A commentary on the proposed EU AI Regulation in the context of criminal justice

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Lavorgna A, Ugwudike P (2022). Managing risks, passing over harms? A commentary on the proposed EU AI Regulation in the context of criminal justice. JUSTICE, POWER AND RESISTANCE, 5(3), 292-298 [10.1332/NPZK4880].

Availability:

This version is available at: <https://hdl.handle.net/11585/901321> since: 2022-11-10

Published:

DOI: <http://doi.org/10.1332/NPZK4880>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

Managing risks, passing over harms? A commentary on the proposed EU AI Regulation in the context of criminal justice

Mind the gap

Artificial Intelligence (AI) is broadly defined as technologies at the basis of systems that display intelligent behaviour by analysing their environment and taking actions (with some degree of autonomy) to achieve specific goals (European Commission 2018). Such technologies are now increasingly deployed across Western and non-Western societies with socially transformative effects (Esposito 2013; Gruetzemacher and Whittlestone 2019). Nowadays, AI is broadly used to make certain services to the public or some production processes more efficient (among many, consider for instance Cioffi et al. 2020; Meyer et al. 2020). It is also deployed in more sensitive sectors such as criminal justice settings. For instance, AI is used by judicial and police systems to ‘predict’ – among other things – risk of recidivism, crime risk locations, or to implement biometric identification (for a recent overview, see Brownsword and Harel 2019; Author 2020; 2021; Authors 2021). As such, it is somehow surprising to think that, at the moment of writing, AI is still mostly reliant on self-regulation, and subject to few legal regulations.

The current legislative gap seems destined to end after the European Commission published in April 2021 a proposal for a Regulation on the subject, focused on managing and controlling the risks associated with AI systems (from here on, AIS) (European Commission 2021a) – hereafter ‘the EU AI Regulation’. Indeed, the proposal does not regulate AI as such. Instead, it regulates the entry into the market and the use of technological systems that rely on AI, in an attempt to maintain as much neutrality towards the technology under discussion, and not to risk to over rely on an obsolescent definition of AI. The Commission has explicitly recognised that, because the increasing use of AIS at the global level, a legislative intervention from the EU as part of shaping Europe’s digital future would ensure ethical use of new technologies. It will also allow the region to be consolidated as a centre of world importance in the sector. This is seen as a precondition for the prosperity and competitiveness of the EU (European Commission 2021b,c; Smuha 2021).

In this short contribution, we argue that such a proposal should be welcomed as the first comprehensive effort to regulate AIS via a legal framework while trying to establish potentially

global standards. But the EU's approach – which is mostly grounded on the concept of technological risks – fails to adequately address some of the major concerns surrounding AIS, particularly in sensitive settings such as criminal justice. We suggest that a different or at least a complementary approach focusing on social harms might be useful to overcome some of the limitations of the current regulatory attempt. As such, we also advocate for the relevance of criminological imagination, and specifically for social harm approaches, in the multidisciplinary debates on AIS and datafication, as they can be of great utility in highlighting unintended consequences and vulnerabilities.

A brief overview of the proposed EU AI Regulation and its risk-based approach

If approved, the EU AI Regulation would be applicable to a wide range of AIS, offering common rules and a series of key principles that aspire to be proportionate and flexible. The challenge, unsurprisingly, is to balance the need to mitigate the risks, and in particular those inherent the inappropriate use of AIS, with the need to support innovation.

Risk is a concept that has long received attention by several disciplinary perspectives (Lupton, 1999). From a sociological perspective, risk may be defined as ‘a systematic way of dealing with the hazards and insecurities induced and introduced by modernization itself’ (Beck 1992: 21). Risks can be interpreted in a more objective way – as relative likelihood of a certain event occurring – or in a more subjective or epistemic way – such as beliefs about the likelihood of a certain event occurring (Finckelstein 2003: 973). Either way, discussions about risks are usually about forecasting future harms (including the social harms at the core of this special issue), in the attempt to avoid or mitigate them. But the relationship between risks and harms is more complex than this. In some cases, for instance, risks are not a sufficient or necessary cause of harm (Livingstone 2010). On the other hand, certain risks can be conceptualised as harmful since exposure to certain risks can undermine a legitimate interest (Finckelstein 2003).

As already noted, the Regulation follows a *risk-based approach*, differentiating between uses of AIS depending on the severity of risk to the health and safety or fundamental rights of natural persons they create or might create.

First, AIS that can create an *unacceptable risk* are generally forbidden. Among those, are systems using ‘real time’ remote biometric identification in publicly accessible spaces. However, the draft provides for several exceptions that seem to allow their use (art. 5(2-4) for

instance for law enforcement purposes. These are based on case-by-case assessment parameters that take into account the nature of the situation, the consequences of using the specific AIS, and proportionality criteria. Second, some AIS are considered *high-risk* depending on their intended purpose, in line with existing product safety legislation. Such AIS are permitted on the European market subject to compliance with certain mandatory requirements and an *ex-ante* conformity assessment. Most AIS with a direct criminal justice application are likely to fall into this category. Third, some provisions target AIS because of the specific *risks of manipulation* they pose; as such, even if they are considered as bearing only limited risk, there are a number of transparency obligations for their use. Generally, when an individual interacts with an AIS recognising their emotions or characteristics, they must be informed. Further, if an AIS is used to generate/manipulate image, audio or video content resembling authentic content (as in the case of deepfakes), this needs to be disclosed.

