Human dignity in the EU Artificial Intelligence Act: A case of chasing shadows in the dark?

Teo, Sue Anne

Published in:
Human Dignity and the Law

2025

Document Version:
Peer reviewed version (aka post-print)

7

Human dignity in the EU Artificial Intelligence Act

A case of chasing shadows in the dark?

Sue Anne Teo

1. Introduction

Human dignity, as Nowak observes in his contribution to this book, plays a foundational role within the human rights framework.[1] In this role, human dignity relates to the inherent worth of human beings – namely that this worth is intrinsic and not dependent upon status, race, ethnicity or other markers tied to the human person. At the same time, human dignity is also a self-standing right and a standard towards which specific human rights should aim.[2] It is a concept invoked when humanity is faced with novel challenges. When biotechnological innovations opened up the possibility of cloning and genetic modification, the field of biotechnology had to grapple with drawing red lines to prevent the abuse of these technologies in ways that can undermine the inherent worth humans are said to possess. In a similar vein, the development and use of artificial intelligence (AI) technologies are also forcing human dignity considerations to the forefront.[3] The existing harms caused by AI on human rights – such as impacts on the right to privacy, equality and non-discrimination, freedom of expression and other human rights – have galvanised policymakers around calling for regulation on AI. The world's first attempt to do so in a comprehensive manner was in the European Union's Artificial Intelligence Act, which entered into force in August 2024. The Act was underpinned by the need to promote innovation while

1 Manfred Nowak, Chapter 2 this volume.

2 Christopher McCrudden, 'Human Dignity and Judicial Interpretation of Human Rights' (2008) 19 European Journal of International Law 655.

3 Mark Latonero, 'Governing Artificial Intelligence' (*Data & Society*, 10 October 2018) https://datasociety.net/library/governing-artificial-intelligence/; Sue Anne Teo, 'Human Dignity and AI: Mapping the Contours and Utility of Human Dignity in Addressing Challenges Presented by AI' (2023) 15 Law, Innovation and Technology 241. See also references to human dignity in the UNESCO Recommendations on the Ethics of AI, which was adopted in 2021.

respecting and protecting fundamental rights and EU values, such as human dignity, and can be contrasted with the free market approach towards AI by the United States, but also the state-controlled approach favoured by China.[4]

This chapter takes the AI Act as its point of departure in examining human dignity in the context of AI for two reasons. First, the Act is the first in the world to comprehensively regulate AI. This may confer a first-mover advantage, potentially influencing the approaches, form and priorities of other legislative and governance efforts currently under way in other jurisdictions as well as internationally. Second and relatedly, as in the case of the General Data Protection Regulation (GDPR), there is the possibility of the 'Brussels effect' of EU legislation taking hold in relation to the EU AI regulation as well.[5] The Brussels effect pertains to how other legislative efforts to regulate a given subject area are shaped by the approaches taken by the EU in the legislated area, with the latter having an indirect extra-territorial effect.[6] Data protection legislation in various jurisdictions such as Brazil and India has been influenced by the EU's GDPR, which has been called the 'gold standard' of data protection.[7] Moreover, the EU AI Act may also act as a catalyst for companies developing AI to adhere to these sets of rules as the highest common denominator rules, thus doing away with the burden of adhering to various different regulatory requirements.

This chapter examines to what extent the EU AI Act takes human dignity into account, what role it plays in the legislation, in terms of framing and substance, and the gaps related to human dignity that are still present. It finds that it is not clear how human dignity has been interpreted in the Act, rendering protection of human dignity inconsistent and uneven. It is also unclear in terms of how the regulation can adequately address existing challenges or new threats that AI poses to human dignity.

The chapter is structured as follows. Section 2 examines the human dignity concerns raised by AI and unpacks the role and meaning of human dignity in the AI Act. Section 3 looks into the risk-based approach of the AI Act, which focuses upon high-risk as well as prohibited use cases of AI and analyses where human dignity plays a role, as well as gaps that are present. Section 4 sketches a research agenda for dignity

---

4 Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford University Press 2023).

5 Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press 2020) https://oxford.universitypressscholarship.com/10.1093/oso/9780190088583.001.0001/oso-9780190088583 accessed 22 February 2022; see however Marco Almada and Anca Radu, 'The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy' (2024) 25 German Law Journal 646.

6 Benjamin Greze, 'The Extra-Territorial Enforcement of the GDPR: A Genuine Issue and the Quest for Alternatives' (2019) 9 International Data Privacy Law 109.

7 Asaf Lubin, 'The Rights to Privacy and Data Protection under International Humanitarian Law and Human Rights Law', *Research Handbook on Human Rights and Humanitarian Law* (Edward Elgar Publishing 2022) https://www.elgaronline.com/display/book/9781789900972/book-part-9781789900972-35.xml accessed 6 June 2023.

and AI going forward, and the final section examines remaining gaps and challenges, including by looking beyond the EU.

