AI and the Crisis of the Self:
Protecting Human Dignity as Status and Respectful Treatment

Ozlem Ulgen*

> *" … a great deal of the work which was formerly done by human beings is now being done by machinery. This machinery belongs to a few people: it is being worked for the benefit of those few, just the same as were the human beings it displaced. These Few have no longer any need of the services of so many human workers, so they propose to exterminate them! The unnecessary human beings are to be allowed to starve to death! And they are also to be taught that it is wrong to marry and breed children, because the Sacred Few do not require so many people to work for them as before!"*[1]

Over a century since Robert Tressell's prescient novel, the unsettling reality of technology replacing humans continues. A tidal wave of messianic worship for AI, robotics, "Big Data", "Internet of Things" is upon us, mainly articulated through the efficiency paradigm - improving productivity, enhancing human capabilities, reducing time spent on mundane tasks. From algorithms that determine student grades, personalise online marketing, approve financial credit applications, assess pre-trial bail risk, and select human targets in warfare, it seems we are willingly complicit in relinquishing decision-making powers to machines. As Tressell reminds us, we need to understand who "These Few" are controlling the technology and to what purpose it is put rather than completely repudiate technological innovation. The Nobel Prize winning economist, Joseph Stiglitz, warns that without governmental policies that support sharing of increased productivity from AI across society, there will be rising unemployment, lower wages, and acute social inequalities.[2] Against this backdrop of political, social, and economic challenges, viewed from a moral philosophical perspective, unfettered use of AI that diminishes human agency and decision-making powers undermines human dignity. AI is not so easily understood as impacting on human dignity, especially when its justification is presented as some sort of a gain for humanity; saving time, energy, or delegating routine tasks. But human interaction that is mediated by technology penetrates the core of what it means to be human; autonomy and agency to engage in free-thinking, and exercise reasoning, judgement, and choice. This is the moral value of human dignity.

In this chapter I argue that human dignity is a universal moral value that should be at the centre of policy formulation and laws governing AI innovation and impact on societies. Part I sets out concerns about AI innovation and its potential adverse impact on human dignity. Part II considers how diverse cultures, international legal instruments, and constitutional laws represent human dignity as innate human worthiness that is a universal moral value, a right, and a duty. Part III develops two distinct dimensions of human dignity which can be concretised in policy and law relating to AI: (1) recognition of the status of human beings as agents with autonomy and rational capacity to exercise reasoning, judgement, and choice; and (2) respectful treatment of human agents so that their autonomy and rational capacity are not diminished or lost through interaction with or use of the technology.

I. AI Innovation and Impact on Human Dignity

It is impressive how AI is being developed for use in different domains and real-life settings - algorithms determining student grades, personalising online marketing, approving financial credit applications, assessing pre-trial bail risk, and selecting human targets in warfare. But is it morally

---

* Dr. Ozlem Ulgen, Reader in International Law and Ethics, School of Law, Birmingham City University, UK.

[1] Robert Tressell, *The Ragged Trousered Philanthropists* (Penguin 2004), 114.

[2] Lecture by Joseph Stiglitz, "The Future of Work", The Royal Society, 18 September 2018.

right to be deploying AI in such scenarios when inanimate deterministic activities have human consequences? In the UK and Europe the ongoing Covid-19 pandemic has meant students were unable to sit exams necessary for entry into university. Instead, predictive algorithms relying on past student performance and averaging were used to determine grades leading to anomalies, bias, and unfair results.[3] With clear consequences for future educational and employment prospects, it seems immoral and reckless to have algorithms performing grading functions that reduce individual students to mere statistics without applying human judgement. Applying data processing and personal data rights contained under the EU General Data Protection Regulation (GDPR),[4] the Norwegian Data Protection Authority claimed the International Baccalaureate Organisation breached Articles 5(1)(a) and 5(1)(d) in using a profiling algorithm which did not process student grades fairly, accurately, and transparently. It requested rectification of grades.[5]

Pre-trial bail risk algorithms used to assist human decision-making may seem good examples of human-machine interaction. But poor dataset reliance and automation bias on the part of the human result in unfair outcomes. In the United States a pre-trial bail risk assessment algorithm, used by judges to decide whether to release a defendant on bail or to remand them in custody, has come under increasing scrutiny. Among others, the Pretrial Justice Institute, a nonprofit organisation previously advocating use of algorithms instead of cash bail, withdrew support for their use because such algorithms perpetuate racial inequities.[6] And at the extreme end of warfare, an algorithm may be determining who should be selected and attacked as a military objective leading to injury and death.[7] Unfairness, inequalities, restrictions on liberty, and life or death decisions form a concerning list of real human consequences as a result of AI systems.

Reflecting on the relationship between man and technology, throughout human history societal changes occurred as a result of new knowledge and technological innovation. Economic historians refer to four phases of innovation shaping economic development: the mechanisation of textile manufacturing; railroads and steam from 1840 to 1890; steel, engineering, and electricity from 1890 to 1930; and automobile, fossil fuel, and aviation from 1930 to 1990.[8] AI-based technologies fall into

---

[3] Laurie Clarke, "How the A-level results algorithm was fatally flawed" (New Statesman Tech, 14 August 2020), https://tech.newstatesman.com/public-sector/how-the-a-level-results-algorithm-was-fatally-flawed.

[4] European Parliament and Council of the European Union, Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, repealing Directive 95/46/EC (General Data Protection Regulation) (GDPR), (OJ L 119, 4.5.2016), 27 April 2016.

[5] Norwegian Data Protection Authority, "Advance notification of order to rectify unfairly processed and incorrect personal data - International Baccalaureate Organization" (7 August 2020), https://www.datatilsynet.no/contentassets/04df776f85f64562945f1d261b4add1b/advance-notification-of-order-to-rectify-unfairly-processed-and-incorrect-personal-data.pdf.

[6] Pretrial Justice Institute, "Updated Position on Pretrial Risk Assessment Tools" (7 February 2020), https://www.pretrial.org/wp-content/uploads/Risk-Statement-PJI-2020.pdf; Open Letter by Academics, "Technical Flaws of Pretrial Risk Assessment Raise Grave Concerns" (July 2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf?source=post_page.

[7] Ozlem Ulgen, "Technological innovations and the changing character of warfare: the significance of the 1949 Geneva Conventions Seventy Years On" (2019) 3-4 *Journal of International Law of Peace and Armed Conflict* (Humanitäres Völkerrecht) 215-228.

