PRoSPEro

Pedagogy of Robotics in the Social Professions in Europe

Scoping Paper [no.]
[Topic]

[Authors]
[Institution]

This scoping paper is intellectual output [Ox] of the PRoSPEro project
11 December 2019

Funded by the
Erasmus+ Programme
of the European Union

Scoping paper:

P. Share, J. Pender & L. Taylor, IT Sligo, Ireland

> Why would anyone need to worry about such issues in relation to AI, isn't it just like any other technology? (Guttman 2018a)

Introduction and overview

This paper addresses artificial intelligence [AI] and robotics in tandem as there are many overlapping and common issues between the two. But it is also important to tease out the specific issues that relate to robotics and to social robotics in particular. Some of these derive from the anthropomorphism of many robots and what that may mean for the interaction between humans and technology. Others relate to how (social) robots co-exist and interact with humans in everyday life situations (such as work, education and care).

Sources of information include ... **[insert]**

The paper is divided into six parts:

Part 1 Background

Part 2 Principles of regulation

Part 3 Regulatory challenges of autonomous technologies

Part 4 Regulatory spaces - transnational

Part 5 Regulatory spaces - nation-based

Part 6: Transnational NGOs and corporations

Part 1 Background

The significance of AI/robotics

On the face of it, if one relied on the interest displayed by key societal actors (eg media, academia, government, investors), then we would have to recognise the immense significance of recent developments in artificial intelligence. As the European Union's High Level Group on AI (2018) suggests:

> Artificial Intelligence (AI) is one of the most transformative forces of our time, and is bound to alter the fabric of society … AI is key for addressing many of the grand challenges facing the world, such as global health and wellbeing, climate change, reliable legal and democratic systems and others expressed in the United Nations Sustainable Development Goals

[1] This paper deals with algorithms, AI and robotics (at times collectively termed 'autonomous technologies') together. While these terms are not synonymous, similar issues range across all these fields. Wachter et al (2017) argue that: 'concerns about fairness, transparency, interpretability, and accountability are equivalent, have the same genesis, and must be addressed together, regardless of the mix of hardware, software and data involved'.

As the theory and practice of AI has developed in recent years (Boden 2018) the prospect of autonomous robotics and artificially intelligent systems has become more viable, albeit that the reality (especially in relation to social robotics) can often be at (disappointing) variance with media, business or even academic depictions (Mols 2018).

While the developments in AI and robotics will potentially provide great economic and social opportunities, they are also likely to have significant impacts upon the functioning of society, posing practical, ethical, legal and security challenges - many of which are not yet fully appreciated or understood (Judge Business School 2017). The Royal Society (2017, p. 84) notes that:

> the disruptive nature of machine learning brings with it challenges for society. Its technological capabilities enable new uses of data, which challenge existing data governance systems. Its new applications raise questions about public confidence and acceptability.

Raso et al (2018) similarly argue that in many cases the 'formal and informal institutions that govern various fields of social endeavour are ill-suited to addressing the challenges posed by AI'. There is a strong view, shared by many governments, NGOs and supranational bodies, and individual critics, that autonomous technologies already pose significant social and political challenges.

Analysts have identified and explored numerous broad questions about the influence of automated digital systems on society. Beer (2017) points to the following in relation to algorithms, but a similar observation could be made in relation to AI in general, or to robotics:

> when thinking of the power of the algorithm, we need to think of the impact and consequences of code, we also need to think about the powerful ways that notions and ideas about the algorithm (AI, robotics) circulate in the social world

In other words, it is not only the technologies themselves that have social impacts, but our beliefs, attitudes and metaphorical structures that relate to those technologies.

This points undeniably to the need for a social analysis of these novel technologies. There is now extensive discussion of both the social shaping of AI and robotic technologies (eg Markoff 2015; Broussard 2018) and of the impact of those technologies at global, community, workplace and individual level (eg Kiggins 2018; Vincent et al 2015; Willcocks & Lacity 2016).

Commentators and researchers offer widely diverse assessments of the nature and dimensions of the impact of autonomous and robotic technologies, but most conclude that there is a significant technological shift in train with future societal impacts that exceed those of many other recent and contemporary technologies. This 'technological determinist' view (Neven & Leeson 2015, p. 85) is most dominant in writing about robots and society. There is a smaller literature on how these technologies themselves have been socially and culturally shaped (see for example Sone 2017 on Japanese responses to robots) and less still on the 'mutual shaping' processes at play (Neven & Leeson 2015; Winkle et al 2019).

By contrast, there is a strand of thought that denies the novelty and significance of these developments. It is pointed out by critics that AI is not particularly new, but has had a long

history, dating back in the contemporary technological context to the 1950s (Boden 2018). Further, the fascination with thinking and autonomous machines can be traced back to at least the eighth century BC (Cave & Dihal 2018), while the construction of relatively sophisticated automatons and mechanical humans has been taking place since at least the fifteenth century (Mols & Vergunst 2018, pp. 19-20). It is argued by some that, despite the 'hype', genuine progress towards real (or 'general') artificial intelligence is negligible, notwithstanding that some of the achievements with 'narrow' AI are impressive.

Critics point out that much of the success of AI's signature achievements (such as the defeats of leading chess, *Jeopardy* or Go players) is at least partly dependent on 'brute force' computing and/or significant human input (Broussard 2018). Data-mining and pattern recognition (amongst the more common uses of AI) are personified by some as little more than the application of sophisticated existing statistical techniques, and then often poorly interpreted (Smith 2018). As one online commenter remarks:

> stop the anthromorphisation of what should be called 'automation'. There is very little 'intelligent' in 'artificial intelligence' and zero real learning in 'Machine Learning'... using both terms is pure hype (comment by Collard on article by Guttman 2018a)

It is indeed a feature of the history of AI that there have been numerous failures and dashed expectations over the last seven decades of AI research and application (Boden 2018). This contributes to continued scepticism of its novelty or its breakthroughs. For Sloane (2018), it is necessary to guard against buying into AI 'hype'. She notes that 'critics outline the prevailing limits of deep learning and the unreliability of machines completing tasks'. She predicts the 'AI hype to cool off into an AI winter soon'. It remains to be seen how the future of AI plays out.

When it comes to *robots*, Mols and Vergunst (2018, pp. 11-12) note that there has been a similar volatility in levels of enthusiasm:

> [since] the mid-20th century [society's] interest ... has waxed and waned. Sometimes we experience episodes of 'peak robot', and the promises about their potential are sky-high. But then we are inevitably disappointed when these promises don't come true, and robots fade into the background

It appears - if the amount of academic and popular publishing on the topic is any guide - that we are currently in a phase of 'peak robot' (and AI) with daily publication of stories in both traditional and social media; academic articles and books on the topic. But, as alluded to above, often the actual performance of social robots (in particular), is (in most people's eyes) very limited and far-removed from depictions in robotics companies' promotional materials or the less critical academic sources (Mols 2018). At least three well-funded and high profile consumer robot initiatives (*Kuri*, *Jibo* and *Anki*) have been discontinued (Mayfield Robotics 2018; Crowe 2019; Meekins 2019).

Nevertheless, even if the impacts of these technologies are over-egged, speculative and contradictory, there is a strong emerging argument that societies would be well-advised to at least explore adapted or new types of institutional innovation (eg laws, regulations, regulatory regimes) in order to manage, regulate and ultimately control these developments. Technologies that have already been developed and applied (such as certain types of algorithmic decision-making, or relatively widespread use of some social robots)

have led to important concerns over regulation and have raised key ethical challenges (Eubanks 2018; EPIC 2020).

The debate over the novelty or significance of AI does not only have implications for the development of the technologies; it is also important when it comes to regulation. If the technology is not significantly differentiated or novel, then perhaps it can, or should, be regulated through existing regulatory mechanisms. Data-mining, for example, could be monitored (in Europe) through the GDPR data protection system; or self-driving cars (if and when deployed into public streets) could be governed by existing traffic laws and product liability regimes. Thus any new regulatory intervention of autonomous technologies will need to be predicated on a demonstrated and unmet need for regulation.

The 'AI arms race'

The field is a dynamic one, with many powerful players, such as national governments in major states and the world's largest technology companies - sometimes referred to as GAFA[2]- and their Chinese counterparts[3]. The United Nations has determined that at least four of the five Permanent Members of the Security Council 'are placing AI at the centre of their current grand strategies', leading to suggestions of a global AI arms race (ITU 2018, p. 52).

For Thompson & Bremmer (2018) the stakes are high:

> A country that strategically and smartly implements AI technologies throughout its workforce will likely grow faster, even as it deals with the disruptions that AI is likely to cause. Its cities will run more efficiently, as driverless cars and smart infrastructure cut congestion. Its largest businesses will have the best maps of consumer behavior. Its people will live longer, as AI revolutionizes the diagnosis and treatment of disease. And its military will project more power, as autonomous weapons replace soldiers on the battlefield and pilots in the skies, and as cybertroops wage digital warfare.

Intense international competition for leadership in AI research and application[4] may lead to an over-emphasis on technological development and a neglect of regulatory and ethical issues. Cave and Ó hÉigeartaigh (2018) draw our attention to the impact of the globally competitive AI 'race' in itself and identify three sets of risks:

- risks posed by race rhetoric alone (an insecure environment that hinders dialogue and collaboration)

[2] The world's largest digital technology companies: Google, Apple, Facebook and Amazon. Along with Microsoft, also the world's most powerful brands (Forbes 2019)

[3] BATX (Alibaba, Baidu, Tencent & Xiaomi) (https://www.telegraph.co.uk/news/world/china-watch/technology/new-technology-giants/)

[4] For example, in early 2019 both the UK and Irish governments announced the establishment of major new doctoral training centres in AI, with a view to the production of hundreds of AI scientists.

- risks posed by a race emerging (failure to take proper safety precautions)
- risks posed by race victory (concentration of power in the hands of one group)

The 'race' is often expressed as a global contest between China, the US and the EU. Thompson & Bremmer (2018) describe China's substantial policy shift into the AI arena, even as the US government has been slower to react. The Chinese state has relatively untrammeled access to unprecedented amounts of personal data and has enrolled key Chinese IT companies - Baidu, Alibaba, Tencent, iFlytek - into the national AI effort.

AI and robotics - 'hopes and fears'

See also:

the concept of "sociotechnical imaginaries" aims to capture "collectively held, institutionally stabilized, and publicly performed visions" (Jasanoff 2015)

In order to understand some of the motivational drivers of regulation of AI and robotics it is useful to consider some of the dominant public discourses in relation to these technologies. Robots and AI have long been a staple of science fiction comics, books, TV shows, games and films. Carter (2018) notes that AI and robots have been portrayed in film in ways that range from 'benevolent companions' to 'hostile machines bent on total destruction of humankind'. Such positive/utopian and negative/dystopian orientations can be thought of as 'hopes' and 'fears'.

Fast and Horvitz (2016) analyse references to AI in the globally-influential *New York Times* newspaper. They list reported 'hopes' as: improvements to work, education, transportation, health-care, decision-making and entertainment, as well as a beneficial singularity event and beneficial merging of humans and AI. As 'fears' they identify: the loss of control of powerful AI, negative impact on work, military applications, a lack of ethics in AI, a lack of progress in AI, a harmful singularity event, and harmful merging of humans and AI (Fast & Horvitz 2016).

Cave & Dihal (2019) argue that people's broad conceptions of AI and robotics have implications outside of their media or cultural interest: 'perceptions of AI's possibilities, which may be quite detached from the reality of the technology, can influence how it is developed, deployed and regulated'. The range of fiction and non-fiction work that engages with the possibility of intelligent and/or autonomous machines is extensive and expanding. To this end, they have 'categorises[d] some of the fundamental hopes and fears expressed in imaginings of artificial intelligence (AI), based on a survey of 300 fictional and non-fictional works'.

It is unremarkable that they identify perceptions of AI/robotics to be either utopian or dystopian (Amos & Page 2014). Most are reflections of an imagined future, albeit often with a contemporary moral resonance. Cave and Dihal identify eight key themes - four positive ('hopes') and four negative ('fears') - in relation to these technologies.

> *Hopes:* Immortality, ease, gratification, dominance
>
> *Fears:* Inhumanity, obsolescence, alienation, uprising

These are mediated by the extent of control of humans over the AI, how far: 'humans believe they are in control of the AI determines whether they consider the future prospect utopian or dystopian'.

Cave and Dihal (2019) argue that this underlying dichotomy helps to explain people's 'extreme' responses to AI:

> the hopeful narratives show the extent to which AI is perceived to be a master tool that can solve problems that have preoccupied humanity throughout history. It represents the apotheosis of the technological dream that humans can use machines to create a paradise on earth. But at the same time, ... the idea of creating tools with minds of their own contains (in our imaginings) inherent instabilities. Losing control over such agential machines, or the world they create, is the primary source of the exaggerated fears

These cultural references relate to the 'Anglophone West'. This suggests that other cultural formations or traditions may involve other perceptions of AI and robotics. Sone (2017) makes this argument in relation to Japanese robot culture.

The social impacts of AI

> The notion of the 'algorithm' is now taking on its own force, as a kind of evocative shorthand for the power and potential of calculative systems that can think more quickly, more comprehensively and more accurately than humans. As well as understanding the integration of algorithms, we need to understand the way that this term is incorporated into organisational, institutional and everyday understandings (Beer 2017)

A case has been made for the regulation of AI based on what are perceived to be negative social impacts, in areas such as invasion of privacy, denial of other fundamental human rights, unintended consequences of algorithmic decision-making and a broad range of other ethical concerns. As Guttman (2018b) points out:

> AIs have a substantial ownership of control of a process whose outcomes would normally be accounted to a human. AIs take actions and make decisions that can alter the course of a person's life dramatically..

