LUND UNIVERSITY

On the Governance of Artificial Intelligence through Ethics Guidelines

Larsson, Stefan

Published in:
Asian Journal of Law and Society

DOI:
10.1017/als.2020.19

2020

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

Link to publication

Citation for published version (APA):
Larsson, S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. *Asian Journal of Law and Society*, *7*(3), 437-451. https://doi.org/10.1017/als.2020.19

Total number of authors:
1

Creative Commons License:
Unspecified

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Peer-reviewed "postprint" from 18 May 2020. To be published in *Asian Journal of Law and Society*

On the Governance of Artificial Intelligence through Ethics Guidelines

Stefan LARSSON[*]

Lund University

Abstract

This article uses a socio-legal perspective to analyse the use of ethics guidelines as governance tool in the development and use of artificial intelligence (AI). This has become a central policy area in several large jurisdictions, including China and Japan, as well as the EU, focused here. Particular emphasis is in this article placed on the Ethics Guidelines for Trustworthy AI published by the EU Commission's High-Level Expert Group on Artificial Intelligence in April 2019, as well as the White Paper on Artificial Intelligence, published by the EU Commission in February 2020. The guidelines are reflected against partially overlapping and already existing legislation as well as the ephemeral concept construct surrounding AI as such. The article concludes by pointing to the i) challenges of a temporal discrepancy between technological and legal change; ii) the need of moving from principle to process in the governance of AI, and iii) and the multidisciplinary needs in the study of contemporary applications of data-dependent AI.

Keywords: AI governance; Ethics guidelines for trustworthy AI; EU Commission; Transparency in AI; AI and law.

[*] Stefan Larsson is a senior lecturer and Associate Professor in Technology and Social Change at Lund University, Sweden, Department of Technology and Society. He is a lawyer (LLM) and socio-legal researcher that holds a PhD in Sociology of Law as well as a PhD in Spatial Planning. His multidisciplinary research focuses on issues of trust and transparency on digital, data-driven markets, and the socio-legal impact of autonomous and AI-driven technologies. Recent publications of relevance include "Transparency in Artificial Intelligence", co-authored with Fredrik Heintz, in *Internet Policy Review* (2020), "The Socio-Legal Relevance of Artificial Intelligence" in *Droit et Société* (2019), and "Conceptions in the Code. How Metaphors Explain Legal Challenges in Digital Times" (Oxford University Press, 2017). Acknowledgments: The research for this paper was enabled by The Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS), within the AI Transparency and Consumer Trust project; The Swedish Research Council (VR; grant no. 2019-00198) in the AIR Lund (Artificially Intelligent use of Registers at Lund University) research environment; and Vinnova (2019-00848) in the Framework for Sustainable AI project.
Email: stefan.larsson@lth.lu.se

I. Introduction and Purpose of Study: Ethics Guidelines as a Tool for Governance

Much like Karl Renner looked for the principal content of property law in times of technology-driven societal transformation in the industrializing western Europe,[1] contemporary society is seeking its proper forms of governance in a digital transformation[2] driven by *platformisation*,[3] *datafication*[4] and algorithmic automation.[5] Much like how Eugene Ehrlich proposed a study of the living law,[6] paralleled by Roscoe Pound's separation of law in books from law in action,[7] contemporary governance of artificial intelligence (AI) is, too, separable in terms of hard and soft law.[8] This article could be read in light of these foundational socio-legal scholars shaping sociology of law as a scientific discipline, that has inspired much thought on the relationship between social change, law and new technology.[9]

In its Communication from April 2018,[10] the EU adopted an explicit strategy for AI and appointed The High-Level Expert Group on AI, consisting of 52 members, to provide advice on both investments and ethical governance issues in relation to AI in Europe. In April 2019, the expert group published Ethics Guidelines for Trustworthy Artificial Intelligence (hereinafter Ethics Guidelines),[11] which – despite explicitly pointing out that the guidelines do not deal with legal issues – clearly indicate issues of responsibility, transparency and data protection as entirely central

[1] See Renner (2010/1949) *The Institutions of Private Law and Their Social Functions*. New Brunswick, USA, and London, UK: Transaction Publishers. First published in German in 1904 as *Die Rechtsinstitute des Privatrechts und ihre soziale Funktion*.

[2] See Larsson (2014).