2. Human dignity and artificial intelligence

2.1 Mapping the relationship

There is no universally agreed definition of artificial intelligence (AI). Attempts to define AI have been mainly referenced in relation to human intelligence. However, the definition of human intelligence itself lacks clarity, with philosophical contestations over its social aspects and its relationship to consciousness. Some have also criticised that the terms themselves are unhelpful – as both its artificial nature, alongside the ostensible lack of intelligence of such technologies, have been called into question.[8] When it comes to defining AI, the EU AI Act takes a functional approach towards its definition, anchoring it to what it can do in the real world. In the AI Act, artificial intelligence is defined as a set of technologies that is able to discern patterns and infer from datasets in order to make predictions, take actions or generate content, doing so with relative degrees of autonomy.[9] To put it another way, AI is distinct in several ways from prior technologies due to its ability to infer patterns and other insights from datasets and its ability to act autonomously (that is, without human intervention), including by taking actions in both the physical and the online world.

The invention of new technologies has propelled questions about human dignity to the forefront. Biotechnological developments, including potential for cloning and genetic medication, were said to threaten human dignity, understood in the sense of undermining human exceptionalism – that is, the supposed special place that humans have in the world. In a similar vein, AI is also posing challenges to human dignity.

In mapping how AI challenges human dignity, I have elsewhere broken down the concept of human dignity into four different conceptualisations that express a recognition of inherent worth.[10] First, human dignity pertains to not treating humans as a means to an end, an idea linked to the categorical imperative expounded by Enlightenment philosopher Immanuel Kant. Second, human dignity relates to the protection of certain vulnerable classes of persons – these may overlap with groups

---

8 Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021); Mohammad Amir Anwar and Mark Graham, *The Digital Continent: Placing Africa in Planetary Networks of Work* (First Edition, Oxford University Press 2022).

9 Article 3 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) 2024. Hereafter referred to as the AI Act.

10 Teo (n 3). To be sure, these conceptualisations of human dignity need not be limited to AI and human dignity.

protected under equality and non-discrimination laws, but need not be a condition precedent.[11] Third, human dignity also relates to the expression and recognition of inherent self-worth through protections afforded under human rights law. Some examples include the right to privacy, which recognises the autonomy of persons, but also the provision of economic, social and cultural rights, which recognises that humans need basic sustenance and services as material conditions to guarantee not only mere survival, but conditions allowing for the other human rights to take hold and be meaningful. Finally, the fourth conception relates to human dignity as human exceptionalism. This tracks to the responses to biotechnology innovations as touched on above, whereby the governance and regulation of bioethics both explicitly and implicitly draws upon the concept of human dignity to prevent certain actions[12] in order to protect humanity as such. These include prevention of human cloning, commercialisation of genetic materials and germline genetic modifications.

Some of these conceptions can be directly applied to AI. AI systems that are biased against racial and ethnic minorities fail to protect classes of vulnerable persons. Where AI systems are first experimented with in environments with low rights awareness, such as in the areas of social welfare and migration and asylum, such deployments in effect treat (certain) humans as a means to an end. Still, gaps remain. In other words, AI may affect human dignity in ways that these conceptualisations don't address. For instance, data and contextual disembodiment through the use of AI is inadequately accounted for, choice architectures that determine how we interact and consume information online has not been examined from a dignity perspective, and finally that human dignity, as a top–down concept, is inadequate in accounting for how technology, specifically AI, impacts us day-to-day, introducing new vulnerabilities through this interaction.[13] Instead, a bottom–up approach, taking into account the lived experiences of persons, including those of vulnerable and minority groups, should shape what dignity means and guide the design of AI.

2.2 Human dignity and the EU Artificial Intelligence Act

Human dignity features both expressly and implicitly in the EU AI Act. Several recitals expressly mention human dignity. Recital 27 links human dignity to human agency and oversight, arguing that AI systems should be 'developed and used as a tool that serves people, respects human dignity and personal autonomy' and does so 'in a way that can be appropriately controlled and overseen by humans'. This is a prescriptive call to ensure adequate human control while using AI and that AI should

---

11 This can be seen, for example, in the *Pretty v The United Kingdom*, no 2346/02, 29 April 2002 case, which pertained to whether Mrs Pretty, a terminally ill woman, could control the circumstances of her death, in other words, to 'die with dignity'.

12 Timothy Caulfield, 'Human Cloning Laws, Human Dignity and the Poverty of the Policy Making Dialogue' (2003) 4 BMC Medical Ethics 3; Steven Pinker, 'The Stupidity of Dignity' [2008] New Republic (New York, NY) https://dash.harvard.edu/handle/1/38822030 accessed 3 April 2021.

13 Teo (n 3).

serve human needs and not vice versa. Recital 28 recognises that AI can be 'misused and provide novel and powerful tools for manipulative, exploitative and social control practices' and goes on to state that these, in turn, 'contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and fundamental rights enshrined in the Charter, including the right to non-discrimination, to data protection and to privacy and the rights of the child'. Human dignity here is listed as one out of many considerations, linked to fundamental values of the European Union. Recital 48 in turn rationalises the classification of the risk-based approach adopted by the regulation, specifically, that AI systems considered as high risk are said to pose high risks to fundamental rights, with human dignity listed as one of those rights. Finally, Recital 58 mentions that AI systems used to determine social security benefits can negatively impact fundamental rights, including human dignity, right to social protection, non-discrimination, or right to an effective remedy.