[8] Nathan Rosenberg and L.E. Birdzell Jr., *How the West Grew Rich: The Economic Transformation of the Industrial World* (Basic Books 1986); Chris Freeman and Francisco Louçã, *As Time Goes By: From the Industrial Revolutions to the Information Revolution* (OUP 2001).

the post-1990 economic development phase, the "fourth revolution", that includes information and communication technologies, AI, and autonomous robotics impacting on every aspect of our lives today.[9] Yet a single invention cannot be the sum of our lives, problems, or solutions. The drive towards greater efficiency and increased productivity precipitates the AI innovation ferris-wheel; a never-ending cycle of innovation to counter human fallibility that rewards slavish adoption and punishes the reticent human mind. Byung-Chul Han refers to this as "psychopolitics"[10]; a form of control of the human psyche exerted by technological domination and use of personal data in the pubic and private spheres that alters our minds and behaviour to an extent that undermines our autonomy and agency. If we are constantly having to sync different platforms, update new software, connect systems with systems so that we can access even bigger systems, we are losing sight of ourselves and getting entangled in a techno-bureaucracy purposely constructed by two strange bedfellows: the regulators and the hackers. Both contribute to the crisis of the self.

i) The techno-bureaucracy of hackers and regulators

Hackers want to explore and exploit new technology vulnerabilities to serve their own illicit purposes, thereby increasing demand for higher security measures from regulators. Regulators (seemingly concerned with human well-being and protection of rights) introduce layers of complexity through overlapping and competing non-legally binding and legally-binding rules, ethical principles, and processes contained in global, regional, and national ethical frameworks, standards, and instruments.[11] Meanwhile, private sector corporate entities, the military, and the State continue to development AI under the radar of any enforceable regulation. It is unclear how divergent ethical/legal initiatives apply across jurisdictions and alongside national legislation. The rules, principles, and processes are often impenetrable to the ordinary person. Take for example the legal concept of "responsibility" determining who or what will be held liable for any harm/damage caused by the technology, AI has potential to disrupt the attribution and causation chains unless there is always a human who will be held responsible throughout AI design, development, and deployment stages. Self-learning algorithms and robots present the spectre of harmful and unattributable behaviours which at the same time undermine human agency of foresight, prudence, and judgement in taking action with consequences in mind. Although responsibility is a priority ethical value and legal requirement contained in several global, regional, and national regulatory frameworks, its interpretation and implementation differs. The UK recognises legal responsibility, accountability, and legal liability as key issues in application of the law to AI, but focuses on developing principles of accountability and intelligibility (which are not the same as legal responsibility or liability) with possible review of the adequacy of existing legislation on legal liability.[12] For China, although responsibility is a core principle applicable at both the AI

---

[9] Luciano Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality* (OUP 2014).

[10] Byung-Chul Han, *Psychopolitics: Neoliberalism and New Technologies of Power* (Verso 2017).

[11] See for example, the GDPR (n 4); 2019 EU AI Guidelines - Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence (EU AI Guidelines), European Commission, https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/2; 2019 IEEE Ethically Aligned Design for Autonomous and Intelligent Systems - The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (Final version, 4 April 2019), https://ethicsinaction.ieee.org; 2019 OECD Recommendation on AI, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449; 2019 G20 Human-Centred AI Principles, https://www.mofa.go.jp/files/000486596.pdf; 2019 The Age of Digital Interdependence, Recommendations of UN Secretary General's High-Level Panel on Digital Cooperation, https://www.un.org/en/digital-cooperation-panel/.

[12] 2018 UK House of Lords Select Committee AI Report - *AI in the UK: ready, willing and able?* (16 April 2018), paras 317-318.

development and deployment stages, it is situated within an ethical framework biased towards commercial exploitation for the purpose of domestic economic growth. It is unclear who or what will be held legally responsible, and future policies/laws may contain a commercial intellectual property/trade secrets exemption preventing disclosure of algorithmic models, datasets, and algorithmic reasoning.[13]

ii) Freeing or enslaving?

Whether AI-based solutions to everyday tasks are freeing or enslaving impacts on the crisis of the self. Does AI free-up the human mind to undertake qualitative judgement-based complex tasks instead of routine memorising numbers, memory recall, and mental arithmetic? Or is more time spent frustrated by the technology (how it works, errors it produces, and rectification of errors and seeking redress)? In theory, more AI-assisting jobs should be available leaving routine tasks to machines. In practice, such jobs are few and far between with not enough training offered by employers to make the transition from displacement by machine to human-machine teaming.[14] Among other mental tasks, recall and mental arithmetic stimulate the brain and if we become dependent on technology for the most simplest of tasks, we are enslaved by the technology and forget how to function. Automation bias is a manifestation of such enslavement whereby in human-machine tasks the human operator favours the machine's response over their own judgement with major repercussions for lives and livelihoods.[15] De-skilling may also occur through automata behaviour exhibited in the human reduced to binary responses without independent critical thinking or judgement. Studies show heavy use of digital technologies cause neurological changes that impede comprehension, retention, and deeper thinking.[16] This diminishes human agency and dignity with potentially serious repercussions for other humans. Remote pilots of unmanned armed aerial vehicles, for instance, thousands of miles away from conflict zones viewing video images of targets to select and attack, have been shown to exhibit lack of deeper thinking and moral disengagement. They are less fearful of being killed and less inhibited to kill. They have problems identifying targets, and reduced situational awareness in complex scenarios resulting in civilian fatalities.[17]

---

[13] 2018 China's AI Standardisation White Paper - *China's White Paper on Artificial Intelligence Standardisation* (January 2018, Standards Administration of China) - 人工智能标准化白皮书（白皮书）（2018年1月，中国标准管理局.

[14] See for example,"Millions of Americans Have Lost Jobs in the Pandemic - And Robots and AI Are Replacing Them Faster Than Ever", *Time Magazine,* 6 August 2020.

[15] Mary Cummings, "Automation Bias in Intelligent Time Critical Decision Support Systems" (2004) Collection of Technical Papers - AIAA 1st Intelligent Systems Technical Conference. DOI 10.2514/6.2004-6313; Parasuraman Raja and Manzey Dietrich, "Complacency and Bias in Human Use of Automation: An Attentional Integration" (2010) 52 *Human Factors* 381-410. DOI 10.1177/0018720810376055.

[16] Gary Small and Gigi Vorgan, *iBrain: Surviving the Technological Alteration of the Modern Mind* (Collins 2008); John Sweller, *Instructional Design in Technical Areas* (Australian Council for Educational Research 1999); Erping Zhu, "Hypermedia Interface Design: The Effects of Number of Links and Granularity of Nodes" (1999) 8(3) *Journal of Educational Multimedia and Hypermedia* 331-358; Diana DeStefano and Jo-Anne LeFevre, "Cognitive Load in Hypertext Reading: A Review" (May 2007) 23(3) *Computers in Human Behaviour* 1616-1641.