[3] See van Dijck, Poell & de Waal (2018); Poell, T. & Nieborg, D. & van Dijck, J. (2019).

[4] Mejias & Couldry (2019).

[5] Larsson (2018).

[6] Ehrlich (1913).

[7] Pound (1910).

[8] For an insightful analysis of Ehrlich's and Renner's theoretical contributions to sociology of law, see Nelken (1984). For discussion on the reuse and reinterpretation of socio-legal classic theory, see Nelken (2007), and for a particular digital context, see Larsson (2013).

[9] For an extensive account of this trichotomy, see Larsson (2017).

[10] EU Commission (2018) Artificial Intelligence for Europe. Brussels, 25.4.2018 COM(2018) 237 final.

[11] The High-Level Expert Group (2019a).

parts of the development of trustworthy AI. Over the last few years, a number of ethics guidelines have been developed relating to AI; by companies, research associations and government representatives.[12] Many overlap in part with already existing legislation, but it is often unclear how the legislation and guidelines are intended to interact more precisely. In particular, the way in which the standpoints in principle are intended to be implemented is often unclear. In other words, the Ethics Guidelines focus on normative standpoints, but are often weak from a procedural perspective. The Ethics Guidelines of the EU Commission's expert group are a clear sign of an ongoing governance challenge for the EU and its member states. Interestingly, during her candidature, Ursula von der Leyen, the new President of the EU Commission, stated that during her first 100 days in office, she would "put forward legislative proposals for a coordinated European approach to the human and ethical implications of AI".[13] Consequently, in February 2020, the European Commission issued a digital strategy including proposals for empowering excellence and trust in AI and a White Paper on Artificial Intelligence.[14] At the same time, The EU Commission's take on AI development and governance signifies a global trend on governmental and jurisdictional approaches to seeing both societal and industrial benefits with AI in tandem with ethical and legal concerns that need to be addressed and governed. This notion of development and governance of high-potential/high-risk have earlier been described as that they are "inevitably and dynamically intertwined", with regards to emerging technologies.[15]

Part of the challenge for the EU, arguably, consists of balancing regulation against the trust that exists in technical innovation and societal development overall, to which AI and its methods can contribute, and which is therefore not desirable to risk undermining with unbalanced or hastily introduced regulation. As societal use and dependency on AI and machine learning is increasing, society increasingly needs to understand any negative consequences and risks, how interests and power are distributed, and what needs exist for both legal and other types of governance.

[12] Cf. Jobin et al. (2019); Hagendorff (2020); Mittelstadt (2019).

[13] Van Der Leyen (2019, p. 13).

[14] European Commission (2020) White Paper on Artificial Intelligence: a European approach to excellence and trust, Brussels. COM(2020) 65 final.

[15] Mandel (2009), p. 75, (analysing advancements in biotechnology, nanotechnology and synthetic biology from a regulatory perspective).

This article focuses on ethics guidelines as tools for governance, points to the interplay with legal tools for governance and discusses the particular features of AI development that have led to ethics issues gaining such a prominent position. Particular focus is placed on the Ethics Guidelines for 'trustworthy AI' as well as the Commission's White paper on AI. *Firstly*, the article focuses on the definitional struggles around the concept of AI, in order to clarify the relationship between the definition and the governance of AI. Since the actual definition of AI is highly debatable, and may depend on the disciplinary field in which the person making the definition is based, it will arguably have an effect on its governance. Here is argued for the need to regard the technologies in their applied context, and in their interplay with human values and societal expressions, which is not least underlined by the dependence of machine learning on large amounts of data or examples as its foundation. *Secondly*, the key features of the ethical approach on AI governance is outlined, addressing some of its critique, with a particular focus on the EU. This must arguably be placed in a broader context of governance tools that nevertheless often share some principle-based central values relating to the control of data, the degree of reasonable transparency and how responsibility should be allocated, and a brief comparison to Chinese and Japanese guidelines are provided. *Finally*, the article concludes with a socio-legal perspective on ethics guidelines as a form of governance over the AI development. The governance using ethics guidelines is highly dependent on recent insights from critical AI studies about the societal challenges relating to fairness, accountability and transparency.[16] At the same time, the governance issue must inevitably deal with temporal aspects of the difference between how legislation is formed and how rapid development has been for the underlying elements of AI.