With the exception of the account of human dignity under Recitals 27 and 28, which invoke it as an underlying value of the EU, human dignity in the other recitals relate to human dignity as a right under the EU Charter of Fundamental Rights (CFR). However, whether invoked as a discrete right or as a foundational value, human dignity as a concept is functionally approached by the courts and through treaty interpretation as consisting of various conceptions – as mentioned, these include the non-instrumentalisation of humans, the protection of certain classes of vulnerable groups, the recognition and expression of inherent worth, and finally as a notion of human exceptionalism. Thus, rather than asking whether dignity plays a foundational or discrete role in the AI Act, it might be more fruitful to examine which of the conceptions discussed above are captured and reflected in the AI Act.

2.2.1 Finding human dignity in the risk-based approach of the AI Act

The AI Act takes a risk-based approach towards regulating AI. Plainly, this means that AI is regulated according to its use cases, where use cases that present an unacceptably high risk to health, safety and fundamental rights are prohibited, while high-risk use cases are subjected to many obligations before AI systems can be deployed. The Act also regulates limited-risk AI use cases, while those with minimal risk are not subjected to legal obligations. Approaching this as a pyramid, minimal-risk AI would form the bottom-most and hence thickest part of the pyramid, followed upwards by limited-risk AI and high-risk AI use cases, while prohibited use cases of AI form the tip of the pyramid. In other words, most AI use cases today are of the limited- or minimal-risk variety and hence are subject to few legal obligations, while prohibited and high-risk AI systems form a smaller, albeit highly regulated, set of use cases.[14]

[14] General Purpose AI models (GPAIs), encompassing generative AI and large language models such as ChatGPT, are also regulated in the AI Act. Article 53 states that GPAIs are required to keep up-to-date technical documentation on testing and evaluations, provide information and documentation to AI providers who wish to integrate it into their systems, put in place a copyright policy and make publicly available content used

2.2.1.1 Prohibited AI use cases

Human dignity is mentioned in relation to prohibited use cases under Recital 28. There are a total of eight prohibited use cases, covering a wide range of possible uses of AI, namely: real-time remote facial recognition systems; social scoring; AI that is subliminally or purposely manipulative; AI systems that exploit age, disability, social and economic status or related vulnerabilities; certain predictive policing systems; indiscriminate scraping of images for facial recognition databases; emotion recognition systems in workplaces and educational institutions; and finally, biometric categorisation systems based on or meant to deduce sensitive characteristics.[15] These varied use cases are said to violate EU values, including human dignity, alongside certain other fundamental rights protected in the Charter.[16]

However, a closer examination of the disparate examples of banned use cases does not seem to reveal an underlying thread tying them together. Pulling two discrete examples out to illustrate this point, we see that while it is undoubtedly detrimental for society to have AI systems that purposely set out to manipulate, and systems that indiscriminately scrape every image from the internet for purposes of building a facial recognition identification database, these two disparate examples engage different conceptions of human dignity (and fundamental rights). In the former, manipulation undermines autonomy, whereby AI systems bypass the engagement with the autonomous capacity of the individual to think, reason and self-govern. Manipulative AI is hidden in the way that it can be difficult to know who is doing the manipulation, for what ends it is pursued and who stands to benefit from it – as a long string of third-party actors typically also benefit from data-driven online interactions. The ability of data-driven AI to specifically target individuals according to their vulnerabilities and proclivities rightly makes manipulative AI a serious threat to autonomy and human dignity.

In the latter example of the prohibited use case of indiscriminate scraping of images with the purposive intent of building powerful image databases, this arguably opens the door to potential widespread abuse where anyone can identify anyone else no matter who or where they are. This 'Clearview'[17] style of scraping and facial recognition

for the training of the AI model. Article 55 regulates the obligations of GPAI models with systemic risk, requiring such providers to perform model evaluation, testing and mitigating systemic risks, document and report to the AI Office incidents of systemic risk and ensure cybersecurity. Systemic risks are defined under Article 3(65) as 'risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.'

15 Article 5 AI Act.

16 Recital 28 AI Act.

17 This is in reference to the service offered by Clearview.ai, a company that scraped millions of images and, powered by its facial recognition technology, is able to identify individuals at a high level of accuracy. See Kashmir Hill, 'The Secretive Company

can open the door for use by law enforcement, but also potentially by nefarious actors out to do harm. Either way, privacy may become nullified at this point, though in different ways. Even as criticism abounds that privacy exercised through consent in online platforms and websites is increasingly meaningless, this bypasses even these lax standards as consent for the scraping of these images is not even sought from individuals in the first place. Further, there are no other legitimate legal bases under EU data protection law that could stand up to scrutiny allowing for this form of data processing. The potential for abuse this practice might raise arguably engages the human dignity concern in relation to using people as a means to an end. With a powerful database at hand, anyone can be potentially targeted for scams and other wrongdoing. On the other hand, despite such a database, the potential for misidentification is non-negligible. Research has shown that minorities are more frequently misidentified by facial recognition systems. When deployed in consequential areas such as policing, the adverse effects, including at the community and societal level, can be compounded. Another possible human dignity argument is that scraping for this purpose leads to individual loss of control, an autonomy concern that undermines human dignity.