[17] Lambèr Royakkers and Rinie van Est, "The Cubicle Warrior: the Marionette of Digitalized Warfare" (2010) 12 *Ethics and Information Technology* 289-296; Heather Linebaugh, "I Worked on the US Drone Program. The Public Should Know What Really Goes On", *The Guardian*, 29 December 2013; Matthew Power, "Confessions of a Drone Warrior", *GQ*, 22 October 2013; Chris Woods, *Sudden Justice: America's Secret Drone Wars* (Hurst & Co 2015).

As for being frustrated by the technology, accessing your online bank account in 2020 can be an exercise fraught with technical and security glitches. You need at least two different devices; one to receive a verification code, another to enter the code in order to gain access. Accessing other online accounts for home, work, or personal purposes requires memorising codes, online storage of codes and passwords, or using facial/voice/fingerprint recognition technology. The technology-based solutions have flaws such as high error rates, non-recognition of dialects and accents, bias, and security breaches of stored biometric data.[18] Personal data divulged and stored across different platforms and devices actually leads to a loss of control over what is happening.

The crisis of the self will continue unless we confront issues of control and use of AI, and determine what supports rather than undermines human dignity. Let us now consider how diverse cultures, international legal instruments, and constitutional laws represent human dignity as innate human worthiness that is a universal moral value, a right, and a duty.

II. Human Dignity as a Universal Moral Value, Right, and Duty

There is a long history of philosophical, religious, and legal thinking on human dignity; what it entails and how it manifests. Such thinking reflects on what it means to be human and shows a sensibility towards articulating human worthiness. The ancient Romans used the concept of *dignitas* to differentiate persons of rank and elevated social status from the common people.[19] In Christian theology human dignity developed from the idea that human beings are created in the image of God and therefore possess worthiness and deserve to be treated with reverence.[20] Augustine of Hippo (354-430), the North African bishop, influential to early Christian thinking, considered it important to nurture and value the inner self in order to enable moral rules to emerge. In the thirteenth century St Thomas Aquinas identified rational nature as an intrinsic human quality which leads to personhood and dignity.[21] In Hinduism human dignity is conceptualised as individual for all living things, and not just humans, albeit with different approaches as to how it is attained; ranging from a non-inclusive, class-based conception of human dignity to Gandhi's conception of the equality and dignity of all humans.[22] In Confucianism human dignity functions as ethical conduct and relates to three qualities: benevolence,

---

[18] Rahhal Errattahia, Asmaa El Hannania, Hassan Ouahmanewe, "Automatic Speech Recognition Errors Detection and Correction: A Review" (2018) 128 Procedia Computer Science 32–37; Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (2018) 81 Proceedings of Machine Learning Research 1–15; "Many Facial-Recognition Systems Are Biased, Says U.S. Study", *The New York Times,* 19 December 2019; "Major breach found in biometrics system used by banks, UK police and defence firms", *The Guardian,* 14 August 2019.

[19] Teresa Iglesias, "Bedrock Truths and the Dignity of the Individual" (2001) 4 *Logos: A Journal of Catholic Thought and Culture* 114-134.

[20] The Bible, Genesis 1:26-27 "Then God said, "Let Us make man in Our image, according to Our likeness … So God created man in His own image"; 5:1 "In the day that God created man, He made him in the likeness of God"; 9:6 "Whoever sheds man's blood, By man his blood shall be shed; For in the image of God He made man".

[21] Thomas Aquinas, *Summa Theologica*, Part II-II (Secunda Secundae) Translated by Fathers of the English Dominican Province (Project Gutenburg, 2006), http://www.gutenberg.org/cache/epub/17611/pg17611-images.html.

[22] Jens Braarvig, "Hinduism: the universal self in a class system" in Marcus Düwell, Jens Braarvig, and Roger Brownsword (eds), *The Cambridge Handbook of Human Dignity* (CUP 2014), ch 15; Letter addressed to UNESCO by Mahatma Gandhi, 25 May 1947, in *Human Rights: Comments and Interpretations* (A Symposium edited by UNESCO), UNESCO/PHS/3(rev.) Paris, 25 July 1948, 3.

righteousness, and integrity.[23] Islam recognises human dignity as a status bestowed by God on pious individuals who fulfil their obligations towards God. Human dignity means security and safety in the life of society, sanctity of life, and honour in the conduct of one's public and private life.[24] The African tradition of *ubuntu* is a communitarian-based notion of human dignity relating to social honour, group moral standing, and the capacity to form communal relations.[25]

But the most sophisticated and secular notion of human dignity comes from the eighteenth century deontological philosopher, Immanuel Kant.

i) Kantian human dignity

Kant makes an explicit connection between human existence and human dignity. His categorical imperative urges us to "act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end".[26] For Kant, human dignity is a special status conferred on humans by virtue of their innate worthiness as sentient beings with the capacity to engage in rational thinking to create and abide by rules. From this special status flows certain rights and duties towards development of one's own free will, fettered to avoid gratuitous encroachment on others' free will.[27] A point of objection here is that human dignity appears to exclude those who lack rational thinking capacity (e.g. children, mentally disabled, wrongdoers, criminals, the deceased). But Kant's formulation is not intended to create an elite human class or exclude the vulnerable; rather, it rationalises a secular human-centric approach that distills core elements of humanity which are capable of universalisation.[28] Thus, it is the *capacity* for rational conduct rather than actual rational conduct that entitles all to human dignity. The capacity of others to act rationally to create and abide by rules that protect the vulnerable is a manifestation of human dignity, and Kant provides specific rules regarding the treatment of wrongdoers, criminals, and the deceased.[29]

A conception of human dignity based on human innate worthiness and rational capacity affords universal application and grounding to recognise it as a universal moral value. Innate worthiness and capacity are not dependent on societal, national, State hierarchical structures to confer status in order to set rules governing human exchange and interaction. In recognising the intrinsic worth of humans, Kantian human dignity does not require formal recognition of personhood by any institutional structure, and discounts arbitrarily-determined extrinsic considerations of nationality, religion, wealth, gender, birthplace, or family connections. Neither does wrongdoing nor criminality deny a person's human dignity. Intrinsic worth pre-exists in all humans and is the basis for their special status with rights and duties attached. Some such as Waldron criticise Kant's

---

[23] Luo An'Xian, "Human dignity in traditional Chinese Confucianism" in Düwell et al, *ibid*, ch 17.

[24] Miklós Maróth, "Human dignity in the Islamic world" in Düwell et al (n 22), ch 14.