In a recent article (Author and colleague 2021), a number of problematic issues in criminal justice matters pertaining to the draft Regulation were identified and are summarised in the following typology:

- (1) *scope* issues. For instance, there are grey areas regarding what could be considered as AIS, and the very broad use of exceptions, especially in sensitive areas such as criminal justice matters, could undermine the safeguards in place;
- (2) *implementation* issues. The draft seems to underestimate the role of private companies in criminal justice matters (see Byrne et al. 2019; Corda and Lagerson 2020). But depending on how it is implemented, the Regulation could allow a wide range of facial recognition technologies designed by private sector companies to identify and isolate individuals in public contexts such as criminal justice settings;
- (3) *spatial* issues. To be successful, the Regulation will have to be accepted as the rightful standard by a majority of global actors, which in part explains the ‘shyness’ of the approach; but it will also have to ‘fit’ with very different criminal justice systems within the EU, which could prove difficult. Additionally, the Regulation carries a problem of ‘maximum harmonisation’ and residual competences, which might have, potentially, a preemptive effect, creating a situation where Member States abilities to act in the area (e.g., regarding their capability to further restrict the use of AIS in policing matters) would be disabled (Veale and Zuiderveen Borgesius 2021);

(4) *temporal* issues. After all, the proposal... it still a proposal, and we might need a long time before it becomes a Regulation; also, the current draft envisages a 18-month period for its implementation. Hence, the regulatory gaps at the moment remains;

(5) *intervention* issues. The sanctioning system envisaged by the current draft proposes only economic sanctions as a deterrent and punishing tool. Overall, this leaves considerable freedom for self-regulation.

On this latter point, we need to keep in mind that the legal accountability of commercial actors, and especially tech companies, is still very porous, although the issue is receiving increasing attention. There is now political pressure to better regulate the sector but there is still no coherent international legal framework on the subject, and the draft Regulation falls short on this aspect. Intervening to find new balances between the drive to maximise innovation and technology profits, and society's interests is probably one of the main (and trickiest) challenges today, but this does not mean it is one we can evade.

Shifting the focus

The risk-based approach summarised above has been generally welcomed, even by commentators that were otherwise critical on some aspects of the draft Regulation (see, for instance, Veale and Zuiderveen Borgesius 2021). But we believe that it has some inherent deficiencies. For instance, the focus on risk adopted puts the spotlight on developers and deployers as 'risk managers'. As mentioned above, this facilitates self-regulation and might not be a sufficient deterrent. By meeting such limited regulatory requirements, developers and procurers can minimise the legal, commercial and even reputational risks of using AIS. This approach overlooks the third element of the equation – that is, users (in most cases, individuals) – leaving them as an afterthought although users, both as active or passive subjects, are the more vulnerable to digital and technological harms (e.g., Wood 2021). Of course, the proposed Regulation is based on article 114 of the Treaty on the Functioning of the European Union (TFEU), governing shared competence on the internal market with the aim to create a practical framework that is fit for Europe's digital future (Dufor et al. 2021); a focus on users would require a different basis (for instance, art. 83 TFEU). Nonetheless, this omission of users needs to be pointed out: after all, using article 114 is a choice in itself.

In our digitised societal context moral panics originate from a vast number of relatively unfamiliar (real or perceived) threats including crime, and risk-based approaches have become central to managing such threats (Beck 1992; Simon 2007; Ungar 2001; Hier 2003; Author 2019). Harms, on the other hand, tend to remain elusive and unspecified, especially when it comes to digital and technological harms, whose quantification can be difficult to calculate because of the intrinsic dark figure of negative events, and because it can be daunting to isolate the digital or tech element from a complex real-life scenario.

But focusing on risk rather than on harm gives a false sense of control. Risk, after all, is about the probability and severity of a certain harm manifesting itself, with the added complication that it implies a sort of forecast of the future, in a context where technology evolves at a very fast pace. Why then not to put the harms – and those impacted by them – at the centre, or at least why not to give them more attention? After all, the merits of harm-based approaches have been long debated in criminology (see, among many others, Pemberton 2007; Dorling et al. 2008; Paoli and Greenfield 2013), and even in more recent literature focusing specifically on the harms of data-driven AIS. Indeed, a fast-growing scholarship on such harms now exists and can as such expose affected populations to excessively punitive penal intervention (e.g., Authors 2021).