Similarly, Wachter et al (2017) note that:

> [AI] systems can make unfair and discriminatory decisions, replicate or develop biases, and behave in inscrutable and unexpected ways in highly sensitive environments that put human interests and safety at risk

Sloane (2018) notes that numerous states, think tanks and private companies are seeking to develop positions in relation to ethics and AI and that this reflects a 'global discourse on the ethical and social issues evolving around data, automated systems, artificial intelligence technology and deep learning more generally'.

Examples and illustrations of these (often negative) social processes are outlined in some detail in book-length studies such as O'Neil (2016), Noble (2018) and Eubanks (2018). They

are systematically explored in a report (prepared for the government of Canada)(Raso et al 2018) on AI and human rights.

O'Neil (2016) examines the operation of algorithms in a number of key areas of social life: higher education; on-line advertising particularly as related to private on-line colleges; employee selection and recruitment; employee surveillance; credit checks and insurance services; and voting/polling/electoral behaviour. These are examples of the 'politics of algorithmic sorting' (Beer 2017) that relates to 'the capacity of the algorithm to create, maintain, or cement norms of abnormality' with the power to 'make choices, classify, sort, order and rank' and to determine what matters to decide and what to render (in)visible.

Noble (2018) examines the operation of search engines (such as Google) and argues that they incorporate and reflect racist and sexist assumptions and misrepresent, in particular, women of colour. These in turn can contribute to the escalation of gender and racially-driven harassment and violence and contribute to further marginalisation of those already marginalised.[5]

Eubanks (2018) analyses the emergence of the 'digital poorhouse' through an extensive analysis of three algorithm-based social interventions in the US: provision of welfare services in Indiana; allocation of homelessness services in Los Angeles County, California; and child and family services in Allegheny, Illinois[6]. She also reveals the social impacts of algorithmically-shaped social systems and suggests how they can be challenged, primarily through social movements.

But inevitably and necessarily the 'power of algorithms/AI/robotics' must be linked to existing social patterns of power:

> to understand the sorting power of the algorithm … we need to understand the associations, dependencies, and relations that facilitate those algorithmic processes and their outcomes – rather than seeing the algorithm as carrying social power (Beer 2017)

[more needed on this background argument]

PART 2: Principles of regulation

The governance of technology

What is 'governance'?

> (it) include(s) the full range of processes of control and management that take place within and between states, in public agencies and private firms, and other social organisations. Governance involves directing or setting goals, selecting means, regulating their operation,

[5] For more on Noble's book, see video at https://nyupress.org/books/9781479837243/

[6] For an alternative analysis of the Allagheny Family Screening Tool (AFST) see Chouldehova et al (2018). Some of the authors of that paper were directly involved in the development of the algorithm-based tool.

> and verifying results. It therefore encompasses formal, legal processes and structures of regulation, as well as forms of institutional oversight and informal processes of self-governance (Chubb et al 2018, s. 2)

Some of the recent innovative technologies/applications that raise major issues of regulation and governance include: nanotechnology, genetics, geoengineering, fintech, autonomous vehicles, algorithms and social media platforms

New technologies are often located in an 'institutional void' and unregulated. When broader society responds, it is often in ways that do not adequately address the issues: 'there are too many examples of where governance has come too late and failed to ensure the public value and legitimacy of new technologies (Chubb et al 2018, s. 2) For example, the 'reactive' approach to social media (where Silicon Valley tech companies have supported a 'hands-off' stance on regulation) is now leading to major social, economic and political issues for individuals, communities and governments. What is frequently called for now is 'anticipatory regulation' (Mulgan 2017) which sees a 'closer relationship between regulators and innovators, greater experimentation and more open dialogue' (Chubb et al 2018, s. 2) that takes place earlier in the cycle of research and innovation and has potentially a greater capacity to shape the development of a technology.

An important point in relation to *power*: 'on issues of emerging science and technology, members of the public are not just concerned with questions of risk and safety. They are also interested in matters of political economy: who is likely to win and lose? (Chubb et al 2018, s. 2).

Chubb et al (2018) identify ten dominant themes and approaches in the literature on the regulation of new technologies. These are briefly outlined below. All of these can also be discerned in the emergent regulation of AI and robotics.

1. *Regulation, laws & standards*

   There is extensive discussion of the need for anticipatory and adaptive regulation; for more experimentation in regulatory approaches; for a stronger focus on outcomes; and for greater collaboration between countries on the development and alignment of regulatory responses.

2. *Risk, risk assessment & cost-benefit analysis*

   The dominant approach in these studies is based on an assessment of risks, costs and benefits, expressed in primarily economic terms. Another group of studies argue that emerging science and technologies pose fundamental challenges to traditional cost-benefit models because relevant risks and benefits are normally uncertain and impossible to calculate in advance

3. *Ethics*

   ethics were frequently mentioned in the literature, either as a justification for a particular approach to governance, or in a more reflective mode, as in the ways that ethical considerations shape and affect governance approaches in particular contexts and cultures.

4. *Public engagement and public understanding*

   'there is a sharp divide between those studies which focus on the need for *improved public awareness and understanding* (often of perceived benefits of the technologies in question); and those that place the emphasis on more deliberative forms of *public engagement and dialogue*'

5. *Anticipatory governance*

   a stronger focus on anticipatory or'upstream' modes of governance and technology assessment, earlier in the cycle of research and innovation.

6. *Government, business and university strategy and incentives*

   promoting or incentivising the development of new science and technologies

7. *Imaginaries, narratives, norms, values and trust*

   how societies develop perspectives on the development of new science and technology, and the role that public norms, values and imaginaries plays in these processes

8. *Self-regulation and self-governance*

   governance can be seen to be internalised within the science and technology development process

   Also CSR

9. *Precaution*

   an approach to governance shaped by a set of questions: First, are the costs and risks of the new technology acceptable and does it have significant benefits? Second, do these benefits solve important problems and could these problems be solved in some other, less risky, way? Finally, what are the long term economic and political consequences of introducing the technology?

10. *Experimentation*

    often related to participatory initiatives of various kinds, designed to test and develop new, more inclusive models of governance.

uncertainty

Soft law:

Marchant and Allenby (2017) explore the role of soft law in governing emerging technologies, arguing that there are at least ten different reasons why nations may seek to harmonise their oversight of a specific technology. A new generation of more informal international governance tools are being explored, often grouped under the term "soft law." They include private standards, guidelines, codes of conduct, and forums for transnational dialogue.(Chubb et al 2018, s. 2).

Public awareness:

Marris and Rose (2010) review numerous initiatives which have sought to engage members of the public in decisions concerning bioscience and biotechnologies. These initiatives have multiple motivations. In some cases, the justifications are normative – that participation is a right. In other cases, they reflect a desire to reduce conflict, (re)build trust, and smooth the way for new innovations. In others the motivations are substantive: the assumption being that participation could lead to innovations that perform better in complex real-world conditions, or that may be more socially, economically, and environmentally viable.

(this could relate to participatory UX for example)

Also see RRI - Responsible Research and Innovation (EU): anticipation, inclusion, reflexivity and responsiveness

We will discuss three approaches to regulation:

a) Human rights approach (Europe)
b) Ethics approach (US)
c) Capability approach

Ethics and human rights approaches to regulation

Arguably, ethics and human rights based approaches to regulation derive from separate (though potentially linked) philosophical principles. These can also be connected to opposing views about the relationship between technology, innovation and the state/civil society.

Sloane (2018) argues that:

> A focus on just 'ethics' and 'bias' does not necessitate an acknowledgement of the historic patterns of unequal power structures, discrimination and multi-facetted social inequalities that *cause* algorithmic and data 'bias'.

This approach points to structural inequality and human rights issues.

Raso et al (2018) argue that: 'Human rights law provides an agreed set of norms and a shared language and institutional infrastructure for helping to ensure that the promises of AI are met and its greatest perils are avoided'.

SO: what would human rights based approaches to the regulation of AI/robotics look like? How might they be an alternative/improvement on 'ethics'-based approaches?

Human rights frameworks of regulation offer 'agreed set of norms', 'shared language' and 'global infrastructure'

The AI Now Symposium addresses many of the human rights issues

RASO the key source also AI Now

Ethics-based approaches

Discussions of ethics in relation to innovative technologies derives in part from previous discussions in inevitably contested and emotive areas such as reproductive technology, genetic manipulation and rationing of drug treatment. [NB see the other scoping paper that is specifically on ethics]

In the US context, there is a specific and identifiable strand of research on the "ethical, legal and social implications" (ELSI) of new technologies - reflecting the development and investment in this field through the 1990s, in parallel with the Human Genome Project. The same model was then extended to nanotechnology, synthetic biology and other emerging technologies, and has also been mirrored in other countries, including Canada, Finland, Netherlands and South Korea. In response to ELSI-type initiatives, there is also a more critical strand of work, suggesting that this approach can lead to a reductionist and narrow framing of public values and concerns, and play an instrumental function in limiting or ameliorating societal opposition to new technologies. A further set of studies build on an ethics-based approach to link to wider currents of anticipatory governance. (Chubb et al 2018)

Capability approach

The capability approach has emerged from the works of Sen (1984) and Nussbaum (2000; 2011). It has been advanced as a possible ethical framework within which AI and robotics may be regulated. Coeckelbergh (2009) refers to 'co-care', 'co-caring' and the 'co-cared for' as the type of caring relationships that may emerge through future human-robot care provision interactions.

It may seem odd - in the context of providing a scoping review of the myriad ways in which AI and robotics may be regulated in the provision of future health and social care - to invoke a philosophical framework that draws its intellectual and theoretical inspiration from

predominantly Aristotelian philosophical ideas and concepts that seek to promote government backed policies geared towards fostering human fulfillment and nourishment and what it means to live a good human life in the year 21st century, but this is precisely what a growing number of contributors to discussions on AI, robotics and care have been advocating in recent years (Coeckelbergh, 2009; Borenstein and Pearson, 2010; Laykyte, 2013; Kamishima, Gremmen and Akizawa, 2018).

Interest in social justice, improving the quality of life and well-being are central components of the capability approach. The capability approach, at least as developed by Nussbaum, identifies ten core principles that need to be underpinned by states so that humans are equipped and supported in their quest to achieve personal fulfillment throughout the life course. For a detailed list of these click here. The central tenet of these is that they are human centric and human centred and intended to maximise human fulfillment, improved quality of life and well-being.

Specifically in relation to the widely predicted negative impacts AI and robotics may have on human employment and employability, Nussbaum's tenth capability principle (b) may provide a future regulatory guidance regarding the manner in which social robots are designed and deployed in work settings as assistants to human carers because humans "...hav[e] the right to seek employment on an equal basis with others". Moreover, states must ensure that: "...being able to work as a human, exercising practical reason and entering into meaningful relationships of mutual recognition with other workers" is facilitated institutionalised (Nussbaum, 2011, p.78). Kamishima, Gremmen and Akizawa (2018) suggest that the capability approach ought to be used to help identify the ways in which AI and robots enhance and/or diminish human capabilities. Attempting to highlight the limits of AI robotics entrepreneurship, the authors opine that this suggestion is, inter alia, built of the work of Coeckelbergh (2009) who recommends that discussions around replacing human health care provision with AI alternatives should be premised on a re-calibrated capability approach framework that seeks to re-imagine and delineate what in fact constitutes good care: "The problem is not the technologies themselves, are not replaceability as such, and not the (potential violation of) the principles of privacy or autonomy alone, but the question what good care and the good life is: for us as humans, for us in this context, and for us as the unique persons that we are" (Coeckelbergh, 2009, p. 190).

For Coeckelbergh (2010), defining what actually constitutes care is the critical determinant of how humans should design and deploy AI and social robots providing or contributing to the delivery of human care. He identifies what he refers to as "deep" and "shallow" care. For Coeckelbergh (2010), there is no good reason why robots will not be capable of providing "shallow" care in the guise of performing manual tasks but that "deep" care - that care he defines as necessitating emotional, intimate and personal engagement - is beyond the current design possibilities (and quite possibly desirability) of social robotics. Coeckelbergh invites a more exhaustive deconstruction of the notion "care" and how this may, if all, be reconstituted along the lines he suggests in a future world of care mediate by human-robot interactions.

An exploration of the manner in which the regulation and definition of what amounts to and characterises care is a potentially fruitful arena to unpack in terms of future regulation of personal care/social robots.

What constitutes human to human care?

Part 3 Regulatory challenges of autonomous technologies

Questions of definition

In order to regulate something, it must first be clearly defined. This is particularly the case in relation to highly controlled technologies such as medical devices (O'Dwyer & Cormican 2017).

Dickinson et al (2018, p. 3) note the difficulty of defining the relevant terms in this debate, particularly that of 'robot'. Similarly, Sloane (2018) notes that:

> the discourse employs a problematic confusion of the terms 'AI', 'deep learning', 'machine learning', 'automated systems' and so on. This prevents more productive conversations about the abilities and limits of such technologies

Confusion or ambiguity over terms such as 'AI' and 'social robot' do and will provide a barrier to effective regulation.

Defining AI

For Raso et al (2018) AI is an umbrella term that covers:

> a variety of computational techniques and associated processes dedicated to improving the ability of machines to do things requiring intelligence, such as pattern recognition, computer vision, and language processing

Russell and Norvig (1995), in an influential textbook, categorise AI as:

1. systems that think like humans (eg, cognitive architectures and neural networks);
2. systems that act like humans (eg, pass the Turing test, knowledge representation, automated reasoning, and learning),
3. systems that think rationally (eg, logic solvers, inference, and optimisation); and
4. systems that act rationally (eg, intelligent software agents and embodied robots that achieve goals via perception, planning, reasoning, learning, communicating, decision-making, and acting)

The European Commission (2018, p. 1) offers the following definitions and observations:

1. Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.
2. AI-based systems can be purely software-based, acting in the virtual world (eg voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (eg advanced robots, autonomous cars, drones or Internet of Things applications).
3. We are using AI on a daily basis, eg to translate languages, generate subtitles in videos or to block email spam.