[25] Thaddeus Metz, "Dignity in the ubuntu tradition" in Düwell et al (n 22), ch 32.

[26] Immanuel Kant, *The Moral Law: Kant's Groundwork of the Metaphysic of Morals* (HJ Paton tr, Hutchinson & Co 1969) 91 para 429.

[27] *Ibid*, 90-91, paras 64-66 [428-429]; 101-102, paras 87-88 [440]; 96-97, paras 77-79 [435-436].

[28] Ozlem Ulgen, "Kantian ethics in the age of artificial intelligence and robotics" (2017) 43 *QIL, Zoom-in (Questions of International Law/Question de Droit International/Questioni di Diritto)* 59-83.

[29] Immanuel Kant, *The Metaphysics of Morals* (Mary Gregor tr and ed, CUP 1996) 105-109, 209-210, 211-213.

emphasis on respect for the innate rather than the person.[30] But Kant situates human worthiness in something innate in order to avoid contested notions of formalised personhood dependent on extrinsic recognition (e.g. by the State, or a community) and which may exclude certain categories of persons. Recognition of innate worthiness then leads to autonomy, and rational capacity to exercise reasoning, judgement, and choice.

ii) Human dignity in international legal instruments and constitutions

References to innate worthiness, human value, and rational capacity are contained in international legal instruments and State constitutions which recognise human dignity as a universal moral value, a right, and a duty. The 1948 Universal Declaration of Human Rights (UDHR)[31] provides an understanding of human dignity based on the Kantian notion; that it is intrinsic to all humans endowed with reason and conscience, and recognisable in humanity as a whole. The Preamble states, "Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world", and "the dignity and worth of the human person", also repeated in the Preamble to the 1966 International Covenant on Civil and Political Rights (ICCPR)[32], and the 1966 International Covenant on Economic, Social and Cultural Rights (ICESCR).[33] Article 1 of the UDHR states, "All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood." Article 22 protects a person's "economic, social and cultural rights indispensable for his dignity and the free development of his personality." Article 23(3) protects the right to just and favourable remuneration for work in order to ensure "an existence worthy of human dignity".

These international legal instruments show that human dignity is a pre-existing status of all humans by virtue of their innate worthiness, providing a rationale for protection of human rights. Human dignity also operates as a guiding principle for interpreting and applying rights. The rationale and guiding principle aspects can be seen in the ICCPR and ICESCR. Article 10 of ICCPR requires that "All persons deprived of their liberty shall be treated with humanity and with respect for the inherent dignity of the human person." The right to education, contained in Article 13 of the ICESCR, is necessary for "the full development of the human personality and the sense of its dignity".

Beyond the universal moral value, several States recognise human dignity as a right and a duty. It is represented as a pervasive norm as well as a duty under German constitutional law. Article 1(1) of the 1949 German Basic Law provides that "Human dignity shall be inviolable. To respect and protect it shall be the duty of all state authority." Article 79(3) prohibits any amendment to human dignity as a State duty.[34] The State has both a negative and positive obligation. "Respect" requires

---

[30] Jeremy Waldron, 'Dignity, Rank, and Rights' The Tanner Lectures on Human Values (University of California, Berkeley, 21-23 April 2009).

[31] UN General Assembly Resolution 217 A(III), 3d Sess, Supp No 13, UN Doc A/810 (1948) (adopted 10 December 1948).

[32] International Covenant on Civil and Political Rights, 999 U.N.T.S. 171 (adopted 16 December 1966, entered into force 23 March 1976).

[33] International Covenant on Economic, Social, and Cultural Rights, 993 U.N.T.S. 3 (adopted 16 December 1966, entered into force 3 January 1976).

[34] 1949 German Basic Law, Basic Law for the Federal Republic of Germany in the revised version published in the Federal Law Gazette Part III, classification number 100-1, as last amended by Article 1 of the Act of 28 March 2019 (Federal Law Gazette I, 404).

the State to refrain from acts that violate human dignity, and it must "protect" individuals "against humiliation, branding, persecution, outlawing" from third party acts.[35] During proceedings of the Parliamentary Council that debated the content of constitutional provisions, Theodor Heuss referred to Article 1(1) as a "non-interpreted thesis" that was an important value yet open to different interpretations.[36] But against the backdrop of acts of dehumanisation experienced under the Nazi regime, the drafters had a clear sense that individual humans needed to be at the centre of State legislation and protection. By placing human dignity in the first article of the constitution and before exposition of fundamental rights, the drafters achieved this objective. It means human dignity is woven into the fabric of legislative interpretation and State structures. It requires constant reference and application to give it substantive meaning and effect in practice.

Aside from representing an overriding constitutional norm and one that is fundamental to protecting rights, human dignity has increasingly been interpreted as a standalone substantive right that guarantees a "dignified minimum existence". It encompasses both the physical existence of a human being as well as the possibility to maintain interpersonal relationships and a minimal degree of participation in social, cultural and political life.[37] The Constitutional Court recognises the right to human dignity means all human beings possess this dignity as persons, irrespective of their qualities, their physical or mental state, their achievements and their social status, or any wrongdoing.[38]

Human dignity as a fundamental value and right has been central to the development of South African constitutional jurisprudence. Section 1 of the 1996 South African Constitution provides that the Republic of South Africa is one, sovereign, democratic State founded on "Human dignity, the achievement of equality and the advancement of human rights and freedoms." It is affirmed as a "democratic value" of the Bill of Rights, and specifically identified as a right under Section 10 which states, "Everyone has inherent dignity and the right to have their dignity respected and protected." The South African Constitutional Court has declared that "dignity is not only a value fundamental to our constitution, it is a justiciable and enforceable right that must be respected and protected."[39] The Court has held that human dignity inherently includes protection of the family;[40] requires protection of the social and economic conditions of vulnerable populations so that State-funded educational benefits extend to certain non-citizens;[41] and requires the State to provide substantial resources in order to realise the right to adequate housing.[42]

---

[35] Bundesverfassungsgericht [BVerfG - Federal Constitutional Court] 1, 97 (104) Order of 19 December 1951; BVerfG 102, 347 (367), Judgement of 12 December 2000.

[36] Theodor Heuss, 4th session of the Committee for Fundamental Constitutional Questions on 23 September 1948, in Deutscher Bundestag and Bundesarchiv (eds), Der Parlamentarische Rat, vol 5/I. Harald Boldt Verlag, Boppard, 1993, 72.

[37] BVerfG 125, 175, Judgement of 9 February 2010; BVerfG 132, 134, Judgement of 18 July 2012. See also, Dieter Grimm, "Dignity in a Legal Context: Dignity as an Absolute Right", in Christopher McCrudden (ed), *Understanding Human Dignity* (OUP 2013) 381-391.