However, a human dignity definition within these eight prohibited use cases is elusive. First, there is no direct reference, either within the provisions of the Act nor in its recitals, to any kind of definition of human dignity. Instead, these prohibited use cases are said to pose an unacceptable risk – drawing from a wide gamut of justifications, one of which includes the invocation of human dignity. These use cases, which started with four prohibitions in the AI Act proposal of the EU Commission[18] in 2021, were gradually increased as the Act went through the legislative process at the Council, EU Parliament and finally towards the inter-institutional negotiations and compromise reached between these three institutions – the trilogues – in the final stage of the EU legislative process. Due to the lack of specific reason attached to the prohibitions, scholars have argued that the more likely scenario that legislators had already decided these use cases should be banned, blanketing them under the wide rubric of going against Union values without linking which values to what and which fundamental rights would be negatively impacted.[19] Ebers concurs, arguing: 'The AI Act does not establish criteria for when AI poses an unacceptable risk to society and individuals. Instead, it merely provides a set list of categories of AI systems that are

That Might End Privacy as We Know It' *The New York Times* (18 January 2020) https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html accessed 11 December 2024.

[18] 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 2021.

[19] Ljupcho Grozdanovski and Jerome De Cooman, 'Forget the Facts, Aim for the Rights! On the Obsolescence of Empirical Knowledge in Defining the Risk/Rights-Based Approach to AI Regulation in the European Union' (2022) 49 Rutgers Computer and Technology Law Journal 207.

considered to pose "unacceptable risks" and are therefore banned in the EU.'[20] The same concern persists when pulling from other examples from the prohibited use cases – while social scoring, namely a score attached to every individual within a polity based upon one's actions, is prohibited, social welfare AI systems have been used to monitor individuals and profile them for purposes of fraud monitoring. The latter, which falls under the high-risk AI use case, has disproportionately impacted vulnerable and marginalised populations, who have been wrongly suspected of fraud and been made to account for funds they were allegedly not entitled to.[21] In other words, while a literal score might not be attached to persons in the latter example, it appears to be in practice a 'social score' by any other name – as individuals are measured, grouped and sorted according to status and trustworthiness, determining their ability to access social services.[22]

Thus, in relation to prohibited use cases, human dignity is expressly named, but it is not clear what role it plays. Further, it is hard to find a thread across all banned use cases as it is not clear if the prohibitions are due to challenges to health, safety and an unacceptable risk to fundamental rights, a combination of these factors, or its cumulative effects. The prohibited use cases category lacks a coherent strand justifying exactly why these eight use cases are prohibited. This is the case for both intra-category and inter-category (prohibited and high risk), as seen in the social scoring and social welfare AI system example above. This lack of clarity also means that while some aspects of human dignity may be protected through these prohibitions, other conceptions of dignity may be neglected or not protected at all.

2.2.1.2 High-risk AI systems in the AI Act

High-risk AI systems form the bulk of the legal obligations and is thus the main regulatory objective of the AI Act. These are use cases said to pose a high risk to health, safety and fundamental rights but which are still allowed within the EU provided that legal obligations, attaching to both the provider and deployer of the AI system, are adhered to. The high-risk AI use cases are found under Annex III of the Act and cover AI systems used in the following areas: biometrics systems, critical infrastructure, education and employment settings, access to essential private and public services, law enforcement contexts, migration, asylum and border control management, and finally, AI used within administration of justice and democratic processes. In adopting a risk-based approach, it may be assumed that dignitarian considerations shaped the determination of what counts as acceptable or unacceptable risks. However, much like the analysis of the prohibited AI systems analysis above, it is similarly not clear where and what role human dignity plays in relation to determining, guiding and

---

[20] Martin Ebers, 'Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AI Act' [2024] European Journal of Risk Regulation 1, 9.

[21] 'Suspicion Machines' (*Lighthouse Reports*) https://www.lighthousereports.com/investigation/suspicion-machines/ accessed 29 June 2023.

[22] Coded Injustice: Surveillance and Discrimination in Denmark's Automated Welfare State, Amnesty International, 2024, accessible at: https://www.amnesty.org/en/documents/eur18/8709/2024/en/

shaping high-risk AI systems. This is examined from three angles: the lack of any explanation of the selection criteria in the categories of high-risk AI, tracing the 'essence of the right' approach, and, finally, discrete exclusions within the high-risk AI use cases.

First, the high-risk use category within the EU AI Act, much like its prohibited categories, lacks a coherent thread. The risk-based approach has been commended for introducing a regulatory approach towards AI that aims to correctly strike the balance between innovation and risk to health, safety and fundamental rights. This means that while certain AI use cases are prohibited, high-risk uses of AI are not banned but are subjected to more obligations in order for the AI system to be deployed into the market. The risk-based approach can be contrasted with a precautionary approach, such as through licensing or outright bans of high-risk use cases, and a market-driven approach enabling goods and services to enter the market and establishing liability for harms post hoc, namely after harm occasions.