4. Many AI technologies require data to improve their performance. Once they perform well, they can help improve and automate decision making in the same domain. For example, an AI system will be trained and then used to spot cyber-attacks on the basis of data from the concerned network or system.

Defining a 'robot'/'social robot'

Hasse et al (2019) note that robots are 'notoriously hard to define, both due to rapid changes in their material components and to conceptual diversities over time and across disciplines'. They refer to the 'dual history' of robots: as automated labour saving technology, and as real or imagined automata in literature, film and other media. They suggest that these two strands of robot conceptualisation are now beginning to merge and blur, such that 'modern understandings of the robot are now caught at the confluence of imagination and machination'.

For Dickinson et al (2018, p. 3) a robot must be embodied to some extent (so cannot be a software programme, or a smartphone).

The European Parliament (2017) has proposed a definition and classification of 'smart' robots that takes into consideration the following characteristics:

- the capacity to acquire autonomy through sensors and/or by exchanging data with its environment (inter-connectivity) and the analysis of those data
- the capacity to learn through experience and interaction
- the form of the robot's physical support [or embodiment]
- the capacity to adapt its behaviour and actions to the environment.

For COMEST (UNESCO 2017) Contemporary robots can be characterised by four central features:

- mobility, which is important to function in human environments like hospitals and offices
- interactivity, made possible by sensors and actuators, which gather relevant information from the environment and enable a robot to act upon this environment;
- communication, made possible by computer interfaces or voice recognition and speech synthesis systems
- autonomy, in the sense of an ability to 'think' for themselves and make their own decisions to act upon the environment, without direct external control.

The EU-funded project *RoboLaw* conceptualises a social robot as 'a personal care robot' (PCR). It recommends that International Standard Organisation (ISO) standards be used to provide soft regulation of PCRs, in particular standard ISO13482: 2014. This defines a PCR as a 'service robot that performs actions directly towards improvement in the quality of life of humans, excluding medical applications (cited in Robolaw, 2017, p.179). ISO13482: 2014

identifies three types of PCRs: i) mobile servant robots; ii) physical assistant robots; iii) person carrier robots (Robolaw, 2017, p.179).

Robolaw (2017) suggests that the most relevant EU legislative instrument is contained in Directive 93/42/EEC as amended by Directive 2007/47/EC. These Directives were generated to define and describe medical devices, rather than robots *per se*. A medical device is an 'instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer' (cited in Robolow, 2017, p.178).

COMEST (2017) suggests the disaggregation, on the grounds of ethical distinctions, of deterministic and cognitive robots. Deterministic robots are defined as a robot whose behaviour is determined by a programme (algorithm) that controls its actions. Responsibility for the actions is, according to the authors, 'clear, and regulation can largely be dealt with by legal means' (COMEST, 2017, p.7). Cognitive robots' behaviour can only be 'estimated' by statistical analysis and is thus unpredictable. Therefore, the authors suggest: 'the responsibility for the robot's actions is unclear and its behaviour in environments that are outside those it experienced during learning (and so in essence 'random') can be potentially catastrophic. So assigning responsibility for the actions of what is partly stochastic machine is problematical' (COMEST, 2017, p. 7).

[summarise this section]

Targeted or generic regulation?

If the need to regulate in some way is established, there will then arise the question as to whether the most appropriate approach is *generic* (applied to AI/robotics in general), or *targeted* (applied to particular uses of AI/robotics in, say, the elder care field). Are there some novel aspects of AI that do merit special regulation (such as AI in the credit sector; or AI in the care sector) and would more targeted types of regulation, rather than generic ones, based on general ethical or human rights principles, be more useful? Or, are both generic principles and sector-specific types of regulation required? Also, as Wachter et al (2017) usefully ask: how can parallels between emerging systems be identified to set accountability requirements'? [expand] This seems to call for close attention to both generic and sectoral issues.

Key regulatory challenges

According to Mamun (2018) 'understanding what counts as not only legal but also ethical and moral in the field of artificial intelligence is a question which confounds both the state and the market' - implying that there are numerous challenges ahead. This is complicated by the forces of innovation and market-driven change that underpin this emergent field of activity.

Notwithstanding the caveats outlined earlier about novelty and uniqueness, which do merit systematic interrogation and debate, we can identify some of the regulatory challenges posed by AI and robotics. These include the speed of change; the correct balance between 'innovation' and regulation; public participation and the challenges of AI's inscrutability; the

challenge of human/machine 'meshing'; and, consequently, identification of the most appropriate and effective methods of regulation.

It has been argued that 'the speed at which technologies such as AI grow and develop makes effective regulation incredibly difficult' (Consumers International 2018). Certainly, the pace of AI research and application has increased dramatically in recent years, particularly as giant global technology corporations (such as Alphabet/Google, Facebook, Microsoft and Amazon) have aggressively entered the field.

[so, need to look at the implications of this]

Danaher (2017, p. 121), from a legal/ethical perspective, notes the:

> well-known 'control dilemma' associated with the launch of any new technology with significant impact potential. During the early phases of development, the technology will be easy to control and change in response to feedback, but its social effects will be poorly understood But during later phases, as the technology becomes more ubiquitous and its social effects (possibly) better understood, it will be effectively impossible to control and change. This presents policymakers and innovators with a difficult choice. Either they choose to encourage the technological development, and thereby run the risk of profound and uncontrollable social consequences, or they stifle the development in the effort to avoid unnecessary risks

As Mamun (2018) asks: how can we maintain both the 'trustworthiness' and 'innovation-friendliness' of AI? What is an appropriate level and style of regulation that will protect the public, but not stifle development and innovation - and associated investment of economic and human capital - in the field? Guttman (2018b), from within the AI industry), calls for 'minimum viable regulation/ethics' to 'ensure that we keep a sustainable strong pace in our AI research and development'.

An important question posed by Mamun (2018) is: how can we involve the public in the design of AI (given its opacity and technical difficulty)? Studies of social robotics, for example, have argued for the importance of participatory user-centred design (Winkle et al 2019) ('participatory UX'), as has the IEEE in relation to automated and intelligent systems [A/IS] more generally (IEEE 2019b, p. 3), but this can be challenging within a rapidly developing, highly technical and commercially competitive and sensitive field of technological development: particularly when companies hold proprietary rights over algorithms (EPIC 2020).

The inscrutability and opacity of AI-based processes renders the regulation of AI difficult. For Pasquale (2015, in Beer 2017) 'we are living in a black box society where power is increasingly expressed algorithmically'. This is a world where 'software may be taking on some constitutive or performative role in ordering the world on our behalf'.

Another challenge to regulation is the meshing of human/AI agency. Where does automated judgement end and human judgement begin? How do we address 'co-decision making' when both humans and AI/robots are involved? (Wachter et al 2017). How and when can one override the other? This becomes more complex as algorithmic modes of thought are 'folded back' into human thought and practices (Christian & Griffiths 2016).[see issues related to 'legal personality' - Turner 2019 writes on this]

A fundamental, if often highly technical question is: which is a more effective approach to regulation of AI/robotics? Is it a set of binding laws/regulations, such as GDPR or existing consumer law? Or is it to be found in codes of ethics and self-regulation (as favoured by the technology industries). Or does the answer lie in a fundamental human rights approach? Is the answer to be found in a combination of these?

Philosopher Thomas Metzinger of Mainz University in Germany, has been amongst those highly critical of what he terms 'ethical greenwashing' by companies such as Facebook. They can, he suggests, 'organise and cultivate ethical debates in order to delay, postpone, avoid…policymaking and regulation. And that is something you find everywhere right now' (Matthews 2019). He argues that regulation needs to be removed from the control of industry and returned to the public realm, including universities.

Who should develop and implement any regulation? The international NGO Consumers International suggests that to ensure 'greater transparency, stronger enforcement and more corporate accountability … collaboration will … be key, with consumer organisations, civil society, business and policy-makers working together to develop answers to the questions raised by AI' (Consumers International 2018).

There are important national and sectoral differences in emphasis (explored in detail below). It is argued (Wachter et al 2017, Bedoya 2018) that in the USA there is a focus on design, education and self-regulation and/or sectoral regulation whereas in Europe there is a clear trend towards a 'hard' regulatory approach with legally-enforceable rights. [this may reflect broader differences in regulation between USA/Europe]

The necessary focus points of AI/robotics regulation

According to Guttman, the following are seven aspects of AI that need to be regulated:

1. *Algorithmic Bias and Fairness.* When an AI makes decisions and takes actions that reflect the implicit values of the humans who are involved in coding, collecting, selecting, or using data to train the algorithm. [for the complexity involved in notions of 'fairness' see Liu et al (2108)]

2. *AI Safety.* An example here are adversarial attacks. For example, Neural networks can be fooled. How can we manage such vulnerabilities in AI?

3. *AI Security.*Hacking a self-driving car or a fleet of delivery drones poses a serious risk. Whole electricity nets and transport systems benefit from autonomous decision making and optimisation, they need to be secured at the same time. How can we secure AI systems? [see also concerns about 'fake news', forgeries and 'deep fakes' (Chesney & Citron 2018)

   – also data access, privacy, sharing and security a huge issue in health and social care

4. *AI Accountability.* Who is accountable when an entire process is automated. For example, for self driving cars, when accidents occur, who can be accounted for? Is it the manufacturer of the car, the government, the driver of the car, or the car itself?

5. *AI Quality Standardisation*. Can we ensure that AI behaves in the same way for all AI services and products?

6. *AI Explainability.* Can or should an AI be able to explain the exact reasons of its actions and decisions?

7. *AI Transparency.* Do we understand why an AI has taken certain actions and decisions? Should there be a requirement for automated decisions to be publicly available?

There are also issues to do with sustainability (the production chain) and IP issues for materials created by AI (see *AI Now* on sustainability issue)

A crucial issue is **Privacy** and invasion of same by algorithmic systems

Also, for example, should there be a law to require any AI system to identify that it is NOT a human? (Walsh 2016)

Hybrid systems, where AI/robotics mesh with human action, are going to be challenging to regulate

'Deep fakes':

> an incredibly easy-to-use application for DIY fake videos—of sex and revenge porn, but also political speeches and whatever else you want—that moves and improves at this pace could have society-changing impacts in the ways we consume media. The combination of powerful, open-source neural network research, our rapidly eroding ability to discern truth from fake news, and the way we spread news through social media has set us up for serious consequences (Cole 2018)

Also important to assess how best to capture the **positive** applications of AI: eg voice activated AI to provide financial services to those with literacy challenges

Taxation of robots - financial measures to ameliorate/prevent/pay for the displacement of human employees by robot alternatives (see report from Mady Delvaux MEP)

See Abbott & Bogenshneider (2018) on taxation of robots and automation (argue that taxing robots is justified as large proportion of current taxation comes from labour)

Argue that tax allowances for accelerated capital depreciation of machinery (including robots and automation) should be discontinued/reduced

[can/should we think of taxation as a form of regulation?]

The particular regulatory challenges of robotics

Within the overall field of AI, there are specific challenges in the regulation of robotics and, within this, of social robotics.

The Robolaw report posits that there are a number of indirect rights ramified by the use of PCRs, including fundamental human rights contained in the EU Charter of Fundamental Rights. For example, Articles 25 and 26 refer to the right to participate in social and cultural

life. Introduction of a PCR into the life of a person with a disability or an older person may impinge on these rights. Article 20, advancing the principle of equality, may be undermined around people's access to and use of PCRs. Other human rights that may be threatened include those to: independent living; participation in community life; equality and access; privacy; and bodily integrity. This suggests that PCRs may be regulated by the laws of unintended consequence.

A range of new human rights will most likely emerge around PCRs. In particular, the right not to be cared for or to use personal care robots. The authors of the report are unequivocal: 'no rights may be acknowledged to robots' (2017, p. 190).

An arena of possible secondary regulation of personal care robots is that of the insurance sector. Existing EU-wide legislation on product liability can readily be transposed to robots. Robolaw (2017) recommends future compulsory third or first party insurance on personal care robots.

Entities/spaces of regulation

Needs a general introduction. Can also be linked to broader histories of technology regulation: eg communications; air travel &c

Who are some of the key entities likely to be involved in the regulation of AI and/or social robotics, and in what types of spatial contexts do they operate? These include:

1. Supranational bodies, such as the EU, UN and the OECD
2. National/state government agencies, with regulatory powers based in legislation (for example, data protection) (in Ireland, Office of the Information Commissioner)
3. National/state government advisory/consultative bodies. For instance, the United Kingdom House of Lords Select Committee Report on AI in the UK (2018) or the Infocomm Media Development Authority (IMDA) in Singapore
4. Government agencies more broadly (such as healthcare bodies, environmental bodies, police, taxation, military &c)
5. Tech companies (such as Google and Facebook) who may develop their own codes of ethics, principles &c
6. Employees within tech companies, who seek not to work on particular areas (eg Microsoft employees writing open letters against work on ICE projects in US)
7. Consumer bodies/coalitions - such as Consumers International
8. Standard-setting organisations - in tech areas (eg robotics) and in non-tech areas (eg social work)
9. Professional and industry bodies - eg IEEE code of ethics
10. Data protection law - particularly GDPR

11. Intellectual property law (copyright, patents, trade secrets, proprietary knowledge) and other legal instruments
12. Legal actions (eg against algorithms)
13. Impact assessment processes (ag Algorithmic Impact Assessment)

As AI and robotics are relatively novel and emergent technologies, it is likely that regulatory initiatives and regimes will vary across different nation states. At the same time, there is likely to be international convergence, particularly as the companies and research groups involved are likely to be global in scope. In addition, transnational entities, whether 'political' or based in civil society (such as international professional bodies) are increasingly becoming involved in matters of regulation. It is thus most likely that, as in areas such as telecommunications and air travel, globally-recognised regulatory mechanisms and structures will be developed.