[38] BVerfG 87, 209 (228), Order of 20 October 1992; BVerfG 115, 118 (152), Judgement of 15 February 2006.

[39] *Khosa v Minister of Social Development* 2004 (6) SA [Constitutional Court of South Africa] 505 (CC) para 41.

[40] *Dawood, Shalabi and Thomas v Minister of Home Affairs* 2000 (3) SA [Constitutional Court of South Africa] 936 (CC).

[41] *Khosa v Minister of Social Development* 2004 (6) SA [Constitutional Court of South Africa] 505 (CC).

[42] *South Africa v Grootboom* 2001 SA [Constitutional Court of South Africa] 46 (CC).

Clearly, human dignity is a universal moral value and, in some jurisdictions, is also understood as a right and a duty. This provides justification for placing human dignity at the centre of policy formulation and laws governing AI technologies and innovation. To understand how this can be achieved, let us now turn to developing the content of human dignity as a status and as respectful treatment.

III. Human Dignity as Status and Respectful Treatment

Kant's secular theory provides the basis for developing two distinct dimensions of human dignity which can be concretised in policy and law relating to AI: (1) recognition of the status of human beings as agents with autonomy and rational capacity to exercise reasoning, judgement, and choice; and (2) respectful treatment of human agents so that their autonomy and rational capacity are not diminished or lost through interaction with or use of the technology.

i) Recognition of the Status of Human Beings

a) Agents with autonomy

Recognising that human agents have autonomy relates to how they perceive situations, their ability to take independent action, and to exercise choice. Maintaining human autonomy is clearly of concern for policy formulation and law governing AI. Autonomy as a philosophical concept refers to the capacity for self-government to make decisions and take action. Individual rights represent autonomy as individual freedoms to take action. But autonomy does not operate in isolation from rules or others and therefore needs to be contextualised within a moral framework. Kant's notion of autonomy, which he refers to as "autonomy of the will"[43], involves individual freedom to self-govern by taking morally-informed decisions and actions. Human agents act autonomously to provide reason for taking action, and decipher what is moral and what is not. It is this internal capacity for morally-informed conduct, rather than sanctions imposed by the State, which leads to freedom and inculcates a sense of duty to act morally. This has implications for individuals formulating policy/law, as well as individuals using or interacting with AI.

A "technology-biased approach"[44] to regulation, focusing on AI capabilities and limitations to improve performance, optimise operational efficiency, and identify and rectify any errors or failures, will prove inadequate to recognising human autonomy. Human wants, needs, and values should be incorporated into the AI system design, development, and deployment in order to maintain and recognise human autonomy. Moreover, the AI system should not erode the human internal capacity for morally-informed conduct by imposing technology-only solutions, or altering thinking and behaviour to induce immoral conduct. The EU's AI Guidelines provides a specific example of how autonomy can be protected when it requires that humans should be able to "keep full and effective self-determination over themselves, and be able to partake in the democratic process."[45]

Another way to protect human autonomy is to allow for non-deterministic influences on decision-making such as environment, learning, and critical thinking. Prior to deployment and regulation of AI, rationale needs to be provided for context and appropriateness of use taking account of non-

---

[43] Kant (n 26), 94 para 432, 101 para 440.

[44] Ozlem Ulgen, "User Rights and Adaptive A/IS – From Passive Interaction to Real Empowerment" in HCII Conference Proceedings, in LNCS Series, (Springer 2020), R. A. Sottilare and J. Schwarz (Eds.): HCII 2020, LNCS 12214, 205–217.

[45] EU AI Guidelines (n 11), 12.

deterministic influences. It is at this stage that policy-makers, legislators, and regulators have autonomy to decide what is legally and morally acceptable, and therefore bear responsibility for their actions. AI designers and developers would need to take a "human-centric approach"[46] to think about user awareness, rights, and to represent non-deterministic influences on decision-making. If the latter is not possible, users should be informed of the deterministic decision-making beforehand and given the option for human decision-making.

The potential human user's autonomy is protected if they are able to act influenced by reason; if they can identify the motivations prompting their action; or they can change their motivations if they cannot identify with them. If the human takes action not based on reason; cannot identify the motivations prompting their action; or cannot change their motivations, then these would indicate that autonomy has been lost. Automation bias and human automata behaviour induced by the AI system are examples of how this can occur. Thus, an AI system that relies on binary conditions being met without consideration of context, personal circumstances, or judgement ends up undermining autonomy.

b) Agents with rational capacity to exercise reasoning, judgement, and choice

Rational capacity is a human characteristic that manifests in the ability to exercise reasoning, judgement, and choice. An infinite number of scenarios, human characteristics, circumstantial evidence, environmental factors and combinations of these influence whether and how a person acts and whether the act is moral. Human perception and social interaction enables deciding whether the act is moral, and requires applying rules or principles to that particular situation, not simply as a calculative or performative process but as part of reflective thinking.

To exercise reasoning is to draw conclusions from a set of premises. It is a dynamic, ongoing process that may rely on common-sense presumptions as well as synthetic *a priori* judgements[47] where the predicate is external to the subject and adds something new to our conception of it. Reasoning involves practical reasoning (i.e. what to do), and pure reasoning (i.e. mulling over or pondering in abstract form). To cope with the infinite number of influencers on moral conduct, both types of reasoning are necessary. This is apparent in Kant's conception of reasoning as requiring universality (that a rule guiding moral conduct into action must be capable of being used by others, or universalised), internal capacity for morally-informed conduct, and an ability to engage in deliberative and reflective thinking.[48] Whilst machines and algorithms can engage in practical reasoning and work well with pre-programmed premises and assumptions, they cannot engage in pure reasoning that is whimsical, inconclusive, and resurfacing at a later stage to enable decision-making. The tendency to revert to common-sense presumptions in practical reasoning is a shortcoming of both humans and machines programmed by humans (e.g. "AI is a new innovation that will lead to new challenges"), which Kant criticised as an "emergency help" "when one knows of nothing clever to advance in one's defense."[49] On the other hand, engaging in critical and reflective thinking as to what these AI challenges might be and how to mitigate them is to engage in pure reasoning in order to reach synthetic *a priori* judgements. It is this pure reasoning capacity of humans that should be protected and ring-fenced from AI intrusion.

---

[46] Ulgen (n 44).

[47] Immanuel Kant, *Critique of Pure Reason* (Paul Guyer and Allen Wood trs, CUP 1998) A7/B11.