II. What *is* AI?

Despite – or perhaps because of – the increased attention that AI and its developed methods are receiving in multidisciplinary research, media and policy work, there is no clear consensus on how AI should best be defined. This seems to be the case not only with regards to public perceptions,[17]

[16] Larsson (2019).

[17] Fast & Horvitz (2017).

but also to computer science[18] and law.[19] For example, Gasser and Almeida establish that one cause of the difficulty of defining AI from a technical perspective is that AI is not a single technology, but rather "a set of techniques and subdisciplines ranging from areas such as speech recognition and computer vision to attention and memory, to name just a few".[20] A number of definitions have been expressed, both within research and in government agency reports, but a major challenge is that the methods express a moveable and changing field. I would here like to emphasise the dynamic of the concept construct as it has been discussed within traditional AI research, and also offer some central aspects that can still be highlighted, as well as show what The High-Level Expert Group is concentrating on.

In conjunction with The High-Level Expert Group publishing the Ethics Guidelines, a definition document was also published, aimed at clarifying certain aspects of AI as a scientific discipline and as a technology.[21] One purpose that it highlights is to avoid misunderstanding and to achieve a commonly shared knowledge about AI that can be used fruitfully also by non-experts, and to indicate details that may contribute to the discussion about the Ethics Guidelines. The High-Level Expert Group uses as its first starting point the definition provided in the EU Commission's communication on AI in Europe, published in April 2018, which is then developed further.

> 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.
>
> AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications). (The High-Level Expert Group, 2019b, p. 1)

This definition concentrates particularly on autonomy, i.e. that there is a measure of agency in AI systems, and points out that the systems can consist of both physical robots and also pure software systems. At the same time, the examples provide a clear indication of what they are aiming at, and, extrapolating, what the governance objects of the Ethics Guidelines consist of. As a

[18] Monett, Lewis & Thórisson (2020).

[19] Martinez (2019).

[20] Gasser & Almeida (2017), p. 59.

[21] The High-Level Expert Group (2019b).

software-based category, they point to voice assistants, image analysis software, search engines, and speech and face recognition systems, while for hardware-based applications they indicate advanced robots, autonomous cars, drones and the linked-up devices that are seen as part of the Internet of Things. As autonomy is emphasised, this can in combination be interpreted as not applying to all drones or all linked-up devices, only those that have an autonomous or even learning element. What characterises an ‘advanced’ robot does not necessarily entail a simple demarcation, which we can expect to be changing over time. This is clearly a “moving target” that seems to be an inherent element of AI, sometimes described as an “odd paradox” or the “AI effect”.[22]

The High-Level Expert Group also notes that an explicit part of AI is the intelligence concept, which is a particularly elusive element that has been included since the area was originated. Legg and Hutter, for example, gather together more than 70 different definitions of the intelligence concept in itself.[23] In addition to listing a number of psychological definitions, they also indicate how the definitions used in AI research have focused on different inherent aspects, with differing emphasis on problem-solving, improvement and learning over time, good performance in complex environments, or the generalisability of achieving domain-independent skills that are needed to manage a number of domain-specific problems. The intelligence concept also leads to a number of human associations, such as the ability to have feelings and self-awareness that cannot be said to be a living part of the methods and technologies that are causing the explosion of applied AI today, and thus not a central object for governance through ethics guidelines. It can therefore be established that contemporary AI primarily includes a number of technologies and analysis methods that have been gathered together under the umbrella concept of ‘artificial intelligence’, namely machine learning, natural language learning, image recognition, ‘neural networks’ and deep learning. Machine learning in particular – a field that expressed in simple terms is about methods for making computers ‘learn’ based on data, without the computers having been programmed for that particular task – is a field that has developed rapidly in just the last few years through access to historically incomparable amounts of digital data and increasing analytical processing power. This has led to that contemporary AI generally refers to “the computational capability of interpreting huge amounts of information in order to make a decision, and is less

[22] Cf. Stone et al. (2016).