Part 4 Regulatory spaces - transnational

While a number of supranational bodies have begun to focus on the use of new technologies, their concerns have so far primarily focused on issues related to privacy, security and data protection. Both the United Nations Charter based institutions, and treaty-based monitoring bodies have, in parallel with regional institutions such as the European Parliament and the Council of Europe, begun to address the role of Artificial intelligence in society.

UN human rights bodies, particularly through the focused work of Special Rapporteurs to the United Nations, have begun to look at the impact of social robotics and AI. This process of responding will intensify as human rights bodies and other transnational institutions begin to give greater consideration to the uses of social robotics in society. The process of developing an overarching legal framework, derived from human rights principles and norms, to regulate the use of social robots, might well develop in parallel with the more significant attention currently given to the regulation of the commercial and military use of robotics and Artificial technology. While the latter fields of commercial and military uses of robotics and AI is examined and highlighted in this survey, the priority here is to locate human rights responses to social robotics in their social care and civilian uses.

United Nations (UN)

When António Guterres, the UN Secretary General first took office, some of his first remarks to the General Assembly highlighted the need for a more comprehensive response to challenges brought by new technologies. As a candidate for the position of Secretary General, his “concise written vision statement" affirmed the requirement to understand global mega-trends, while recognising that ‘globalization and technological progress fostered extraordinary economic growth and created conditions for unparalleled reduction of extreme poverty and generalized improvement of living standards’ (Guterres, 2016). The UN Centre for Artificial Intelligence and Robotics was established in 2016 as ‘the first Centre on Artificial Intelligence and Robotics within the United Nations system’ (UNICRI, 2016). Furthermore, a number of new initiatives from the United Nations, including the UN Secretary-General’s Strategy On New Technologies (2018), which was followed by a High-level Panel on Digital Cooperation was established by United Nations Secretary-General António Guterres on 12 July 2018, and more recently a Report of the UN Secretary-General’s High-level Panel on Digital Cooperation, The Age of Digital Interdependence (2019).

One recent proposal to this panel offered that international governance of AI should be anchored to a regime under the UN which is 'inclusive (of multiple stakeholders), anticipatory (of fast-progressing AI technologies and impacts), responsive (to the rapidly evolving technology and its uses) and reflexive (critically reviews and updates its policy principles)' (A Proposal for International AI Governance, 2019). They have recommended that the international governance of AI ‘could help coordinate existing international law on AI, forecast future developments, risks and opportunities, and fill critical gaps in international governance’ (A Proposal for International AI Governance, 2019).

The UN System Chief Executives Board for Coordination, which is the coordination forum of the United Nations system, and is chaired by the UN Secretary-General, has also begun to

focus on the development of AI technologies. Through the High-Level Committee on Programmes they have decided to proceed with the three-step approach to pursue UN system-wide engagement on AI capacity development, and to draft a system-wide AI engagement strategy across the UN system (High-Level Committee on Programmes. Report of the 35th Session, 2018). They have proposed to,

1. develop a common UN system position and shared guiding principles on AI technologies that would help define the internal strategic direction of the UN system in its assistance to Member States.

2. based on the guiding principles, articulate a system-wide framework on AI technologies in order to encourage and guide integrated action within the UN system.

3. based on this system-wide framework, elaborate recommendations and concrete actions towards a capacity building programme for developing countries.

This three-step approach proposes development of a UN system-wide engagement on AI capacity building. The Board for Coordination have requested the International Telecommunication Union (ITU) to present a draft system-wide framework for the Committee's consideration at its 36th session in 2019.

The UN has also recognised the role of AI technologies in achieving the Sustainable Development Goals (AI for Good, 2017). António Guterres has suggested that digital and technological transformations 'offer us powerful new ways to achieve our shared commitments to each and every one of the Sustainable Development Goals (SDGs)' (Strategy On New Technologies, 2018).

UN Charter based bodies

UNESCO

The World Commission on the Ethics of Scientific Knowledge and Technology of UNESCO (COMEST) produced a their 'Report of COMEST on Robotics Ethics' (2017), to highlight the ethical challenges inherent in the futures of robotics and humans. The Report's recommendations pertain to ethical issues, and includes a focus on robots used in healthcare settings, including with concern for children with autism, and for the elderly (COMEST on Robotics Ethics, 2017). The Report identified a role for robots in stimulating the cognition of a dementia patient, executing day to day tasks, and also assuring security (risk of falling, heart failure) at home or while in care (COMEST on Robotics Ethics, 2017: 129).

The Human Rights Council

A Report of the United Nations High Commissioner for Human Rights to the Human Rights Council, identified the concept of a "digital divide", which refers to,

> the gap between individuals, households, businesses and geographic areas at different socioeconomic levels with regard to their opportunities to access

information and communications technologies (ICTs) and to their use of the Internet for a wide variety of activities (HRC, 2017).

The Report focused on the "gender digital divide", which 'refers to the measurable gap between women and men in their access to, use of and ability to influence, contribute to and benefit from ICTs' (HRC, 2017, para 3). The Report observed that the 'offline population is disproportionately poor, rural, older and female, and the gap between them and those who have access to the Internet is widening steadily' (HRC, 2017, para 4).

Furthermore, women are among those marginalized groups that 'may remain trapped in a disadvantaged situation, thereby perpetuating inequality', because of the absence of access to new technologies, including internet access (HRC, 2017, para 12). A human rights based approach proposes, 'establishing and maintaining key principles such as accountability, equality and non-discrimination, participation, transparency, empowerment and sustainability' (OHCHR, Leaving No One Behind in the 2030 Development Agenda, 2016). One concern identified among the issues raised for women in the Report is algorithmic discrimination and bias, with an acknowledgement that 'there may be disproportionate and disparate impacts on certain groups facing systemic inequalities, including women within those groups' (HRC, 2017, para 41). The Report observed there was, 'evidence of gender-based discrimination in the targeting of job-related advertisements online', and is an issue that needs to be address in the development and deployment of new technologies (HRC, 2017, para 41).

A subsequent Report to the Human Rights Council detailed the right to privacy in the digital age, and focused on the human rights implications of the uses of new technologies by states, including the use of Artificial technologies (HRC, The right to privacy in the digital age, 2018). The Report proposed that the international human rights framework 'provides a strong basis for shaping the responses to the manifold challenges arising in the digital age' (HRC, 2018, para 58). Focusing on data protection and privacy, the Report also recommended to states that they, recognize 'the full implications of new technologies, in particular data driven technologies for the right to privacy but also for all other human rights' (HRC, 2018, para 61).

Special Rapporteurs

A number of significant developments of Special Rapporteurs mandated under the UN Charter have responded to the development of new technologies, and have focused on issues related to care for the elderly, freedom of expression and privacy, and concern for those experiencing poverty and exclusion.

Care for the Elderly

The Independent expert on elder persons, provided one of the more significant Reports in relation to the social impact of the use of Artificial technologies, and observed 'Robots and artificial intelligence will radically transform our lives, including the concept of care of older persons' (HRC, 21 July 2017,para 12). Furthermore, the Independent Expert stated, new technologies, including assistive devices, built-in environmental applications and robotics 'are gaining traction as cost-effective and efficient solutions to the increased need for individualized support for older persons' (HRC, 21 July 2017,para 14). Three areas are

projected to be transformed by the advent of these innovations, namely, 'to help monitor the behaviour and health of older persons; to assist them or the caregiver in their daily tasks; and to provide for social interactions' (HRC, 21 July 2017,para 15).

The COMEST Report (2017) has also recognised that an 'area where the use of social robots and robots as companions is in a considerable rise concerns the area of healthcare, especially for elderly people. In the context of demographic projections' (COMEST Report (2017: 32). Similarly, Nils Muižnieks, the Council of Europe Commissioner for Human Rights, outlined her concerns regarding the right of older persons to dignity and autonomy in care (Muižnieks, 18 January, 2018). This followed on foot of a recommendation from the Committee of Ministers of the Council of Europe on care of the elderly, including defining principles of care, and models of good practice that involved integration of assistive technology into care provision (CoE, CM/Rec(2014)2, 19 February 2014).

Freedom of expression

The Special rapporteur on freedom of expression, David Kaye, provided guidance on the implications of AI technologies and the uses of robotics in society (UNGA (29 August 2018) UN Doc. A/73/348). The report prepared by the Special Rapporteur sought to identify the human rights legal framework relevant to AI, and also to 'present some preliminary recommendations to ensure that, as the technologies comprising AI evolve, human rights considerations are baked into that process' (UN Doc. A/73/348, para 2). In a similar vein, the Special Rapporteur on freedom of expression, has highlighted concerns with private companies ability to provide online moderation processes and tools, and self regulation around online safety (HRC (6 April 2018) UN Doc A/HRC/38/35).

UN treaty based bodies

The Convention on the Rights of Persons with Disabilities (CRPD)

The Convention on the Rights of Persons with Disabilities has become a source for defining what role Artificial technologies might play in social care, and a source of legal definition in international law (Human Rights Council, 21 July 2017: para 16). New emphasis has been given to the text of the Convention which states that,

> [...] persons with disabilities have access to a range of in-home, residential and other community support services, including personal assistance necessary to support living and inclusion in the community, and to prevent isolation or segregation from the community (CRPD, 2006: art 19.b).

The Text further states 'Community services and facilities for the general population are available on an equal basis to persons with disabilities and are responsive to their needs' (CRPD, 2006, art 19.c). Similarly, states that there is an obligation on States to take effective measure to ensure personal mobility and afford persons with disabilities the greatest possible independence, by facilitating 'access by persons with disabilities to quality mobility aids, devices, assistive technologies and forms of live assistance and intermediaries, including by making them available at affordable cost' (CRPD, 2006: art 20.b). These provisions within the Convention have contributed to the conclusion that with the advancement of new technologies, there may emerge a 'right to assistive technology',

including that afforded by Artificial technologies and robotics (Human Rights Council, 21 July 2017: para 17).

Rosa Kornfeld-Matte, the Independent Expert on the enjoyment of all human rights by older persons, detailed a number of related measures across the human rights framework, that may shape a 'right to assistive technology' for older persons in the future. The independent expert proposed that effectively designed robots 'could help meet the increasing demand for care in a safer and more responsible, sustainable manner by reducing the prevalence of elder abuse and violence against older persons' (Human Rights Council, 21 July 2017: para 89). Furthermore, that assistive technology is an essential measure 'to enable older persons to live independently and to participate fully in all aspects of life, on an equal basis with everyone, everywhere, needs to be affordable and accessible' (Human Rights Council, 21 July 2017: para 89).

The Committee on the Rights of Persons with Disabilities (CRPD) has noted 'with concern the challenges faced by persons with disabilities in acquiring necessary mobility aids and assistive devices, including assistance technologies' (CRPD, 25 September, 2017). The Committee 'recommends that the State party adopt measures to facilitate the acquisition of necessary mobility aids and devices, including assistance technologies' (CRPD, 25 September, 2017). Similarly, the Committee recommended to Luxembourg in compliance with the Convention on the Rights of Persons with Disabilities, that based on grounds of equality and non-discrimination (art. 5), and accessibility (art. 9), that the state party ensure that 'information and communications, including information and communication technologies, are accessible to persons with disabilities, on an equal basis with others' (CRPD, 10 Oct 2017).

The European Union and other pan-European bodies

Europe may be ahead of other major AI/robotics research centres (eg USA, China) in relation to progress with regulation.

At a conference in Helsinki in February 2019, Jan Kleijssen, the Council of Europe's Director of the Information Society and Action against Crime Directorate, proposed three stages in the legal and ethical response to AI technologies for the Council of Europe. Firstly, to 'carefully examine existing regulatory frameworks and assess where gaps exist'. Secondly, to 'continue to develop sector-specific recommendations and other non-legislative measures (such as recommendations guidelines, professional codes of conduct)' (Kleijssen, 2019). And finally, to 'explore the feasibility of a legal instrument that sets a general framework for the development, design and deployment of AI in conformity with our standards' (Kleijssen, 2019).

EU Parliament

Significantly, the Parliamentary Assembly published a Recommendation about Technological convergence, artificial intelligence and human rights, in which it explicitly calls on the Committee of Ministers to, 'define the framework for the use of care robots and assistive technologies in the Council of Europe Disability Strategy 2017-2023 in the framework of its objective to achieve equality, dignity and equal opportunities for people with disabilities' (EU Parliament, Rec n°210, 2017).

The Council of Europe Parliamentary Assembly's Committee on Legal Affairs and Human Rights decided to create a new Sub-Committee on artificial intelligence and human rights (Council of Europe 2019).