However, Kaminski argues that the risk-based approach of the AI Act and AI regulation in general carries with it 'policy baggage', in the sense that it inherits characteristics from risk regulation applied in other fields. Further, a risk-based approach towards AI proceeds from the assumption that AI forms an inescapable part of our new technologically mediated reality.[23] In other words, AI is here to stay but its risks should be managed. Malgieri and Santos in turn argue that the risk-based approach may be incompatible with the dignitarian and rights-based considerations that underpin some of the harms, such as discrimination, privacy, and freedom of expression, highlighted in relation to AI.[24]

The risk-based approach requires providers and deployers of an AI system under the high-risk category to carry out certain obligations before the systems are deployed. Providers are defined under Article 3 as those who develop, put into the market, or put the AI system into service under its own name or trademark. Deployers are in turn those who use an AI system under its authority. The AI Act does not cover AI systems used in personal non-professional capacities. The obligations placed on providers are found under Section 2 of the Act, covering obligations to put in place a continuous and iterative risk management system throughout the lifecycle of the high-risk AI (Article 9), taking robust data governance measures (Article 10), having technical documentation (Article 11) and record keeping (Article 11) in place, transparency and information to deployers (Article 12), ensuring the provision for adequate human oversight (Article 14), and, finally, take measures to ensure adequate robustness, accuracy and cybersecurity protections (Article 15). The provider can demonstrate compliance by conducting a conformity assessment with standards developed, by delegated

[23] Margot Kaminski, 'Regulating the Risks of AI' [2023] Boston University Law Review https://scholar.law.colorado.edu/faculty-articles/1621.

[24] Gianclaudio Malgieri and Cristiana Santos, 'Assessing the (Severity of) Impacts on Fundamental Rights' (Social Science Research Network, 25 June 2024) https://papers.ssrn.com/abstract=4875937 accessed 1 January 2025.

private standardisation bodies, for each of these obligations. Deployers of high-risk AI systems for the public sector as well as private actors using AI for public services are in turn required to conduct a fundamental rights impact assessment before using the AI system. This assessment should identify the risks the high-risk AI poses to fundamental rights and address them, including through mitigation measures.[25] There is, in other words, an underlying rationale between the risk-based approach that aims to prevent harms and human dignity. As Grozdanovski and De Cooman argue, 'prevention of harm entails "the protection of human dignity [Art. 1 ECFR] as well as mental integrity [Art. 3 ECFR]"'.[26] Further, the measures required under the Act arguably aims to protect the human dignity of persons – through better representation of individuals in datasets, to the preservation of human agency through human oversight to an impact assessment to gauge and address risks to fundamental rights, including human dignity.

However, there are also gaps within the Act in accounting for human dignity. The first critique raised pertains to the lack of a coherent thread in the high-risk AI categorisation. While eight high-risk use cases, found under Annex III, have been singled out, there is no clear thread that ties these disparate situations together. The classification rightly includes areas where individual rights and freedoms can be impacted to a high degree, such as through the access of public services, but it is unclear why exactly these eight were chosen and not others. At the same time, the high- versus limited- and minimal-risk classification denotes a harm typology of AI that cannot be so readily fixed.[27] Relatively 'harmless' AI use cases such as chatbots, which are considered as having limited risk, may turn out to be manipulative.[28] AI as a sociotechnical system depends very much upon how such systems are used and adopted in societies – as it can also shape social behaviour or introduce new risks.

Second, human dignity is mentioned as one of many fundamental rights considerations that shape the high-risk AI classification. Direct references to human dignity are found in Recitals 48 and 58 in relation to high-risk AI, whereby risks to fundamental rights, including towards human dignity, were invoked to justify the categorisation. Thus, while human dignity features, it is not clear it is the dominant concern, nor does it provide the justification for the categorisation. Recitals are also non-binding. They function to provide interpretive clarity where the substantive provisions in the Act itself are unclear. In other words, while the risk-based approach underlies the AI

---

25 Article 27 AI Act.

26 Grozdanovski and De Cooman (n 19) 299 (the authors cited the work of the High-Level Expert Group on AI in this quote).

27 Michael Veale and Frederik J Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22 Computer Law Review International 97; Ebers (n 20).

28 Kevin Roose, 'Can A.I. Be Blamed for a Teen's Suicide?' *The New York Times* (23 October 2024) https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html accessed 18 January 2025.

Act, rights-based considerations are also predominant in setting apart these high-risk use cases. However, risk and rights cannot be so easily reconciled, as risk-based approaches are primarily assessed through the lens of a cost–benefit analysis while rights approaches are dignitarian in nature. Malgieri and Santos, in their work reconciling the incompatibilities between the risk versus rights approach, proposed reconciling this disjuncture by that we looking for what the core or essence of a right is.[29] They argue that when trying to distil this core or essence, it boils down to what the right in question aims to uphold. They provide the example of the right to privacy as comprising the need to protect 'values such as personal autonomy, human dignity, physical and mental integrity, and identity' alongside the general observation that the 'essence of freedom rights includes personal autonomy, pluralism, and democracy'.[30] In gauging whether or not an interference to a right, which may be variable when it comes to AI systems, amounts to a serious interference to fundamental rights, Malgieri and Santos argue that this is when the AI practices in question 'compromise the dignity and the flourishing of individuals'.[31] Thus, this argument offers the position that many, if not all, of the fundamental rights can be traced towards a dignitarian core. In other words, all roads lead to human dignity in the sense of human flourishing.

However, two explicatory positions can be offered here. First, Malgieri and Santos's position pertains not to the justification in clarifying dignity as the common denominator that binds together the disparate use cases under the high-risk AI category. Instead, it was meant to clarify the operationalisation of the fundamental rights impact assessment to be carried out by deployers of high-risk AI systems. As such, not every interference to fundamental rights is problematic – measures can be taken to mitigate those interferences. It is only serious interferences that are considered to impact upon human dignity.