[23] Legg & Hutter (2007).

concerned with understanding human intelligence, or the representation of knowledge and reasoning", according to Virginia Dignum, a professor in AI and ethics that also is a member of the High-Level Expert Group.[24]

The complexity of the concept construct has led The High-Level Expert Group to put forward a fairly complex definition, which thus expands the EU Commission's first definition. It also includes the AI functionality in its systemic context, i.e. the fact that it is often part of a larger whole,[25] includes the division of machine learning into structured and unstructured data, and the fact that AI systems are primarily goal-driven to achieve something that a human being has defined:

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.[26]

There are thus differing aspects of AI to be considered in the definition of AI as a challenge to regulation, where the most central ones for today's development and use of AI tend to concern a) autonomy/agency, b) self-learning from large data amounts (or "adaptability"), and c) the degree of generalisable learning. Finally, as a step towards a wider social sciences-based discussion and in the light of the challenges that AI has displayed in its implementation and interaction with society's values and structure, it can be argued that there are multidisciplinary advantages of not leaning too heavily towards a computer sciences-based definition of AI. The definition is in itself a form of conceptual control that impacts on the regulation debate, and we therefore need to be both careful and take a multidisciplinary approach when making definitions.[27]

[24] Dignum (2019), p. 3.

[25] Also emphasized in Larsson & Heintz (2020).

[26] The High-Level Expert Group 2019b), p. 6.

[27] For definitional and metaphoric aspects of new technologies and their regulatory implications, see Larsson (2017).

III. EU – Trustworthy AI

If we first look at the discussions about AI and ethics that are held in the global arena, we can establish that it is currently a lively subject among academics and policy-oriented bodies. Ethics Guidelines in particular, as a governance tool, have seen very strong development over the last few years. For example, a study of the global AI ethics landscape, published in 2019, identified 84 documents containing ethical principles or guidelines for AI.[28] The study concluded that there is relative unanimity globally on at least five principle-based approaches of ethical character: 1) transparency, 2) justice and "fairness", 3) non-harmful use, 4) responsibility, and 5) integrity/data protection. At the same time, the study establishes that there are considerable differences in how these principles are interpreted, why they are considered important, what issue, domain or actors they relate to, and how they should be implemented. The single most common principle is "transparency", a particularly multifaceted concept it seems.[29]

Meanwhile, the ethics researcher Thilo Hagendorff considers that the weak point of Ethics Guidelines is that AI ethics – like ethics in general – lack mechanisms for creating compliance or for implementing their own normative claims.[30] According to Hagendorff, this is also the reason why ethics is so appealing to many companies and institutions. When companies and research institutes formulate their own ethics guidelines, repeatedly introduce ethical considerations or adopt ethically motivated own undertakings, Hagendorff argues for that this counteracts the introduction of genuinely binding legal frameworks. He thus places great emphasis on the avoidance of regulation specifically as a main aim of the AI industry's ethics guidelines. Mark Coeckelbergh, professor of media and technology philosophy, who is also a member of the High-Level Expert Group, expresses similar risks, "that ethics are used as a fig leaf that helps to ensure acceptability of the technology and economic gain but has no significant consequences for the development and use of the technologies".[31] Even if this reminder has merits, and the risk is real –

[28] Jobin et al., (2019).

[29] For a conceptual analysis of transparency in AI, see Larsson & Heintz (2020), and a socio-legal commentary in Larsson (2019) proposing seven aspects of socio-legal relevance for transparency in applied AI.

[30] Hagendorff (2020).

[31] Coeckelbergh (2019), p. 33.

it is indubitably an incentive for many companies to avoid tougher regulation by pointing to "self-regulation" and the development of internal policies with weak actual implementation – there may yet be other reasons for ethics as a tool for governance to have been emphasised so heavily within AI development. Even though self-regulation is surely used as an argument for avoiding the intervention of concrete legislation, the question is still whether the rapid growth of the AI field in particular does not play just as important a role in the conclusion that this particular field has required a softer approach while waiting for critical research to catch up and offer a stable foundation for potent regulation. The question is, however, what codification of AI ethics would involve, and which parts of it would be best suited for legislation.

A. Ethics Guidelines for Trustworthy AI

In April 2018, the EU adopted a strategy for artificial intelligence (AI), and appointed The High-Level Expert Group with its 52 members, to provide advice on both investments and ethical governance issues in relation to AI in Europe. In December 2018, the Commission presented a coordinated plan – "Made in Europe" – which had been developed with the member states to promote the development and use of AI in Europe. For example, the Commission expressed an intention that all member states should have their own strategies in place by the middle of 2019, which did not completely materialise. The expert group was appointed via an open call, and consists of a fairly mixed group of researchers and university representatives (within areas such as robotics, computer science and philosophy), as well as representatives of industry (such as Zalando, Bosch and Google), and civil society organisations (such as Access Now[32], ANEC[33] and BEUC[34]). The composition has not avoided criticism, however. For example, in May 2019, Yochai Benkler, a professor at Harvard Law School – perhaps most famous for his optimistic writings on collaborative economies, focusing on phenomena such as Wikipedia, Creative Commons and open source code – expressed a fear that representatives of industry were allowed too much control over regulatory issues governing AI.[35] Benkler drew parallels between the EU Commission's expert group, Google's failed committee for AI ethics issues, and Facebook's investment in a German AI

[32] An international non-profit, human rights, public policy, and advocacy group dedicated to an open and free Internet.