A 2017 European Parliament report called for the creation of a European Agency for robotics to supply public authorities with technical, ethical, and regulatory expertise, and a voluntary ethical Code of Conduct. (this probably has been superseded by much of what is here)

The European Parliament's Committee on Legal Affairs (JURI) published a report with recommendations in 2017 (EP report with recommendations to the Commission on Civil Law Rules on Robotics [27 January 2017]). This report highlights the following important measures, initiatives and possible legislative provisions required to regulate and safeguard Europeans in a world increasingly mediated by AI and robotics. It calls for the EU Commission to consider the designation of a European Agency for Robotics and Artificial Intelligence 'in order to provide the technical, ethical and regulatory expertise' to public actors throughout the member states on technological developments in robotics. Identified issues in the report include:

- classification and definition of 'smart robots'
- registration of smart robots to be managed by a new EU Agency for Robotics and Artificial Intelligence.
- civil law liability measures such as an obligatory insurance scheme based on an obligation of producers of AI and robots to take out insurance for autonomous robots
- access to code and intellectual property rights
- a Charter on Robotics ought to be developed 'in consultation with a European-wide research and development project dedicated to robotics and neuroscience, must be designed in a reflective manner that allows individual adjustments to be made on a case-by-case basis in order to assess whether a given behaviour is right or wrong in a given situation and to take decisions in accordance with a *pre-set hierarchy of values* (our italics)
- development of Ethical Conduct for Robotics Engineers
- right to privacy: robotics engineers and designers should guarantee that individuals are not personally identifiable and that 'human informed consent should be pursued and obtained prior to any man-machine interaction'.
- researchers should maximise the benefits of their work and minimise harm. Risk analysis should be undertaken and should be precautionary and proportional.
- a Code for Research Ethics Committee (REC) should be established. This needs to be established on the following principles: independence, competence, transparency and accountability.
- the REC should be composed of men and women with a broad experience of and expertise in the area of robotics research. Those appointed should be a mixture of

people with ‘an appropriate balance of scientific expertise, philosophical, legal or ethical backgrounds, and lay views’. Moreover, at least one member of REC must have a specialist knowledge in ethics. Additionally, members with specialist knowledge of health, education or social services with a focus on research and methodological activities should be recruited.

- REC will ensure that there is a continuous and rigorous monitoring on AI and robotics research throughout the lifetime of the research
- designers of AI and robotics should take into consideration *European* values of dignity, autonomy and self-determination, freedom and justice before and after the design process. The principle of no harm should prevail. Trustworthiness, privacy and traceability ought to be at the centre of design principles. Considerations around uncertainty and unpredictability in human-robot interactions need to be taken into account. Design protocols need to be devised emphasising health and safety safeguards in human-robot engagements. AI and robotics designers should obtain an positive opinion from a REC before testing a robot on humans.
- users of AI and robots shall be allowed to use a robot without risk or fear of physical or psychological harm. Users have expectancy rights about the purpose and function of AI and robots. Users should know that AI and robots have perceptual, cognitive and actuation limitations. Human frailty should be respected. Privacy rights of individuals engaging with AI and robots must be respected including ‘the deactivation of video monitors during intimate procedures’. Consent must be obtained from users of AI and robotics before any use or disclosure of personal data undertaken. Users cannot use AI and robots to contravene ethical or legal principles and standards. Users are not permitted to modify any robot to enable it to function as a weapon.

EU Commission

An early initiative of the EU Commission was to commission the RoboLaw Collaborative project, with the aim ‘to offer an in depth analysis of the ethical and legal issues raised by robotic applications and to provide the European and national regulators with guidelines to deal with them’ (RoboLaw, 2014). One focus was to ‘offers an analysis of the key concerns in the field of care robots’(RoboLaw, 2014: 168). It’s focus on the ethical and legal consequences considered both the governance of care robots and the social structure of care, and suggests ‘the introduction of robots into social systems will force us to review – in terms of respecting fundamental rights – the way we guarantee care and justice for people who live or work in conditions of vulnerability’ (RoboLaw, 2014: 171). The Report addressed those questions through the adaptation of the Capability approach (discussed elsewhere), and propose that Personal Care Robots (PCR) are a means of realising capabilities to reach a greater range of functioning, which is a ‘person's real freedom or opportunities’ (RoboLaw, 2014: 173). Furthermore, the Report expands on this reflection to review some critical concerns regarding safety, responsibility, autonomy, independence, enablement, privacy, social connectedness, distributive justice, and future scientific research, points that cannot be expanded upon here. On the overarching issue of regulation of robots per se, the authors cite the work of Leenes et al (2017) where four regulatory dilemmas are identified:

1. keeping up with rapid technological advances
2. striking a balance between stimulating innovation and protecting human fundamental rights and values
3. affirming existing social norms or nudging those norms in new and different directions
4. balancing between effectiveness and legitimacy in techno-regulation (COMEST, 2017, p.6).

The High-Level Expert Group on Artificial Intelligence has produced *Draft Ethics Guidelines for Trustworthy AI* (latest draft Dec 2018, document to be finalised in March 2019). This aims to be a pragmatic document that aims at implementation:

> In contrast to other documents dealing with ethical AI, the Guidelines hence do not aim to provide yet another list of core values and principles for AI, but rather offer guidance on the concrete implementation and operationalisation thereof into AI systems. Such guidance is provided in three layers of abstraction, from most abstract in Chapter I (fundamental rights, principles and values), to most concrete in Chapter III (assessment list).

Importantly: the Guidelines 'are not intended as a substitute to any form of policymaking or regulation'.

Council of Europe (COE) Initiatives

Council of Europe adopted its first European Ethical Charter on the use of artificial intelligence in judicial systems, and identified a number of core principles to be respected in the field of AI and justice (CoE, 3-4 December 2018). This includes the principles respect of fundamental rights, non-discrimination, the quality and security with regard to the processing of judicial decisions and data, the principles of transparency, impartiality and fairness, and finally, the principle "under user control".

In 2017 the Council of Europe issued the report: Algorithms and Human Rights: study on the human rights dimensions of automated data processing techniques and possible regulatory implications. In this report the Council recommended:

- Trans-disciplinary research that is evidence-based and problem-orientated should be undertaken on the human rights, ethical and legal implication of algorithmic decision-making
- Human rights impact assessments should be conducted before making use of algorithmic decision-making in all areas of public administration
- Public entities should be held accountable for the decisions they adopt based on algorithmic processes.
- Experimental regulatory approaches on how best to protect rights because of the use of algorithms should be considered.

- Public awareness campaigns should be developed to inform and engage the general public to that they can 'critically understand and deal with the logic and operation of algorithms'.
- Certification and auditing mechanisms for automated data ought to be developed to ensure compliance with human rights.
- States should not insist that internet intermediaries use automated technique to monitor data they transmit about users as this has 'a chilling effect on the freedom of expression'.
- Standards and guidelines on the challenges of decision-making through algorithms ought to be developed by sectoral regulators such as insurance, credit reference agencies, banks etc.).
- Additional institutions are needed to analyse and assess the impacts of algorithmic decision-making.
- The Council of Europe is the appropriate venue to develop 'standards-setting instruments for guidance to the member states'.

CoE Commissioner for Human Rights

Dunja Mijatović, CoE Commissioner for Human Rights offered her views on the human rights issues at stake in AI development and use (Mijatović, 2018). Separately, Mijatović stated, 'existing human rights framework must apply and the concerns and rights of everyone put at the centre of AI systems' design, deployment and implementation' (CommDH/Speech(2019)1, 26 – 27 February, 2019). In 2019 the Commissioner plans 'to publish a document on AI and human rights to help member states handle the multifaceted impact that AI can have on human rights' (CommDH/Speech(2019)1, 26 – 27 February, 2019). In her Report following her visit to Estonia from 11th to 14th June, 2018, the Commissioner identified strong focus on digitalisation, new technologies and artificial intelligence, and urged 'the authorities to support and empower older persons in the use of information and communications technology, so that they can exercise fully their right to participate in social and public life' (CommDH(2018)14, 28 September 2018). Furthermore, she highlighted the need for careful consideration given 'to the ethical, legal and human rights implications of using robots and artificial intelligence in the care of older persons' (CommDH(2018)14, 28 September 2018: 2). Focusing on the use of robots and artificial intelligence in care, the Commissioner, urged 'the authorities to conduct an investigation into errors related to the use of algorithms in decisions about social benefits', and recommended 'that the authorities consider developing specific human rights-based guidelines regarding the use of robot and artificial intelligence in long-term care' (CommDH(2018)14, 28 September 2018: para 96). The Commissioner also encouraged concerns for the rights of the elderly be included as Estonia moves forward with drafting an artificial intelligence strategy and legislation.

*The EU Fundamental Rights Agency (*FRA)

In 2018, FRA launched a research project on artificial intelligence, big data and fundamental rights. This project aims to assess the positive and negative fundamental rights implications

of new technologies, including AI and big data. The question of AI and human rights is addressed in the most recent Fundamental Rights Report (2019), however it did not address its application to the health care sector.

The FRA also provided a Report, '#BigData: Discrimination in data-supported decision making' (2018), and observed that using data and algorithms can contribute to discriminatory decision making, and proposed how we can move towards fundamental rights compliance in the development and use of algorithms. This included, transparency to give greater scrutiny about the way the algorithms were built; to conduct fundamental rights impact assessments to assist in identifying 'potential biases and abuses in the application of and output from algorithms'; to check the quality of data, and 'make quality assessments of the correctness and generalisability of the data'; and to develop meaningful procedures to explain the build and construction of algorithms to those who challenge data-supported decisions, and to facilitate access to remedies (FRA #BigData, 2018).

The Committee of Ministers

A declaration by the Committee of Ministers of the Council of Europe member States recently committed 'to building societies based on the values of democracy, human rights and the rule of law', while new focus is given to 'the ongoing process of societal transformation that is fuelled by technological advancements' (Decl(13/02/2019)1, 13 February 2019). The declaration emphasized the importance of protecting personal data, ensuring unbiased application of AI and machine learning technology, and protecting vulnerable people from subliminal manipulation through the exploitation of new technologies.

> Contemporary machine learning tools have the growing capacity not only to predict choices but also to influence emotions and thoughts and alter an anticipated course of action, sometimes subliminally. The dangers for democratic societies that emanate from the possibility to employ such capacity to manipulate and control not only economic choices but also social and political behaviours, have only recently become apparent. In this context, particular attention should be paid to the significant power that technological advancement confers to those – be they public entities or private actors – who may use such algorithmic tools without adequate democratic oversight or control (Decl(13/02/2019)1, 13 February 2019. Para 8).

Pursuant to the The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108),[7] the Consultative Committee of the Convention has developed new Guidelines On Artificial Intelligence And Data Protection.[8] The Guidelines provide a range of recommendations for AI developers, manufacturers and service providers, and also for legislators and policy makers, to provide a set of baseline

[7] The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108), opened for signature on 28 January 1981. This Council of Europe treaty was the first legally binding international instrument in the data protection field.

[8] Consultative Committee Of The Convention For The Protection Of Individuals With Regard To Automatic Processing Of Personal Data (Convention 108) *Guidelines On Artificial Intelligence And Data Protection.* T-Pd(2019)01. Strasbourg, 25 January 2019.

measurements. The key elements of this approach from the Convention for AI development relying on the processing of personal data are: lawfulness, fairness, purpose specification, proportionality of data processing, privacy-by-design and by default, responsibility and demonstration of compliance (accountability), transparency, data security and risk management.[9] Respect for fundamental rights, accountability, vigilance and means of redress are also included in the Guidelines of the Consultative Committee while 'developing and adopting AI applications that may have consequences on individuals and society'.

[9] Ibid. 1.

Part 5: Space of regulation the nation state

National strategies

There is now an economic imperative for national governments to seek to intervene in and drive development in AI and robotics. According to *Forbes* (Baron 2019) nineteen countries now have an 'AI strategy' (Canada being the first and the US the 19th). The governments of advanced economies, led by those such as France, China, the US, Germany and the UK, are now making very substantial investments of public money in AI research and development (Sloane 2018), boosting heavy investment by the private sector.

Nevertheless, some researchers have already suggested that individual governments seeking competitive advantage through a national strategy may need to consider alternative approaches, due to the risks inhering in these new technologies. Esposito et al (2018) observe that:

> data flows align with geographic boundaries only incidentally, not fundamentally. Geopolitically, nation-states are sovereign entities; but in the digital economy, they are sovereign in name only, not necessarily in practice.

As a consequence, national strategies may suffer from country-specific biases, and contribute to the fragmentation of knowledge and talent in resolving the issues AI technologies present. As Baron (2019) notes, 'AI simply can't be limited by geographic boundaries, and global cooperation will be crucial'. Notwithstanding that, we can outline some of the strategic direction within individual nation states.

Europe

Crider (2018) maps and analyses strategies and proposals for regulation on artificial intelligence (AI) in Europe. [check and use to enhance this section]

France

France has developed an ethics based approach (Raso et al 2018)

- How can humans keep the upper hand? Report on the ethical matters raised by algorithms and artificial intelligence. 26 December 2017.[10]
- C. Villani, Donner un sens à l'intelligence artificielle, Premier ministre, 2018[11]
- Renouveau de l'intelligence artificielle et de l'apprentissage automatique, Académie des technologies, 2018.[12]

[10] https://www.cnil.fr/en/how-can-humans-keep-upper-hand-report-ethical-matters-raised-algorithms-and-artificial-intelligence

[11] https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

[12] Commission technologies de l'information et de la communication. *Renouveau de l'Intelligence artificielle et de l'apprentissage automatique*. Rapport de l'Académie des technologies. Mars 2018. https://www.academie-

Germany

Federal Government of Germany - Data Ethics Commission. The DCS has recently suggested to add the following to the German National AI Strategy.

- "Upholding the ethical and legal principles based on our liberal democracy throughout the entire process of developing and applying artificial intelligence"
- "Promoting the ability of individuals and society as a whole to understand and reflect critically in the information society"

Germany Industry 4.0, European commission, 2017

UK

London has been positioned as 'the AI growth capital of Europe'

- Government Transformation Strategy 2017 to 2020
- Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights, has expressed concern about the uses of AI in the context of the establishment of a 'digital welfare state':
    - 'Not only will government services become 'digital by default,' as was first announced in 2012, but the inner workings of government itself will be transformed in a push for automation aided by data science and artificial intelligence. There are few places in government where these developments are more tangible than in the benefit system. We are witnessing the gradual disappearance of the postwar British welfare state behind a webpage and an algorithm. In its place, a digital welfare state is emerging. The impact on the human rights of the most vulnerable in the UK will be immense' (Alston 2018, p. 7)

The House of Lords Select Committee report (2018) titled *AI in the UK: ready, willing and able?* documents numerous concerns about the future of AI in Britain. Inter alia, these include: access to, and control of, data; technical transparency; explainability; addressing prejudice and data monopolies.