The second argument pertains to the more general argument about risk-based regulation. Kaminski notes that this regulatory tool often takes on a 'techno-correctionist' nature and 'takes as its starting point that a technology must be fixed so that it can be used'.[32] Thus, even if the fundamental rights impact assessment finds that the interference to rights impact upon human dignity, there is no legal requirement for the AI deployment to be stopped.[33] Seen in this way, dignity is subsumed under risk-based considerations.

The final argument on why human dignity is missing from the AI Act pertains to the discrete exclusions within the high-risk use cases. AI systems used within the field of

---

29 Malgieri and Santos (n 24).

30 Ibid 11.

31 Ibid.

32 Kaminski (n 23) 1397.

33 As long as appropriate mitigation measures are taken, this remains a subjective judgement for deployers as there is as yet no standardised methodology for fundamental rights impact assessments as of January 2025.

migration and law enforcement are only subjected to limited transparency requirements compared with the other high-risk use cases.[34] Instead of being required to register within a publicly accessible EU Commission database, these systems should only be registered in a non-publicly accessible database and requiring only limited details about the AI system to be made available in this database. In turn, emotion recognition systems, namely AI systems that can purportedly gauge the inner emotional states of persons based upon facial or micro gestures, are banned from workplace and educational settings[35] but allowed within areas such as migration, asylum and border management and law enforcement settings.[36] This means that protections are ironically missing where the power disparity is at its most pronounced, where the risk of potential of abuse and impacts upon human dignity is at its gravest.

What these three lines of argument show is that it is not clear what role human dignity plays in determining what falls within the high-risk category use cases, including within intra-category use cases. Its inconsistency also leaves various aspects of dignity unprotected.

A more general conclusion from Section 2 is that human dignity within the AI Act appears to be one out of many justifications in both prohibited and high-risk use cases. Examining dignity in relation to both categories also reveals that there are interpretative and foundational inconsistencies. In other words, finding concrete meanings of dignity in the AI Act is akin to chasing shadows in the dark.

3. From chasing shadows in the dark to a new research agenda

What might the future trajectory of AI and dignity look like, especially when considering the inconsistencies, gaps and challenges identified here? Three key developments in this regard might be instructive.

First, there have been increasing calls for new rights to address the new harms and vulnerabilities enabled through AI affordances. As I have observed elsewhere, AI harms to human rights can be difficult to pin down and thereby account for – variously due to its non-salient nature – for example, due to the lack of tangible harm, temporality gaps and lack of foreseeability, but also due to issues in proving causation of harms allegedly caused by AI systems.[37] At the same time, the call for new rights echoes much of the same conceptual framing taken by the AI Act, namely in focusing on processes and procedures upstream, rather than in harms that result downstream,

---

34 Article 49(4) AI Act.

35 Article 5(1)(f) AI Act.

36 Annex III, AI Act.

37 Sue Anne Teo, 'How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework' (2022) 40 Nordic Journal of Human Rights 216.

when AI is deployed in real-world settings.[38] It is beyond the scope of this chapter to cover the plethora of new rights called for or to examine each one here in detail, but a sampling includes calls for a right to reality,[39] a right to be free from digital coercion,[40] a right to opt out of automated decision-making,[41] a right to human decision-making[42] and a right to human-to-human interaction.[43] At the same time, Kaminski observes that the risk-based governance around AI signals a normative change – namely that AI brings enormous potential and that its risks should be balanced against the potential for AI to revolutionise healthcare, education, transportation, scientific development and other areas for the better. It might ironically even lead to a right *to use* AI. These calls for rights signal a gap in accounting for and in respecting dignity in the age of AI. Rights are not called for their own sake – rights act as a means to attain dignity.

Second, we can look to the AI Act itself in fleshing out this new research agenda. Article 7 of the AI Act provides that the EU Commission can add to or modify the use cases of high-risk AI systems if they pose a risk to health, safety and fundamental rights by taking into account the following grounds – the extent of the autonomy of AI systems and whether a human can override a decision,[44] high intensity of harms and adverse impacts and its capacity to 'affect multiple persons or to disproportionately affect a particular group of persons',[45] the inability to reasonably opt out of outcomes produced by AI systems,[46] the extent of the imbalance of power,[47] the vulnerability of an individual and the deployer of an AI system, due to, amongst others, 'status, authority, knowledge, economic or social circumstances, or age'.[48] This list is not exhaustive but provides a snapshot of the potential that is left open for the EU Commission to add to the existing list of high-risk use cases. In other words, novel or

---

38 The AI Act itself provides for two rights for affected persons: the right to lodge a complaint with a market surveillance authority (Article 85) and the right to an explanation for individual decision-making (Article 86).

39 Kevin Frazier, 'The Right to Reality' (*Lawfare*, 21 December 2023) https://www.lawfaremedia.org/article/the-right-to-reality accessed 26 December 2024.

40 DigitalCourage, 'German Petition Calls for "a Life without Digital Coercion"' (*European Digital Rights (EDRi)*, 20 November 2024) https://edri.org/our-work/german-petition-calls-for-a-fundamental-right-to-a-life-without-digital-coercion/ accessed 26 December 2024.

41 Blueprint for an AI Bill of Rights, The White House Office for Science and Technology Policy, accessible at: https://www.whitehouse.gov/ostp/ai-bill-of-rights/

42 *Artificial Intelligence, Ethics, and a Right to a Human Decision with John Tasioulas* (2023) https://www.youtube.com/watch?v=16E4-rNSBXY accessed 26 December 2024.