[48] Kant (n 26), 84 para 421; Kant (n 29), 157 para 6:395; see Kant's noumenal/phenomenal world distinction, Kant, *ibid*, A235- 260/B294-315.

[49] Immanuel Kant, *Prolegomena to Any Future Metaphysics* (Gary Hatfield tr and ed, CUP 2004) 9, para 4:259.

Judgement is the faculty of thinking the particular is contained under a universal rule, principle, or law, and functions as an "intermediary between understanding and reason".[50] Kant identifies two types of judgement: "determinant" (where the universal rule, principle, or law is already known and the particular is easily subsumed under it); and "reflective" (where the particular is known but the universal rule, principle, or law has to be found for it). An example of determinant judgement is knowing that physically assaulting someone is morally wrong and unlawful. But it is reflective judgement that differentiates humans from machines. The value and purpose of human reflective judgement can be illustrated in the following example.

A State official's sworn affidavit states that prison conditions must satisfy a list of legal and ethical requirements to protect prisoners' well-being and, therefore, if the defendant were to be imprisoned they would not be subject to inhumane, degrading, or life-threatening treatment. The existence of the list and its application to prisons are facts. However, it does not follow that the defendant, if imprisoned, would not suffer ill-treatment. There is a difference between what is stated on the list and what is actually implemented in practice. Without evidence of how the legal and ethical requirements are implemented in prisons generally, and in the particular prison relevant to the defendant, and the effects on prisoners, the State official is in no position to make the determination that that particular defendant would not suffer ill-treatment. In fact, the affidavit is worthless as the list of legal and ethical requirements could simply be reeled off by an algorithm as factors that a judge should take into consideration during sentencing. The point is that the human (State official or judge) is required to go beyond determinant judgement and engage in reflective judgement to consider whether the defendant would suffer ill-treatment. This is something that cannot be performed by algorithms. There are no pre-programmable or pre-existing universal rules that can be relied upon. Through reconciliation and calibration of understanding and reasoning, the human decision-maker is able to reach a judgement.

Others point to more specific features of human judgement which cannot be replicated in machines or algorithms. Suchman intimates consciousness as a requisite to exercising judgement when he refers to it as self-direction that cannot be specified in a rule.[51] Weizenbaum refers to it as wisdom which only human beings possess because they have to "confront genuine human problems in human terms".[52] Judgement is often required in the grey areas, the problematic issue points where there is no precedent to follow and no clear-cut answer or solution. This implies a discretionary aspect devoid of orderly rule formation and adherence. But it also captures an invaluable human faculty leading to human solutions that reinforce human dignity in the person exercising judgement and the person affected by it. For example, in a situation where there is an automated decision-making system deciding on an applicant's eligibility for a health test, the system may automatically reject the applicant because it detects incomplete or unclear information. A human decision-maker can exercise judgement to determine the significance of any incomplete or unclear information and therefore decide on an appropriate response which may not involve outright rejection of the application.

Finally, the exercise of choice, as a manifestation of human rational capacity, is the process by which different desires, pressures, and attitudes compete leading to a decision and action. Kant

---

[50] Immanuel Kant, *Critique of the Power of Judgement* (Paul Guyer ed, Paul Guyer and Eric Matthews trs, CUP 2000) 64 para 5:177, 66 para 5:179.

[51] Lucy Suchman, *Plans and Situated Actions: The Problem of Human-Machine Communication* (Xerox Corporation 1985).

[52] Joseph Weizenbaum, *Computer Power and Human Reason: From Judgement to Calculation* (WH Freeman & Company 1976), ch 8.

referred to choice as a competing process that is controlled by the self for higher purpose such as reason or morality.[53] In regulating the use of AI, the EU refers to "meaningful opportunity for human choice".[54] Thus, if a person does not want to use AI for resources or services, alternatives must be provided. Equally, if prior to use or during use of an AI system a person decides that they no longer want to be subject to automated decision-making, they must be allowed the opportunity to opt out or withdraw consent. This reflects enforceable rights under the GDPR and the Council of Europe's Modernised Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108+).[55]

ii) Respectful Treatment of Human Agents

Through use of and interaction with technology humans should not be subjected to forms of treatment which would undermine their human dignity or diminish their autonomy and rational capacity. This also relates to Kant's core notion of human dignity as not treating humans as mere means to ends. It follows that respectful treatment entails recognition of human agents' autonomy and rational capacity. This can manifest in the AI system in several ways: a) respecting human agent rights; b) respecting AI limitations; and c) respecting prioritisation of human needs.

a) Respecting human agent rights

For an AI system to respect human agent rights requires designers and developers of such systems to adopt a "human-centric approach" taking account of rights to privacy, data protection, and fundamental rights. The privacy and data protection rights provided for under the GDPR and Convention 108+ are an obvious starting point due to their reach across the entire lifecycle of a system. These are rights to: not being subjected to automated decision-making; prior consent; prior notification of right to withdraw consent; notification of automated decision-making; access to personal data; access to information on the logic of an automated decision; information on the significance and envisaged consequences of automated decision-making; object to processing of data; lawful, fair, and transparent processing of data; rectification of inaccurate data; withdraw consent; explanation of automated decision; obtain human intervention; express a point of view; and contest an automated decision.[56] Among the fundamental rights most relevant for an AI

[53] Kant (n 26), 101 para 440; Kant, *Critique of Pure Reason* (n 47) A533/B561.

[54] EU AI Guidelines (n 11), 12.

[55] Ulgen (n 44).

[56] *Ibid*.

system to respect are: freedom from torture, cruel, inhuman or degrading treatment;[57] freedom of expression;[58] freedom of thought, conscience, and religion;[59] and freedom of association.[60]

It has been argued that the use of autonomous weapons in warfare is contrary to human dignity and constitutes a form of cruel, inhuman or degrading treatment because such weapons treat humans as disposable inanimate objects rather than ends with intrinsic value and rational capacity.[61] Autonomous weapons are characterised by their use of AI and robotics in order to achieve varying degrees of autonomy in the critical functions of acquiring, tracking, selecting, and attacking targets. Human involvement, either partially or fully, may be removed in any of these critical functions, and from the lethal force decision-making process. Replacing human combatants with AI and robotics means human moral and legal agency is lost. A hierarchy of human dignity is created whereby certain humans are deemed more valuable and priceless than others. The human combatant is protected from harm and their human dignity is elevated above that of the human target. The human target is treated as an inanimate object without any interests; easily removed and destroyed by a faceless and emotionless machine. All individuals targeted and killed by such weapons are entitled to respect for their human dignity. Whether or not they are designated enemy combatants or terrorists, they have rational capacity, possess a moral value of dignity which cannot be replaced by an equivalent, and they cannot lose such status through immoral acts.