In particular, the report identified five overarching principles for an AI Code:

1. AI should be developed for the common good and benefit of humanity.
2. AI should operate on principles of intelligibility and fairness.
3. AI should not be used to diminish the data rights or privacy of individuals, families or communities.

technologies.fr/blog/categories/rapports/posts/renouveau-de-l-intelligence-artificielle-et-de-l-apprentissage-automatique

4. All citizens have the right to be educated to enable them to flourish mentally, emotionally and economically alongside AI

5. The autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence.

The UK government has also, in 2018, established the CDEI Centre for data ethics and innovation (CDEI). This body would be *advisory and not regulatory* of AI and attendant issues. It is to be organised along the lines of the UK Human Fertilisation and Embryology Authority.

The remit of the CDEI includes:

- Working with experts to develop and ethical framework (and publicising that work)
- Being a public advocate of the benefit of the technology
- Recommending changes to policy where appropriate
- The promotion of standards around the use of data
- Developing data trusts (trusted mechanisms to make it easier for organisations to understand how to use and share data for AI safely and securely)

(Evidence submitted by Matt Hancock, MP to HLSC report (2018), p.108).

More specifically, the CDEI seeks to contribute to emergent 'high level themes' including:

- issues around *targeting* data and artificial intelligence: ensuring that we receive services we want while ensuring the we are not manipulated in the process
- issues around *Fairness* in data generation and utilisation: ensuring that previous biases within data are not reproduced and to prevent the perpetuation of prejudice in future algorithm design
- issues around *Transparency* and the impact future AI technologies may have on human cognitive abilities: what are the circumstances in which non-human decisions ought to be taken and how and in what ways can these be explained
- new modes of *Liability* attaching to AI errors: how will these be assessed? There is a need for the collaboration of public and private sectors and individuals to seek new solutions to the complexity of unintended consequences and AI
- *data access* will require much more complex supply chain networks both nationally highlighted in the Hall-Presenti review of AI reportand internationally. This will require the creation of *data trusts* as per suggestions
- *intellectual property ownership:* AI will throw-up new and challenging issues around intellectual property rights. The requirement to develop new legal systems is prescient (Source:https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation#the-centres-role-and-objectives).

- See also NESTA 2018
- Royal Society report on machine learning

From UK Houses of Parliament (2018)

Legal and Regulatory Concerns

The Engineering and Physical Sciences Research Council funded UK-Robotics and Autonomous Systems Network12 has highlighted the need for international governance and regulation in this area

Legal and regulatory challenges include determining legal personality and legal liability for decisions made by robots.

The aforementioned European Parliament report suggested that autonomous robots could be granted 'electronic personalities' to enable them to be held liable for damages.

However, an open letter to the EC signed by 156 AI experts from 14 European countries warned that this would be 'inappropriate' from a legal and ethical perspective.

The diverse functions of robots may mean that robots are regulated differently. For example, robots that remind users to take medication may be classified as medical devices and regulated by the Medicines and Healthcare Products Regulatory Agency, while those processing personal data are regulated under GDPR. Clarifying ownership of data collected by robotics has been highlighted as an issue of concern. Data gathered from robots may be beneficial to roboticists in developing the technology, improving AI, and for machine learning, but in social care this may include personal or sensitive data.

North America

Canada

There is quite a battle amongst organisations to be the standards-setters for ethical AI (eg Canadian CIO Strategy Council and the OECD 2018)

The CIO Strategy Council is starting work to develop a new standard on the ethical design and use of automated decision systems (CIOSC 2018)

USA

Federal Government

Commencing in Nov 2016 the US Senate held hearings on AI [https://www.commerce.senate.gov/public/index.cfm/hearings?ID=042DC718-9250-44C0-9BFE-E0371AFAEBAB]

A bipartisan Artificial Intelligence Caucus was established (but seems to have disappeared?)

The 2016 National Artificial Intelligence Research and Development Strategic Plan (ref?) includes sections on ethics, standards and regulation.

Obama White House released report *Preparing for the future of artificial intelligence*

Government of California

[https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001]

Senate Bill No. 1001. An 'AI bot' is not allowed to communicate or interact with another human person in California online with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election.

- Other US states and Federal level protections?
- CCPA Privacy in California?

East Asia

Japan

- Society 5.0 [find out more!]

Future Advocacy (Fenech et al 2018) has detailed how the use of robotics may provide a resources to assist older or vulnerable patients in Japan. The Japanese strategy has 'the explicit aim that four in five care recipients will accept having some support provided by robots by 2020'. It has also been proposed that 'robots can take over physical tasks such as lifting patients from their beds' (Szondy 2015)

Japan's 'new robot strategy' (Headquarters for Japan's Economic Revitalization 2015) suggests new measures 'to promote introduction of robots with sensor technologies and artificial intelligence' motivated by 'looking after the aged and preventing them from falling prey to a serious disease such as dementia'.

Furthermore the Strategy (p. 63) proposes,

> to mitigate the workload of health care workers, efforts will be made to produce robotic wheelchairs that, making the best use of sensor and network technologies, allow the aged to move around indoors and outdoors independently and safely.

China

China has become a major player in the field of AI and robotics, emerging as a major global rival to leading developers such as the US and the EU. In July 2017 China released its *New Generation AI Development Plan*. This detailed its approach to becoming world leader in the field by 2030 (State Council of The People's Republic of China 2017).

The European Commission's European Political Strategy Centre (2018) has observed that because of cultural differences between China and Europe, Chinese consumers afford a greater level of trust in sharing personal data with producers of products, and therefore 'Chinese firms can implement more advanced AI technologies and work with larger volumes of granular data'. The Centre (2018, p. 4) argues that Europe 'also lags behind the US and

China on patent submissions and investments'. There are strong indications that China is deeply committed to becoming a global leader in the development of new technologies, including in the social applications of robotic technology.

China's 'Guidelines on Artificial Intelligence Development', has included a focus on 'Intelligent Health and Elder Care Systems' (State Council of The People's Republic of China 2017). This will include strengthening 'community intelligent health management', with an effort to research and develop 'health management wearable equipment and home intelligent health testing and monitoring equipment'. The strategy will also, construct 'intelligent elder care communities and institutions', and 'build a safe and convenient intelligent pension infrastructure system'. Furthermore, it proposes to strengthen, 'the intelligentization of products for elderly persons and intelligent products suitable for the aged'. The strategy includes developing 'audio-visual aid equipment, physical auxiliary equipment, and other intelligent home care equipment' for the aged. Finally, the strategy will encourage the development of 'mobile social and service platform for the elderly and emotional escort assistant to enhance the quality of life of the elderly' (State Council of The People's Republic of China 2017).

In more concrete terms, China has also released a 'Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018–2020)', which details the steps to be taken by various facets of Chinese institutions, and the Chinese economy (MIIT 2017).[13] In relation to social robotics, the action plan has identified achieving 'the large-scale application of intelligent service robots' with a broad range of applications, including, 'enhance the intelligence level of household service robots in applications including cleaning, elder care, rehabilitation, disability, and children's education' (MIIT 2017, Action Goals 3). Furthermore, the Chinese Action Plan has the ambitious target to achieve the following:

> By 2020, make breakthroughs in key technologies [...including] intelligent home service robots and intelligent public service robots should achieve mass production and application in medical rehabilitation, assistance to elderly and the disabled, and fire and disaster relief, perfect technological and functional prototype production, and achieve demonstrations for over twenty applications (MIIT 2017, Action Goals 3).

'Social credit score' - where 'a numerical index of an individual's "trustworthiness" based on a vast array of data points, including social media data, arrest and infraction records, volunteer activity, city and neighborhood records, and more. Those with high social credit scores enjoy benefits such as lower utility rates and more favorable borrowing conditions, while those with unfavorable scores might be unable to purchase airline or high speed rail tickets. National roll-out in 2020?

Singapore

- Has established (2018) an Advisory Council on the Ethical Use of AI and Data with an associated 5-year research programme

[13] See also, MIT Technology Review, '*China has a new three year plan to rule AI*', 15 December 2017.

- It will 'will advise and work with the Infocomm Media Development Authority (IMDA) on the responsible development and deployment of AI. Amongst other things, the Advisory Council will assist IMDA in engaging stakeholders on issues that support the development of AI governance capabilities and frameworks. These include engaging ethics boards of commercial enterprises on ethical and related issues arising from private sector use of AI and data; consumer representatives on consumer expectations and acceptance of the use of AI; as well as members of the private capital community on the need to incorporate ethical considerations in their investment decisions into businesses which develop or adopt AI. The Advisory Council will also assist the Government in developing ethics standards and reference governance frameworks, and publish advisory guidelines, practical guides, and codes of practice for the voluntary adoption by the industry. https://www.imda.gov.sg/about/newsroom/media-releases/2018/composition-of-the-advisory-council-on-the-ethical-use-of-ai-and-data

- Has also developed a 'fact sheet' on the issues: https://www.imda.gov.sg/-/media/imda/files/about/media-releases/2018/2018-06-05-fact-sheet-for-ai-govt.pdf

- FEAT AI and Finance (Nov 2018) –find out more

Other

India

- ethics based approach (Raso et al 2018)

'Developing' and 'low income' countries

https://www.consumersinternational.org/media/154869/ci_connecting-voices_2017_v2.pdf

This document is good for looking at digital rights in non-western societies, including in the Middle East and Africa

See also Buston et al (2018)

Government agencies more broadly

Probably hard to get info on this: this would be the like of health agencies, procurement bodies &c

Part 6: Transnational NGOs and corporations

Regulation of AI and robotics will inevitably, given the influence of multinational corporations and the ubiquity of free trade, be necessary at the transnational and global level.

Standards setting bodies

As Winfield (2019, p. 46) notes: 'standards are a vital part of the infrastructure of the modern world: invisible, but no less important than roads, airports and telephone networks'. Standards can act to reduce harm (and increase safety); seek to create interoperability (phone services; electrical devices); and enhance the quality of devices and services. Winfield notes that standards often express or reflect ethical stances and principles: 'without ethical standards, it is hard to see how robots and AIs will be trusted and widely accepted, and without that acceptance their great promise will not be realized (p. 48).

Organisations that set regulatory standards for the design of social and care robots include the British Standards Institution (BSI) and the International Organization for Standardization (ISO).

Key regulatory standards for robotics in social care include:

- ISO 8373, which provides an overview of robotics terms and vocabularies, notably defining and distinguishing between types of service robots and industrial robots
- ISO 13482, which focuses on minimising the potential risks posed by robots that come into direct contact with people
- BS 8611, which addresses ethical hazards relating to the use of robots

While in essence voluntary, standards can become effectively mandatory through licensing requirements. Demonstrated adherence to standards can confer a competitive advantage in the marketplace, professional bodies can require adherence to standards, and public bodies can require adherence to standards through procurement and funding mechanisms.

Winfield (2019, p. 46) identifies common principles in emergent robotics standards. These are that robots and AIs should:

- do no harm
- be free of bias and deception
- respect human rights and freedoms, including dignity and privacy
- promote well-being
- be transparent and dependable while ensuring that the locus of responsibility and accountability remains with their human designers or operators.

The first 'explicit' ethical standard related to robots was BS 8611.

> BS 8611 is not a code of practice, but instead guidance on how designers can undertake an ethical risk assessment of their robot or system, and mitigate any ethical risks so identified. At its heart is a set of 20 distinct ethical hazards and risks, grouped under four categories: societal, application, commercial and financial, and environmental (Winfield 2019, p. 46)

The societal hazards include … loss of trust, deception, infringements of privacy and confidentiality, addiction, and loss of employment.

International Standards Organisation ISO - JTC 1- SC42

Certification (but of what exactly?) will be a crucial issue

IEEE

Apart from the global technology companies, probably the most influential player in the regulation of AI is the IEEE (Institute of Electrical and Electronics Engineers).

The IEEE has proposed *Ethically Aligned Design* (launched as a major initiative in March 2019).

https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

This initiative places social (rather than commercial) wellbeing at the centre of its ethical stance: its aim is to:

> ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity.

The ethical design, development and implementation of autonomous technologies is to be guided by a set of General Principles:

- human rights: Ensure they do not infringe on internationally recognized human rights
- well-being: Prioritise metrics of well-being in their design and use
- accountability: Ensure that their designers and operators are responsible and accountable.

The IEEE's is a highly collaborative process, for example it established the IEEE P7000 series of standards in April 2016. To date this is the largest suite of standards projects focused on AI ethics. There are now 14 approved standardisation working groups open for anyone to join, with over 1000 volunteers involved (Winfield 2019, p. 47).[update?]

The working groups focus on the intersection of technology, interoperability and applied ethical considerations. They address the full range of autonomous technologies, from diagnostic AI to driverless cars and drones. According to Winfield (2019, p. 47) 'The significance of this initiative cannot be overstated; coming from a professional body with the standing and reach of the IEEE Standards Association it marks a watershed in the emergence

of ethical standards'. Winfield further describes the process of development of explicitly ethically-based standards by this professional body as 'ambitious' and 'unprecedented'.