43 Yuval Shany, 'A Right to a Human-to-Human Interaction' (*Institute for Ethics in AI*, 7 November 2024) https://www.oxford-aiethics.ox.ac.uk/blog/right-human-human-interaction.

44 Article 7(2)(d) AI Act.

45 Article 7(2)(f) AI Act.

46 Article 7(2)(g) AI Act.

47 Article 7(2)(h) AI Act.

48 Ibid.

intensified threats to dignity can be potentially accommodated within the high-risk category and be subjected to heightened obligations in future.

These collectively point to new research agendas that lie on the horizon. Research is needed to clarify the role of human dignity in the AI Act, including on how discrete exclusions in the high-risk category impact upon the most vulnerable groups. Further research is also needed to examine and account for new possibilities of harms or vulnerabilities[49] that underlie the calls for new rights or for protections to be afforded through the expanded grounds under Article 7 of the AI Act. This necessitates research agendas that go beyond traditional doctrinal research to include empirical research, including through interviews and surveys with those most impacted by AI systems. This bottom–up approach may be necessary on account of the lack of foreseeability of certain forms of harms by AI and the possibility of new emergent harms due to the use of AI. The research agenda going forward may also need to entail working with new research methods – as risks from AI are less like the laundry list of familiar, foreseeable and traceable risks within other fields that entail risk regulation, such as within the pharmaceutical, medical or environmental sector.[50] Instead, the category of unknown unknowns in relation to AI risk may call for creative methods such as sociotechnical foresight[51] or engaging with imaginaries and shaping of AI narratives.[52]

---

49 Gianclaudio Malgieri and Maria-Lucia Rebrean, 'Vulnerability in the EU AI Act: Building an Interpretation' (Social Science Research Network, 28 November 2024) https://papers.ssrn.com/abstract=5058591 accessed 27 December 2024.

50 Daniel Carpenter and Carson Ezell, 'An FDA for AI? Pitfalls and Plausibility of Approval Regulation for Frontier Artificial Intelligence' (2024) 7 Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society 239.

51 Shakir Mohamed, Marie-Therese Png and William Isaac, 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence' (2020) 33 Philosophy & Technology 659; 'Strategic Foresight' (*OECD*) https://www.oecd.org/en/about/programmes/strategic-foresight.html accessed 21 January 2025. According to the OECD, foresight as a methodology is a 'structured and systematic approach of exploring plausible futures to anticipate and better prepare for change'. Accordingly, this rejects linearity of trajectories and instead 'cultivates the capacity to anticipate alternative futures and an ability to imagine multiple and non-linear possible consequences'. Sociotechnical foresight in turn pertains to imagining futures of technology and society.

52 Laura Sartori and Andreas Theodorou, 'A Sociotechnical Perspective for the Future of AI: Narratives, Inequalities, and Human Control' (2022) 24 Ethics and Information Technology 4; 'AI Narratives: Portrayals and Perceptions of Artificial Intelligence and Why They Matter' (*Royal Society*, 2018) https://royalsociety.org/news-resources/projects/ai-narratives/ accessed 27 December 2024.

4. Gaps and remaining challenges

While the EU's AI Act is the first in the world to comprehensively regulate AI, other regions and countries are catching up.[53] It remains to be seen if the Brussels effect will play out for AI, as it did for the GDPR. Global governance of AI is shaping up, such as through the Council of Europe's Framework Convention on Artificial Intelligence, which is the first international legally binding treaty on AI.[54] However, overall, international governance on AI remains fragmented,[55] both in relation to scope and consensus.[56] In any case, this chapter has argued that while fundamental rights and human dignity feature highly on the EU's regulatory vision for AI, it is far from clear what interpretations of human dignity were employed, nor is it clear what role that dignity plays in relation to the Act. While the possibilities for new research agendas were briefly mapped out in Section 3, gaps and challenges for human dignity in relation to AI remain.

AI is increasingly pushing the technological possibilities and, therein, boundaries between human and computer interactions. Grey areas include anthropomorphic experiences offered through deathbots, therapy bots, companion bots and other chatbots that offer friendship and companionship, especially using the corpus of human experiences offered by large language models. The ease of availability and accessibility of these 'relationships' may on the one hand address the epidemic of loneliness, but may on the other hand significantly alter the nature of human relationships, including opening up possibilities for dependency, manipulation or deception. Humans are fundamentally social animals, finding dignity in belonging to human communities and the pursuit of human relationships. This also explains why punishments such as solitary confinement are also infringements of dignity, in addition to being a disproportionate form of punishment.[57]

However, human bonds of sociality also include a wider circle of concern, including animals and nature. As humanity evolves with technology, human sociality may by

---

53 Daniel Chiang, 'South Korea Passes AI Basic Act amid Political Turmoil' (*Digitimes*, 27 December 2024) https://www.digitimes.com/news/a20241224PD210/government-2027-development.html accessed 27 December 2024.

54 Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, 2024.

55 Emma Klein and Stewart Patrick, 'Envisioning a Global Regime Complex to Govern Artificial Intelligence' (*Carnegie Endowment for International Peace*, 21 March 2024) https://carnegieendowment.org/research/2024/03/envisioning-a-global-regime-complex-to-govern-artificial-intelligence?lang=en accessed 20 January 2025; Peter Cihon, Matthijs M Maas and Luke Kemp, 'Fragmentation and the Future: Investigating Architectures for International AI Governance' (2020) 11 Global Policy 545.