Common Article 3 of the 1949 Geneva Conventions provides fundamental guarantees (applicable to both non-international and international armed conflicts) that civilians and *hors de combat* "shall in all circumstances be treated humanely".[62] Enemy combatants are protected under Articles 1(2) and 75 of the 1977 Additional Protocol I to the Geneva Conventions, which refer to "the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience"; and if they do not benefit from more favourable treatment under the Geneva Conventions or the Additional Protocol then they must be "treated humanely in all circumstances". These provisions establish obligations to take account of others' interests, including the human dignity of enemy combatants. Use of autonomous weapons to kill

---

[57] Article 5, Universal Declaration of Humans Rights (UDHR); Article 5, African Charter on Human and Peoples' Rights (AfCHR); Article 5, American Convention on Human Rights (AmCHR); Article 27, American Declaration of the Rights and Duties of Man (AmDR); Article 8, Arab Charter on Human Rights; Article 3, European Convention on Human Rights (ECHR); Articles 4, 7, and 10, International Covenant on Civil and Political Rights (ICCPR).

[58] Article 19, UDHR; Article 19, ICCPR; General Comments 10 [19] (Article 19) and 11 [19] (Article 20) of the Human Rights Committee (CCPR/C/21/Rev.1 of 19 May 1989); Article 9, AfCHR; Article 13, AmCHR; Article 10, ECHR.

[59] Article 18, UDHR; Article 18, ICCPR; Article 9, ECHR.

[60] Article 20(1), UDHR; Articles 21 and 22, ICCPR; General Comment 25 (Article 25) of the Human Rights Committee (participation in public affairs and the right to vote); Article 8, International Covenant on Economic, Social and Cultural Rights; Articles 10 and 11, AfCHR; Articles 21 and 22, AmDR; Articles 15 and 16, AmCHR; Article 11, ECHR.

[61] Ozlem Ulgen, "Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an 'Elementary Consideration of Humanity'?" (2019) vol 17 (2017/2018) *Baltic Yearbook of International Law* 169-196.

[62] *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. USA*), Merits Judgment of 27 June 1986, ICJ Reports 1986, 14. The majority decision held that Common Article 3 expresses "minimum rules applicable to international and non-international conflicts" (para 219), and these rules reflect "elementary considerations of humanity" (para 218).

"wrongdoer" human targets completely bypasses such obligations and represents a modern-day example of Kant's "disgraceful punishments" amounting to "outrages upon personal dignity".

Manipulating a human agent's thoughts so as to distort their freedom of expression, beliefs, and actions would not respect rights to freedom of expression, thought, conscience, and religion. As was alluded to in the EU's AI Guidelines, this could impact on participation in political processes and voting rights. The 2020 UN, OSCE, and OAS Joint Declaration on Freedom of Expression and Elections in the Digital Age notes the alarming misuse of social media by both State and private actors to subvert election processes, including through various forms of inauthentic behaviour and the use of "computational propaganda" (employing automated tools to influence behaviour).[63] Recommendation 1(a)(i) requires States to have in place a regulatory and institutional framework that promotes a free, independent and diverse media, in both the legacy and digital media sectors, which is able to provide voters with access to comprehensive, accurate and reliable information about parties, candidates and the wider electoral process. Thus, the State has a duty to protect human agents from manipulation by AI systems by providing a legal framework within which such systems will operate, and regulating the conduct of third party actors. This is similar to the German constitutional law State duty to refrain from acts that violate human dignity, and to protect individuals against harmful acts of third parties.

Recommendation 2(a)(ii) requires non-State actors, such as digital media and platforms companies, to make a reasonable effort to adopt measures that make it possible for users to access a diversity of political views and perspectives. In particular, they should make sure that automated tools, such as algorithmic ranking, do not, whether intentionally or unintentionally, unduly hinder access to election related content and the availability of a diversity of viewpoints to users. The nature of actions to be taken by private companies is somewhat weakened by the wording "reasonable effort" and "unduly hinder". They should make a "reasonable effort" to adopt measures that make it possible for users to access a diversity of political views and perspectives. What constitutes a "reasonable effort"? For this to have any practical meaning and impact on potential users the digital media/platform company would need, for example, to ensure that: it does not prioritise news outlets from whom it receives advertising revenue or direct funding; small, independent, or foreign news outlets are accessible and not blocked; and users are able to clearly see how to access a variety of sites. "Reasonable effort" could not be discharged by pointing to the comments section as representing a diversity of views and perspectives.

In making sure that automated tools do not "unduly hinder" access to election related content and the availability of a diversity of viewpoints to users, companies may argue that there is no undue hinderance where the information transmitted is blocked on the grounds of preventing crime or legal censorship. But this would need to be tested against the source(s) of information justifying the grounds for blocking, and be balanced against public interest access to information, freedom of expression, and the freedom of the press to undertake investigative journalism. For example, Google's announcement that it would de-rank Russia Today and Sputnik falls foul of the diversity principle and engages in censorship to manipulate public opinion.[64] Ironically, the stated aim was to prevent misinformation yet by acting as gatekeeper and arbiter of information the corporate entity is engaging in distortion of information which undermines the human agent's autonomy

---

[63] The United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, and the Organization of American States Special Rapporteur on Freedom of Expression, Joint Declaration on Freedom of Expression and Elections in the Digital Age, 30 April 2020.

[64] "Google to 'de-rank' Russia Today and Sputnik", *BBC*, 21 November 2017, https://www.bbc.co.uk/news/technology-42065644.

and rational capacity to decide for themselves. Similarly, Facebook's manipulation of news feeds, as part of a psychology study to alter the emotional content of users' posts, treated users as mere means and undermined their agency to engage in free thinking and expression.[65]

Recommendation 2(a)(v) also requires digital actors to be transparent about the use and practical impact of any automated tools they use, including data harvesting, targeted advertising, and the sharing, ranking and/or removal of content, especially election-related content. Thus, as a minimum, companies deploying AI systems must make sure that potential users are made aware of the various uses and impact of automated tools. However, this requirement is caveated by the phrase "albeit not necessarily the specific coding by which those tools operate" so that, unlike the GDPR, there is no automatic right for users to access information on the logic of an automated decision, or information on the significance and envisaged consequences of automated decision-making. This points to the divergence and complexity of non-legally binding and legally-binding rules across different regional instruments alluded to earlier.

b) Respecting AI limitations

AI systems should be designed and developed in such a way that recognises and respects their own limitations (e.g. lack of pure reasoning; undesirability of deployment in certain contexts). This is a means of controlling what AI systems are used for and maintaining human agency. AI systems should not assume they are interacting with inanimate and determinative objects that simply require binary responses. The lack of reflective judgement in machines and algorithms is clearly a handicap for many situations in which answers are less clear-cut, exploratory rather than determinative in nature, and need time for deliberation. A certain level of exchange and interaction is needed to gauge what may or may not be appropriate. A person may be unsure about their options, implications of taking a particular option, their ability to change options later, the consequences from this, or about how the AI system will retain and use their personal data. In accordance with rights under the GDPR and Convention 108+, recourse to a human should be available in such circumstances.