Ethically Aligned Design [EAD] covers: 'general (ethical) principles; how to embed values into autonomous intelligent systems; methods to guide ethical design; safety and beneficence of artificial general intelligence and artificial superintelligence; personal data and individual access control; reframing autonomous weapons systems; economics and humanitarian issues; law; affective computing; classical ethics in AI; policy; mixed-reality; and well-being.' (Winfield 2019, p. 47)

The *EAD* standards reflect the complexity of regulation. They operate across numerous autonomous technologies, identify a broad range of principles that need to be addressed, and then identify different types and levels of regulation for different stakeholder groups, from accident investigators to the general public.

Business (IEEE 2019b)

Other transnational NGOs

Other industry/professional bodies such as Software and Information Industry Association (SIIS), Information Technology Industry Council (ITIC) and International Telecommunication Union (ITU)have developed and continue to develop positions on application of AI.

Recently 14 national and international standards organisations (including the IEEE) have come together in a global forum Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS). Through collaboration, members of OCEANIS will share information and coordinate on initiatives to enhance understanding around the critical role that standards can play in facilitating innovation and addressing the ethical use of big data and artificial intelligence.

During the RightsCon human rights convention in Toronto 2018, a number of international human rights NGOs launched the 'Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems'. The Declaration was prepared by Amnesty International and Access Now, and was also endorsed by Human Rights Watch. The Declaration emphasised using human rights law, and key human rights themes, including the right to equality and non-discrimination, and to promote inclusion, diversity, and equity as 'key components to ensuring that machine learning systems do not create or perpetuate discrimination, particularly against marginalised groups' [footnote].

Civil society/activist/academic organisations

- AI Now
- Amnesty International
- Human Rights Watch
- Civil liberties groups
- Future Advocacy [thinktank]

- Welfare coalitions
- Consumer organisations (eg Consumers International)
- International Committee of the Red Cross (ICRC)
    - A special competency and mandate of the ICRC to examine the use of new technologies and modern weapon systems, as it has responsibility to offer oversight on the Geneva Conventions and its protocols. The ICRC can offer valuable sets of principles and guidance of the military and civilian use of new technologies, including Artificial technologies.

Montreal Declaration for Responsible AI

(An initiative of Université de Montréal)

https://www.montrealdeclaration-responsibleai.com/the-declaration

The Montreal Declaration for responsible AI development has three main objectives:

Develop an ethical framework for the development and deployment of AI

1. Guide the digital transition so everyone benefits from this technological revolution
2. Open a national and international forum for discussion to collectively achieve equitable, inclusive, and ecologically sustainable AI development

Commercial bodies

Google, Facebook, Amazon, Apple, Microsoft &c have developed their own ethical frameworks in the light of potential external regulation and political pressure. They also have access to vast reservoirs of internal expertise.

Google/Alphabet

Google's principles

1. Be socially beneficial (benefit/risk analysis) - but also reflect local 'cultural, social and legal norms', which presumably allows for censorship in China &c
2. Avoid creating or reinforcing unfair bias (but allow 'fair' bias?)
3. Be built and tested for safety
4. Be accountable to people (subject to *appropriate* human direction/control)
5. Incorporate privacy design principles (including notice and consent)
6. Uphold high standards of scientific excellence

Also Google will NOT deploy or develop AI that can:

1. cause or is likely to cause overall harm

2. develop weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.
3. gather or use information for surveillance violating internationally accepted norms
4. contravenes widely accepted principles of international law and human rights

Role of tech employees

> Google had to end its collaboration with the Department of Defense last year on Project Maven after thousands of its own employees signed a petition to end the use of their work by the military. However, tech giants like Amazon and Microsoft have pledged to continue to work with the government, and, specifically, the Department of Defense as they see fit (Baron 2019)

Facebook

using AI for monitoring – eg 'nude' or sexual images: 96% automatically identified and deleted (31m images in Q3/2018)

They have been less successful in identifying text-based processes such as 'hate speech' and online bullying

Interesting aspect: human reviewers are being trained to adopt AI-like systems of discretion.

[cf *WIRED* article on this]

[JP to expand on this highlighting, for example, CORU standards and proficiencies of care and other health and social care regulations throughout the EU and beyond]

References/links

Materials that are particularly good sources for students are indicated with the symbol ☝

Abbott, R. & B. Bogenshneider (2018) 'Should robots pay taxes? Tax policy in the age of automation'. *Harvard Law and Policy Review* 12. pp. 145-175. http://harvardlpr.com/wp-content/uploads/2018/03/AbbottBogenschneider.pdf

AI Now Institute (2018) AI in 2018: A year in review: ethics, organizing and accountability. https://medium.com/@AINowInstitute/ai-in-2018-a-year-in-review-8b161ead2b4e

Alston, P. (2018) 'Statement on visit to the United Kingdom, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights', pp. 7-12. https://www.ohchr.org/Documents/Issues/Poverty/EOM_GB_16Nov2018.pdf

Amos, M. & R. Page (eds) (2014). *Beta life: Stories from an A-life future*. Manchester: Comma Press.

Apolitical (2018) 'New York writes new rules to rein in government by algorithm'. https://apolitical.co/solution_article/new-york-writes-new-rules-to-rein-in-government-by-algorithm

António Guterres (12 December 2016) 'Secretary-General-designate António Guterres' remarks to the General Assembly on taking the oath of office'. United Nations. https://www.un.org/sg/en/content/sg/speeches/2016-12-12/secretary-general-designate-ant%C3%B3nio-guterres-oath-office-speech.

António Guterres (2018)UN Secretary-General's Strategy On New Technologies https://www.un.org/en/newtechnologies/images/pdf/SGs-Strategy-on-New-Technologies.pdf

António Guterres (2019) Report of the UN Secretary-General's High-level Panel on Digital Cooperation, 'The Age of Digital Interdependence' https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf

Baron, J. (2019) 'Will Trump's new artificial intelligence initiative make the U.S. the world leader in AI?' *Forbes* 11 February. https://www.forbes.com/sites/jessicabaron/2019/02/11/will-trumps-new-artificial-intelligence-initiative-make-the-u-s-the-world-leader-in-ai/

Bedoya, A. (2018) 'Why Silicon Valley lobbyists love big, broad privacy bills'. *New York Times* 11 April.

Beer, D. (2017) 'The social power of algorithms'. *Information, communication and society*. 20(1). https://doi.org/10.1080/1369118X.2016.1216147

Boden, M. (2018) *Artificial intelligence: A very short introduction*. Oxford University Press.

Broussard, M. (2018) *Artificial unintelligence: How computers misunderstand the world*. Cambridge [MA]: MIT Press.

Buston, O., M. Fenech & N. Strukelj (2018) *Technology forward scan: Future applications for digital technology in low and middle income countries.* Background Paper Series 14. Oxford: Pathways for Prosperity Commission http://futureadvocacy.com/wp-content/uploads/2019/01/nic-futureadvocacy-28nov18_final_to_go_on_website.pdf

Calo, C., N. Hunt-Bull, L. Lewis & T. Metzler (2011) 'Ethical implications of using the Paro robot with a focus on dementia patient care'. *Human-robot interaction in elder care: Papers from the 2011 AAAI workshop*. https://www.aaai.org/ocs/index.php/WS/AAAIW11/paper/view/3808

Carter, S. (2018) '90 years of AI in the movies: what's changed (and what hasn't)'. *Enlightened Digital*. August 9. https://enlightened-digital.com/90-years-of-ai-in-the-movies-whats-changed-and-what-hasnt/

Cave, S. & K. Dihal (2018) 'Ancient dreams of intelligent machines: 3,000 years of robots'. *Nature*. https://www.nature.com/articles/d41586-018-05773-y

Cave, S. & K. Dihal (2019) 'Perspective: Hopes and fears for intelligent machines in fiction and reality'. *Nature Machine Intelligence* Vol. 1 https://doi.org/10.1038/s42256-019-0020-9

Centre for Data Ethics and Innovation (CDEI) https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei

Chesney, R. & D. Citron 'Deep fakes: A looming crisis for national security, democracy and privacy?' *Lawfare.* https://www.lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy

Chouldehova, A. et al (2018) 'A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions'. *Proceedings of Machine Learning Research* 81. http://proceedings.mlr.press/v81/chouldechova18a/chouldechova18a.pdf

Christian, B. & T. Griffiths (2016) *Algorithms to live by: The computer science of human decision*s. New York: Holt.

Chu, M., M. Harryson, J. Manyika, R. Roberts, R. Chung, A. van Heteren & P. Nel (2018) *Notes from the AI frontier: Applying AI for social good*. Discussion Paper. McKinsey. https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good

Chubb, J., J. Montana, J. Stilgoe, A. Stirling & J. Wilsdon (2018) *A review of recent evidence on the governance of emerging science and technology*. Wellcome. https://wellcome.ac.uk/sites/default/files/evidence-review-governance-emerging-science-and-technology_0.pdf

CIOSC [CIO Strategy Council] (2018) [https://ciostrategycouncil.com/2018/08/cio-strategy-council-an-emerging-international-leader-on-developing-standards-for-ethical-use-of-ai-and-big-data/]

Coeckelbergh, M. (2009) 'Health care, capabilities and AI assistive technology'. *Ethical Theory and Moral Practice*, Vol. 13, Issue 2, pp.181-190.

Committee of the Rights of Persons with Disabilities (25 September, 2017) *Concluding observations on the initial report of Morocco*, UN Doc. CRPD/C/MAR/CO/1.

Committee of the Rights of Persons with Disabilities (10 Oct 2017) *Concluding observations on the initial report of Luxembourg*, UN Doc CRPD/C/LUX/CO/1.

Cole, S. (2018) 'We are truly fucked: Everyone is making AI-generated fake porn now'. *Vice Motherboard*. https://motherboard.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley

Consumers International (2018) 'AI for consumers: five things we learnt at the Euroconsumers event on artificial intelligence' https://www.consumersinternational.org/news-resources/blog/posts/ai-for-consumers-blog/

Convention on the Rights of Persons with Disabilities (New York, 13 December 2006) 2515 U.N.T.S. 3 (2006) entered into force 3 May 2008.

Council of Europe (2017) Technological convergence, artificial intelligence and human rights. Recommendation n°2102. Text adopted by the Assembly on 28 April (18th Sitting). https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en

Council of Europe (2019) Sub-Committee on Artificial Intelligence and Human Rights. http://www.assembly.coe.int/nw/xml/AssemblyList/AL-XML2HTML-en.asp?XmlID=SubCommittee-B2

Council of Europe (13 February 2019) *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes.* Adopted by the Committee of Ministers at the 1337th meeting of the Ministers' Deputies. Decl(13/02/2019)1.

Crider, C. (2018) 'Mapping artificial intelligence strategies in Europe: a new report by Access Now', *Access Now*. https://www.accessnow.org/cms/assets/uploads/2018/11/mapping_regulatory_proposals_for_AI_in_EU.pdf

Crowe, S. (2019) 'Anki, consumer robotics maker, shuts down'. *Robot Report* 29 April. https://www.therobotreport.com/anki-consumer-robotics-maker-shuts-down/

Danaher, J. (2017) 'The symbolic-consequences argument in the sex robot debate' In J. Danaher & N. McArthur (eds) *Robot sex: Social and ethical implications.* Cambridge [MA]: MIT Press. pp. 103-131.

Danaher, J., B. Earp & A. Sandberg (2017) 'Should we campaign against sex robots?' In J. Danaher & N. McArthur (eds) *Robot sex: Social and ethical implications.* Cambridge [MA]: MIT Press. pp. 47-72.

Devlin, K. (2018) *Turned on: Science, sex and robots*. London: Bloomsbury.

Dickinson, H., Smith, C., Carey, N. and Carey, G. (2018) *Robots and the delivery of care services: What is the role for government in stewarding disruptive innovation?* Melbourne:

ANZSOG. https://www.anzsog.edu.au/resource-library/research/robots-and-the-delivery-of-care-services ☝

Duffy, M. (2018) 'Building sustainable jobs and supporting human potential in the care sector'. In E. Paus (ed) *Confronting dystopia: The new technological revolution and the future of work*. Cornell University Press, pp. 94-111.

EIII (2019a) Ethically aligned design.

EIII (2019b) A call to action for businesses using AI. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead/ead-for-business.pdf

EPIC [Electronic Privacy Information Center] (2020) Algorithms in the criminal justice system: Pre-trial risk assessment tools. https://epic.org/algorithmic-transparency/crim-justice/ ☝

Esposito, M., T. Tse & J. Entsminger (2018) 'The case against national AI strategies'. *Project Syndicate*, Oct 17. https://www.project-syndicate.org/commentary/case-against-national-ai-strategies-by-mark-esposito-et-al-2018-10

Eubanks V. (2018) *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St Martin's Press.

European Commission (2014) '*Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics (RoboLaw)*' Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013).

European Commission (2018) *Artificial intelligence for Europe*. Brussels: Communication from the Commission to the European Parliament, the European Council, The Council, The European Economic and Social Committee and The Committee of the Regions. COM(2018) 237 final. https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

European Commission (2018b) [High level expert group on artificial intelligence] https://www.euractiv.com/wp-content/uploads/sites/2/2018/12/AIHLEGDraftAIEthicsGuidelinespdf.pdf

European Commission (2018b) *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment.* European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe, Adopted at the 31st plenary meeting of the CEPEJ. Strasbourg.