56 Claire Dennis and others, 'What Should Be Internationalised in AI Governance?' (2024) Policy Paper https://www.oxfordmartin.ox.ac.uk/publications/what-should-be-internationalised-in-ai-governance accessed 20 January 2025.

57 I thank the reviewer for raising this point.

extension also include robots or chatbots of silicon and software in the near future. The impacts on human dignity – for example on human exceptionalism, inherent worth and derivative rights – should be closely studied and targeted measures should be undertaken to enhance rather than adversely affect dignity.

On another note, the expansion of 'military AI' also signals a worrying trajectory for human dignity. Despite its comprehensiveness, the EU AI Act is limited by what is allowed for the EU to regulate within its fields of competence. Matters of national security, defence and military remain within the competence of individual member states and are therefore excluded from the scope of the AI Act.[58] Despite the long-standing efforts to ban 'killer robots'[59] and other lethal autonomous weapons systems, progress on governing military AI remains protracted, contested and slow.[60] In the meantime, AI is deployed within military settings in other capacities, for example, to assist in identifying targets. While the precision afforded by AI in target generation can limit collateral damage from attacks, the speed of target generation ironically increases, rather than decreases, the casualties of war.[61] AI can offer a strategic advantage to states, with some even calling it 'the most powerful weapon of our time'.[62] States possessing and using such systems can have an outsized impact on the dignity of individuals, through machine determination of human worth and human life.

Finally, while the EU AI Act aims to narrow the knowledge and access gap to AI by stipulating that AI literacy is essential for those marketing and using AI,[63] access to AI remains unequal in different parts of the globe, notably between the Global North and Global South. AI literacy, as defined under Article 3(56) AI Act, is meant to enable providers, deployers and affected persons 'to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause'.[64] The requirement for a certain level of AI literacy, however,

---

58 Article 4(2) Treaty on the European Union. See also Recital 24 of the AI Act.

59 'Ten Reasons Why It's Time to Get Serious about Banning "Killer Robots"' (*Amnesty International*, 12 November 2015) https://www.amnesty.org/en/latest/news/2015/11/time-to-get-serious-about-banning-killer-robots/ accessed 25 May 2022.

60 Charukeshi Bhatt and Tejas Bharadwaj, 'Understanding the Global Debate on Lethal Autonomous Weapons Systems: An Indian Perspective' (*Carnegie Endowment for International Peace*, 30 August 2024) https://carnegieendowment.org/research/2024/08/understanding-the-global-debate-on-lethal-autonomous-weapons-systems-an-indian-perspective?lang=en accessed 1 January 2025.

61 Yuval Abraham, '"Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza' (*+972 Magazine*, 3 April 2024) https://www.972mag.com/lavender-ai-israeli-army-gaza/ accessed 29 October 2024.

62 Raluca Csernatoni, 'Governing Military AI Amid a Geopolitical Minefield' (*Carnegie Endowment for International Peace*, 17 July 2024) https://carnegieendowment.org/research/2024/07/governing-military-ai-amid-a-geopolitical-minefield?lang=en accessed 1 January 2025 (quoting the words of UK Cybersecurity and Infrastructure Security Agency Director Jen Easterly).

63 Article 4 AI Act.

64 Art 3(56) AI Act.

presupposes that access to AI, if not widespread within the EU, is at the very least attainable and available to large majorities of the population. The lack of equal access to AI in other parts of the world means that benefits that could flow from AI, for example towards improving medical and scientific research, remain disproportionately concentrated in a few countries. The Global Digital Compact, a 'comprehensive framework for global governance of digital technology and artificial intelligence', is one attempt by the United Nations to narrow the AI divide, and digital divide, more broadly speaking. This Compact was adopted in the historic 'once-in-a-generation' Summit of the Future event in September 2024, aimed at fostering international consensus on how to secure a bright and safe future by tackling ongoing and emerging threats. Ensuring human dignity means nothing less than addressing critical accessibility gaps and ensuring equitable access to technologies such as AI, including through provisions on co-funding, scientific cooperation and benefits sharing.[65]

5. Conclusion

While human dignity features in the AI Act, it is far from clear what role it plays. More worrisome, its internal contradictions and inconsistencies respect dignity in name but not in practice. This chapter proposes several measures that can contribute towards shaping a research agenda going forward. These include clarifying the role of human dignity in the Act and examining new possibilities of harms or monitoring new vulnerabilities that could emerge. Further empirical research is critical to this endeavour, alongside more inclusive forms of participation by those most affected by AI systems, including towards shaping the narratives of AI. The path forward may also need to engage with new research methodologies – such as through creative sociotechnical foresight. Finally, while the EU is the first in the world to comprehensively regulate AI, human dignity concerns also lie beyond the horizon of the EU. These include the proliferating use of military AI and the widening digital divide across the globe. AI systems that are increasingly anthropomorphic may also introduce new human dignity concerns hitherto unencountered. As human–machine interactions become increasingly human-like – blurring the distinction between machines and humans – what might this mean for humans and human dignity going forward?

[65] Objective 5, Global Digital Compact.