AI Regulation as an issue of justice, power and resistance – let's mind the right gap

This intervention does not seek to demonize the use of AIS. We recognise that they are underpinned by the laudable notion that, if well designed, they can perform certain tasks for which computing power is more suitable and effective than solely human activity. However, we should also keep in mind the idea that 'can implies ought' is invalid (Niiniluoto 1990): the harms of AIS should form a crucial part of criminological and interdisciplinary scholarship, and should be addressed via a robust legal framework. This intervention reaffirms this and highlights the importance of expanding the social harm scholarship in criminology, to accommodate theoretical and empirical analysis of the risk and harms increasingly associated with new and emerging AIS, as well as considerations of how best to develop remedial strategies.

References

Beck, U. (1992) *Risk society: towards a new modernity*. London: Sage.

Brownsword, R., and Harel, A. (2019) 'Law, liberty and technology: Criminal justice in the context of smart machines', *International Journal of Law in Context*, vol 15, no , pp 107-125.

Byrne, J., Kras, K.R. and Marmolejo, L.M., (2019) 'International perspectives on the privatization of corrections', *Criminology and Public Policy*, vol 18, pp 477-503.

Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A. and De Felice, F. (2020) 'Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions', *Sustainability*, vol 12, p. 492.

Corde, A. and Lagerson, S.E., (2020) 'Disordered punishment: workaround technologies of criminal records disclosure and the rise of a new penal entrepreneurialism', *The British Journal of Criminology*, vol 60, no 2, pp 245-264.

Dorling, D., Gordon, D., Hillyard, P., Pantazis C., Pemberton, S. and Tombs, S. (eds.), (2008) *Criminal obsessions: Why harm matters more than crime*, London: Centre for Crime and Justice Studies.

Dufor, R., Koehof J, van der Linden, t. and Smiths, J. (2021) 'AI or more? A risk-based approach to a technology-based society'. *Oxford Business Law Blog*. Available at: <https://www.law.ox.ac.uk/business-law-blog/blog/2021/09/ai-or-more-risk-based-approach-technology-based-society?msclkid=c5779db8aacd11eca4a1f68d556809f4>.

European Commission (2018) Communication from the Commission. Artificial Intelligence for Europe, COM(2018)237.

European Commission (2021a) Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM(2021)206 final.

European Commission (2021b) Communication on Fostering a European approach to Artificial Intelligence. Available at: <https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence>.

European Commission (2021c) A European approach to AI. Available at: <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>.

Esposito, E. (2013) 'Digital prophecies and web intelligence', in M Hildebrandt and K De Vries (eds) *Privacy, Due Process and the Computational Turn*, London: Routledge.

Finckelstein, C. (2003) 'Is Risk a Harm Symposium: Preferences and Rational Choice: New Perspectives and Legal Implications'. *University of Pennsylvania Law Review*, vol 151, no 3, pp 963-1003.

Gruetzemacher, R. e Whittlestone, J. (2019) 'Defining and Unpacking Transformative AI', arXiv:1912.00747v2.

Hier, S. (2003) 'Risk and panic in late modernity: implications of the converging sites of social anxiety'. *British Journal of Sociology*, vol 54, no 1, pp 3-20.

Livingstone, S. (2010) e-Youth: (future) policy implications: reflections on online risk, harm and vulnerability. In: e-Youth: balancing between opportunities and risks, Antwerp, Belgium. Available at:
[http://eprints.lse.ac.uk/27849/1/eYouth_\(future\)_policy_implications_\(LSERO_version\).pdf](http://eprints.lse.ac.uk/27849/1/eYouth_(future)_policy_implications_(LSERO_version).pdf).

Lupton, D. (ed.). (1999). *Risk*. London: Routledge.

Meyer, C., Cohen, D. and Nair, S. (2020) 'From automats to algorithms: the automation of services using artificial intelligence', *Journal of Service Management*, vol 31, no 2, pp 145-161.

Niiniluoto, I. (1990) 'Should technological imperatives be obeyed?'. *International Studies in the Philosophy of Science*, vol 4, no 2, pp 181-189.

Paoli, L. and Greenfield, V.A. (2013) 'Harm: a Neglected Concept in Criminology, a Necessary Benchmark for Crime-Control Policy'. *European Journal of Crime, Criminal Law and Criminal Justice*, vol 21, no 3-4, pp 359-377.

Pemberton, S. (2007) 'Social harm future(s): exploring the potential of the social harm approach', *Crime Law & Social Change*, vol 48, no 1-2, pp 27-41.

Simon, J. (2007). *Governing through crime. How the war on crime transformed American democracy and created a culture of fear*. Oxford: OUP.

Smuha, N.A. (2021) 'From a "race to AI" to a "race to AI regulation": regulatory competition for artificial intelligence'. *Law, Innovation and Technology*, vol 13, no 1, pp 57-84.

Ungar, S. (2001) 'Moral panic versus the risk society: the implications of the changing sites of social anxiety'. *British Journal of Sociology*, vol 52, no 2, pp 271-291.

Veale, M. and Zuiderveen Borgesius, F. (2021) 'Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach'. *Computer Law Review International*, vol 22, no 4, pp 97-112.

Wood, M.A. (2021) 'Rethinking how technologies harm', *The British Journal of Criminology*, vol 61, no 3, pp 627-647.