European Parliament (2017) Resolution and recommendations to the Commission on Civil Law Rules on Robotics, including study of 'Ethical Aspects of Cyber-Physical Systems' (16 February) http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN

European Political Strategy Centre (2018) 'The age of artificial intelligence. Towards a European strategy for human-centric machines'. *EPSC Strategic Notes* 29. https://ec.europa.eu/epsc/sites/epsc/files/epsc_strategicnote_ai.pdf

Fast, E. & E. Horvitz (2016) 'Long-term trends in the public perception of artificial intelligence'. https://arxiv.org/abs/1609.04904

Fenech, M., N. Strukelj & O. Buston (2018) 'Ethical, social, and political challenges of artificial intelligence in health'. Wellcome Trust /Future Advocacy. https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf

Forbes (2019) The world's most valuable brands. https://www.forbes.com/powerful-brands/list/

Google's AI principles. https://www.blog.google/technology/ai/ai-principles/

Guttman, C. (2018a) Do you know what artificial intelligence really is? *Pulse* (LinkedIn) https://www.linkedin.com/pulse/learn-artificial-intelligence-from-credible-sources-guttmann/

Guttman, C. (2018b) 'An overview of artificial intelligence ethics and regulations. *Pulse* (LinkedIn) https://www.linkedin.com/pulse/overview-artificial-intelligence-ethics-regulations-guttmann/

Hasse, C., S. Trentmøller & J. Sorenson (2019) 'Robot definitions'. REELER project. http://reeler.eu/reeler-library/robot-definitions/

Headquarters for Japan's Economic Revitalization (2015) 'New robot strategy'. Ministry of Economy, Trade & Industry. www.meti.go.jp/english/press/2015/pdf/0123_01b.pdf

Houses of Parliament (2018) 'Robotics in social care'. *Postnote* 591, December. London: Houses of Parliament, Parliamentary Office of Science and Technology. https://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0591

Human Rights Council (5 May, 2017), *Promotion, protection and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective*, Report of the United Nations High Commissioner for Human Rights, Thirty-fifth session of the Human Rights Council, A/HRC/35/9.

Human Rights Council (21 July 2017) *Robots and rights: the impact of automation on the human rights of older persons*, Report of the Independent Expert on the enjoyment of all human rights by older persons, Rosa Kornfeld-Matte Thirty-sixth session of the Human Rights Council, A/HRC/36/48.

Human Rights Council (6 April 2018) *The regulation of user-generated online content.* Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, submitted to the thirty-eighth session of the Human Rights Council UN Doc A/HRC/38/35.

Human Rights Council (3 August, 2018) *The right to privacy in the digital age*, Report of the United Nations High Commissioner for Human Rights,Thirty-ninth session of the Human Rights Council, A/HRC/39/29.

Inno3Med (2018) 'Phoque émotionnel interactif PARO' [PARO interactive seal][dossier] http://www.phoque-paro.fr/wp-content/uploads/2018/10/Dossier-PARO-021018.pdf

ITU (International Telecommunication Union) (2017) '*AI for Good*' Global Summit in Geneva, Switzerland.

ITU (International Telecommunication Union) (2018) *United Nations activities on Artificial Intelligence (AI)*. Geneva: ITU.

Judge Business School (2017) Workshop: Risks and benefits of artificial intelligence. https://www.jbs.cam.ac.uk/faculty-research/centres/risk/news-events/events/2017/risks-and-benefits-of-artificial-intelligence-and-robotics/

Kamishima, Y., B. Gremmen & H. Akizawa, H. (2018) 'Can merging a capability approach with effectual processes help us define a permissible action range for AI robotics entrepreneurship?'. *Philosophy and Management* 17(1), pp. 97-113.

Kemp, L., Cihon, P., Maas, M., Belfield, H., Ó hÉigeartaigh, S., Leung, J., Cremer, Z., (26 February 2019) *UN High-level Panel on Digital Cooperation: A Proposal for International AI Governance*. Cambridge University's Centre for the Study of Existential Risk and Oxford University's Center for the Governance of AI at the Future of Humanity Institute.

Kiggins, R. (ed)(2018) *The political economy of robots*. London: Palgrave Macmillan.

Kleijssen, Jan (26-27 February 2019) "Governing the Game Changer - Impacts of artificial intelligence development on Human Rights, Democracy and the Rule of Law". Helsinki, Finland.

Lee, J. (2017) *Sex robots: The future of desire*. London: Palgrave Macmillan.

Levy, D. (2008) *Love and sex with robots: The evolution of human/robot relationships*. London: Duckworth Overlook.

Liu, L. et al (2018) 'Delayed impact of fair machine learning'. *Proceedings of Machine Learning* Research 80 http://proceedings.mlr.press/v80/liu18c/liu18c.pdf

Livingstone, S. & A. Third (2017) 'Children and young people's rights in the digital age: an emerging agenda'. *New Media & Society*. http://eprints.lse.ac.uk/68759/

Mamun, S.M. (2018) 'How to get the public excited about the rise of the robots'. *Apolitical*. https://apolitical.co/solution_article/how-to-get-the-public-excited-about-the-rise-of-the-robots/

Marcus, G. (2018) Deep learning: A critical appraisal. https://arxiv.org/pdf/1801.00631.pdf

Markoff, J. (2015) *Machines of loving grace: The quest for common ground between humans and robots*. New York: Ecco.

Matthews, D. (2019) 'Big tech funding AI ethics research to "delay and avoid" regulation'. *Times Higher Education* 18 April. https://www.timeshighereducation.com/news/big-tech-funding-ai-ethics-research-delay-and-avoid-regulation

Mayfield Robotics (2018) 'An important (and difficult) announcement.' https://www.heykuri.com/blog/important_difficult_announcement/

Meekins, A. (2019) 'Jibo's goodbye: social robots hard to build.' *The Robot Report* 11 March. https://www.therobotreport.com/jibo-social-robots-hard-build/

MIIT (Ministry of Industry and Information Technology of The People's Republic of China)(2017) '*Three-year action plan for promoting development of a new generation artificial intelligence industry (2018–2020*)'. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-government-outlines-ai-ambitions-through-2020/

Mijatović, Dunja (3rd July, 2018) *Safeguarding Human Rights In The Era Of Artificial Intelligence*. Human Rights Comment. Strasbourg.

Mijatović, Dunja (26 – 27 February, 2019) *Speech at the High level conference "Governing the Game Changer - Impacts of artificial intelligence development on human rights, democracy and the rule of law"*, Conference co-organised by the Finnish Presidency of the Council of Europe Committee of Ministers and the Council of Europe. CommDH/Speech(2019)1 Helsinki, Finland.

Mijatović, Dunja (28 September 2018) *Report Of The Commissioner For Human Rights Of The Council Of Europe, Following Her Visit To Estonia From 11 To 15 June 2018*. CommDH(2018)14. Strasbourg.

Mols, B. & N. Vergunst (2018) *Hallo robot: Meet your new friend and workmate*. Kingston-upon-Thames: Canbury.

Mols, B. (2019) Moderator, panel discussion on 'The promise and perils of social robots' at Etmaal van de Communicatiewetenschap. Radboud University, Netherlands. http://benniemols.blogspot.com/2019/02/the-promises-and-perils-of-social-robots.html

Muižnieks, Nils (18 January, 2018) 'The right of older persons to dignity and autonomy in care', Strasbourg. https://www.coe.int/en/web/commissioner/-/the-right-of-older-persons-to-dignity-and-autonomy-in-care

Mulgan, G, (2017) 'Anticipatory regulation: 10 ways governments can better keep up with fast-changing industries'. https://www.nesta.org.uk/blog/anticipatory-regulation-10-ways-governments-can-better-keep-up-with-fast-changing- industries/

NESTA (2018) '10 principles for public sector use of algorithmic decision making' https://www.nesta.org.uk/blog/10-principles-for-public-sector-use-of-algorithmic-decision-making/

Neven, L. & C. Leeson (2015) 'Beyond determinism: understanding actual use of social robots by older people'. In D. Prendergast & C. Garattini (eds) *Ageing and the digital life course.* New York/Oxford: Berghahn. pp. 84-102.

Noble, S.U. (2018) *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press.

Nussbaum, M. (2000) *Women and human development: The capabilities approach.* Cambridge: Cambridge University Press.

Nussbaum, M. (2011) *Creating capabilities.* Cambridge: Cambridge University Press.

Obama White House (date?) *Preparing for the future of artificial intelligence* https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

OCEANIS (Open Community for Ethics in Autonomous and Intelligent Systems) https://ethicsstandards.org

O'Dwyer, C. & K. Cormican (2017) 'Regulation – do or die: an analysis of factors critical to new product development in a regulatory context'. *Journal of Technology Management Innovation* 12(1). http://dx.doi.org/10.4067/S0718-27242017000100004

OECD (2018) 'OECD creates expert group to foster trust in artificial intelligence'. http://www.oecd.org/going-digital/ai/oecd-creates-expert-group-to-foster-trust-in-artificial-intelligence.htm

Office of the High Commissioner of Human Rights (2016) 'A Human Rights-Based Approach to Data: Leaving No One Behind in the 2030 Development Agenda'. Geneva, Switzerland. https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf

O'Neil C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.

Raso, F. et al (2018) *Artificial intelligence and human rights*. Cambridge [MA] Berkman Klein Center for Internet and Society at Harvard University https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3259344

Reisman, D. J. Schultz, K. Crawford & M. Whittaker (2018) *Algorithmic impact assessments: A practical framework for public agency accountability*. AI Now Institute. https://ainowinstitute.org/aiareport2018.pdf

Royal Society (2017) *Machine learning: the power and promise of computers that learn by example*. https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf

Russell, S. & P. Norvig (1995) *Artificial intelligence: A modern approach.* Englewood Cliffs [NJ]: Prentice Hall.

Saborowski, M. & I. Kollak (2015) '"How do you care for technology?" Care professionals' experiences with assistive technology in the care of the elderly'. *Technological Forecasting and Social Change* 93, pp. 133-140. http://dx.doi.org/10.1016/j.techfore.2014.05.006

Sen, A. (1984) 'Rights and capabilities.' In *Resources, values and development*. Cambridge [MA]: Harvard University Press, pp.307-324.

Sloane, M. (2018) 'Making artificial intelligence socially just: why the current focus on ethics is not enough' [blog post] LSE British Politics and Policy. https://blogs.lse.ac.uk/politicsandpolicy/artificial-intelligence-and-society-ethics/

Smith, G. (2018) *The AI delusion*. Oxford: Oxford University Press.

Sone, Y. (2017) *Japanese robot culture: Performance, imagination and modernity*. London: Palgrave Macmillan.

State Council of The People's Republic of China (2017) *Guidelines on artificial intelligence development*. https://chinacopyrightandmedia.wordpress.com/2017/07/20/a-next-generation-artificial-intelligence-development-plan/

Strickwerda, L. (2017) 'Legal and moral implications of child sex robots'. In J. Danaher & N. McArthur (eds) *Robot sex: Social and ethical implications.* Cambridge [MA]: MIT Press. pp. 133-152.

Szondy, D. (2015) 'Robear robot care bear designed to serve Japan's aging population', *New Atlas*. https://newatlas.com/robear-riken/36219/

Thompson, N. & I. Bremmer (2018) 'The AI cold war that threatens us all'. *Wired* October 23, https://www.wired.com/story/ai-cold-war-china-could-doom-us-all/

Turkle, S. (2011) *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

Turner, J. (2019) *Robot rules: Regulating artificial intelligence*. London: Palgrave Macmillan.

UN Centre for Artificial Intelligence and Robotics [UNICRI] http://www.unicri.it/in_focus/on/UNICRI_Centre_Artificial_Robotics

UN General Assembly (29 August 2018) *the implications of artificial intelligence technologies for human rights in the information environment, focusing in particular on rights to freedom of opinion and expression, privacy and non-discrimination* Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, submitted to the 73rd session of the General Assembly, UN Doc. A/73/348.

UN System Chief Executives Board for Coordination (May 2018) *High-Level Committee on Programmes. Report of 35th Session* (April 2018, New York). Available at: https://www.unsceb.org/content/action-artificial-intelligence-0

UNESCO (2017) Report of Comest on Robotics Ethics. Paris, 14 September 2017 (SHS/YES/COMEST-10/17/2 REV) 4.

Vincent, J., S. Taipale, B. Sapio, G. Lugano & L. Fortunati (eds)(2015) *Social robots from a human perspective*. Cham: Springer.

Wachter, S. et al (2017) 'Transparent, explainable, and accountable AI for robotics'. *Science Robotics*. http://robotics.sciencemag.org/content/2/6/eaan6080.full

Walsh, T. (2016) 'Turing's red flag: A proposal for a law to prevent artificial intelligence systems from being mistaken for humans.' *Comm. ACM* 59(7). doi: 10.1145/2838729 http://www.cse.unsw.edu.au/~tw/turingredflag.pdf

Willcocks, L. & M. Lacity (2016) *Service automation: Robots and the future of work*. Stratford-upon-Avon: Steve Brookes Publishing.

Winfield, A. (2019) 'Ethical standards in robotics and AI'. *Nature Electronics* 2. pp. 46-48. https://www.nature.com/articles/s41928-019-0213-6.epdf

Winkle, K., P. Caleb-Solly, A. Turton & P. Bremner (2019) 'Mutual shaping in the design of socially assistive robots: A case study on social robots for therapy'. *International Journal of Social Robotics* https://doi.org/10.1007/s12369-019-00536-9

UNESCO (2017) '*Report of COMEST on Robotics Ethics*', The World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) (SHS/YES/COMEST-10/17/2 REV, 2017)

Bits

The Universal Guidelines for Artificial Intelligence, grounded in a human rights framework, set forth twelve principles that are intended to guide the design, development, and deployment of AI, and frameworks for policy and legislation. Broadly, the guidelines address the rights and obligations of: 1) *fairness*, *accountability*, and *transparency*; 2) *autonomy* and *human determination*; 3) *data accuracy* and *quality*; 4) *safety* and *security*; and 5) *minimization of scope*. These principles can also guide the use of algorithms in the pre-trial risk context.