An example of problems relating to respecting AI limitations is an AI-based decision support system designed to certify a person's health status and free movement. The system may have pre-programmed biases that cause rejection of individuals from certain locations or postcodes with high Covid-19 reproduction rates or which are in lockdown. Although the biases are a result of human error, the fact that the system is being deployed in a scenario with high-risk of detrimental outcomes for personal freedoms points to the need to set and respect limitations. Even a human reviewing the refused automated certification may be susceptible to automation bias so that there is minimal exercise of reflective judgement and over-reliance on the AI system's decision that the application should be rejected. The combination of AI and human deficiencies means it may be better to rely completely on a transparent human decision-making process rather than a human-machine process from which it may be difficult to untangle errors, attribute responsibility, and seek redress.

c) Respecting prioritisation of human needs

As already mentioned above under agent autonomy recognition, there should be a "human-centric approach" to AI design and development that prioritises human wants, needs, and values. More specifically, the AI system should enable human agent preferences and choices to curtail its application and use. Van Kleek et al refer to "obstacle respect" in an AI system that understands human agents in a particular way in order to pursue its own goals (e.g. treating the customer as a

---

[65] "Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry", *The New York Times*, 29 June 2014 https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html.

product, either through advertising or data mining in order to generate profit).[66] Such a system would treat the human agent as a means to an end, and prioritise profit generation over data protection, privacy, autonomy, and rational capacity. Unless the human agent is informed of such prioritisation and consents beforehand, it would undermine both the status and respectful treatment dimensions of human dignity.

Another means of respecting prioritisation of human needs is to allocate the distribution of harm resulting from the AI system. Prior to purchase and engagement with the system, there should be full and clear disclosure of how the AI system will distribute harm in unavoidable harm scenarios. For example, in the case of autonomous vehicles, will these be designed to avoid injury to drivers, pedestrians, or other drivers? Or will they adopt a utilitarian approach to minimise casualties and maximise lives saved? Consumers may be reluctant to purchase autonomous vehicles that fail to prioritise the driver's safety, so some car manufacturers have already declared prioritisation of driver safety.[67]

Conclusion

In positioning ourselves ready for the "fourth revolution", we should understand the crisis of the self brought on by the AI innovation ferris-wheel and recognise human dignity represents an important moral and legal norm to focus policy formulation and laws governing AI innovation and impact on societies. Diverse cultures, international legal instruments, and constitutional laws represent human dignity as a universal moral value and, in the case of German and South African constitutional laws, a right and a duty. Kantian deontological ethics provides the most rigorous exposition of human dignity with its consideration of innate human worthiness and rational capacity. It is from these human characteristics that we can begin to understand human dignity as meaning recognition of the status of human beings as agents with autonomy and rational capacity to exercise reasoning, judgement, and choice; and respectful treatment of human agents so that their capacity is not diminished or lost through interaction with or use of the technology. Human dignity means innate human worthiness that justifies the autonomy and rational capacity status of human agents, and their respectful treatment.

AI that diminishes human agency to engage in free-thinking, and to exercise reasoning, judgement, and choice undermines human dignity. Part of the problem stems from a sense of losing control over what human activities will be overtaken or replaced by the technology, and its subsequent impact on human autonomy and rational capacity, which relate to the status aspect of human dignity. Several measures can be adopted to protect human autonomy and rational capacity. First, a "human-centric approach" to regulation whereby AI design and development prioritises human wants, needs, and values, as well as user awareness and rights. Second, representation of non-deterministic influences on decision-making in AI systems, and if this is not possible to inform users of the AI's limitations in terms of deterministic decision-making and allow them to opt out. Third, acceptance and safeguarding of human pure reasoning; that is, the ability to engage in deep and critical reflective thinking to mitigate challenges posed by AI. Reflective rather than determinant judgement differentiates humans from machines and algorithms. Without pre-programmed or pre-existing rules, humans are able to reconcile and calibrate understanding and reasoning in order to reach a judgement. In the exercise of choice, human agents should be

---

[66] Max Van Kleek, William Seymour, Reuben Binns, and Nigel Shadbolt, "Respectful Things: Adding Social Intelligence to Smart Devices", PETRAS IET Living in the Internet of Things, 2018, 1-6.

[67] Michael Taylor, "Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians", *Car and Driver* (online 7 October 2016), http://blog.caranddriver.com/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/.

provided with alternatives if they do not want to use AI for resources or services, including prior to use and during use, and for the opportunity to opt out or withdraw consent.

Respectful treatment of human agents so that their capacity is not diminished or lost through interaction with or use of the technology involves three main categories of policy and legal requirements. First, is to respect human agent rights. Designers and developers of AI systems will need to adopt a "human-centric approach" that takes account of rights to privacy, data protection, and fundamental rights. The GDPR and Convention 108+ provide for extensive privacy and data protection rights throughout the lifecycle of a system, ranging from prior consent for automated decision-making to its contestation. Consideration of fundamental rights to freedom from torture, cruel, inhuman or degrading treatment; freedom of expression; freedom of thought, conscience, and religion; and freedom of association, may necessitate policy decisions not to deploy AI systems in certain circumstances (e.g. warfare; where the AI system can manipulate a human agent's thoughts to distort their freedom of expression, beliefs, and actions, or participation in political processes and voting rights). The Joint Declaration on Freedom of Expression and Elections in the Digital Age recognises particular harmful effects from automated tools used to influence public opinion and access to media, and seeks to include both State and non-State actors in a regulatory framework. Second, recognising the limitations of AI systems (e.g. deterministic decision-making; lack of pure reasoning) may lead to early policy decisions not to deploy them in open-ended, uncontrolled circumstances where there is a high-risk of detrimental outcomes to humans. This can be due to the need for human deliberation, interaction, and reflective judgement (e.g. user's uncertainty about options and their implications; AI-based decision support system certifying a person's health status for free movement purposes). Finally, respecting prioritisation of human needs can be achieved by adopting a "human-centric approach" to regulation and design of AI systems to prioritise information about the system, consent, and allocation of distribution of harm as key features.