[33] The European Association for the Co-ordination of Consumer Representation in Standardisation.

[34] The European Consumer Organisation, bringing together 45 European consumer organisations from 32 countries.

[35] Benkler (2019).

and ethics research centre. Similarly, technology and law researcher Michael Veale criticizes the High-Level Expert Group – focusing on the set of policy and investment recommendations[36] that was published after the Ethics Guidelines – for failing to address questions of power, infrastructure as well as organisational factors (including business models) in contemporary data-driven markets.[37] When the Ethics Guidelines were published, they were also criticised by members of the expert group itself. Thomas Metzinger, a philosopher at the Johannes Gutenberg University Mainz, critically described the process as "ethics washing" in an opinion piece where he described how the drafts produced on prohibitions against certain areas of use, such as autonomous weapons systems, had been toned down by representatives of industry and allies of these, to land in softer and more permissive wordings.[38]

The Ethics Guidelines have had a clear impact on the subsequent White Paper on AI from the EU Commission, see below, but it still remains to be seen what sort of importance and impact all of these sources will have on European AI development. The Ethics Guidelines point out that trustworthy AI has three components that should be in place throughout the entire lifecycle of AI:

a) it should be **legal** and comply with all applicable laws and regulations;
b) it should be **ethical** and safeguard compliance with ethical principles and values, and
c) it should be **robust**, from both a technical and a societal viewpoint, as AI systems can cause unintentional harm, despite good intentions.

The guidelines focus on ethical issues (b) and robustness (c), but leave legal issues (a) outside the explicit guidelines. It does this despite the fact that issues that are fairly well-anchored in law, such as responsibility, anti-discrimination and – not least – data protection, still fall within the framework for ethics. Just as the expert group established, many parts of AI development and use in Europe are already covered by existing legislation. These include the Charter of Fundamental Rights, The General Data Protection Regulation (GDPR), the Product Liability Directive,

[36] The High-Level Expert Group on Artificial Intelligence (2019c).

[37] Veale (2020); see also Koulu (2020) on the shortcomings of human control over automation and the need to broaden the discussion from current focus on technology and ethics to discussions about societal structures and law.

[38] Metzinger (2019).

directives against discrimination, consumer protection legislation, etc. Even though ethical and robust AI is to some extent often already reflected in existing laws, its full implementation may reach beyond existing legal obligations.

The expert group provides with four ethical principles constituting the "foundation" of trustworthy AI: 1.) Respect for human autonomy; 2.) Prevention of harm; 3.) Fairness; and 4.) Explicability. However, for the realisation of trustworthy AI, they address seven main prerequisites, which, they argue, must be evaluated and managed continuously during the entire lifecycle of the AI system:

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

As mentioned, although the guidelines emphasise that they focus on ethics and robustness, and not on issues of legality, it is interesting to note that both anti-discrimination (5) and protection of privacy (3) are developed as two of the seven central ethical prerequisites for the implementation of trustworthy AI. In relation to the investment and policy recommendations also published by the expert group, it recommends features such as a risk-based approach that is both proportional and effective in guaranteeing that AI is legal, ethical and robust in its adaptation to fundamental rights.[39] Interestingly, the expert group calls for comprehensive mapping of relevant EU regulations to be carried out, in order to assess the extent to which the various regulations are still fulfilling their purposes in an AI-driven world. They highlight that new legal measures and control mechanisms may be needed to safeguard adequate protection against negative effects, and to enable correct supervision and implementation.

The Ethics Guidelines argues for the need for processes to be transparent in the sense that the capacities and purpose of AI systems should be "openly communicated, and decisions – to the

[39] The High-Level Expert Group (2019c). See also the Opinion of the German Data Ethics Commission (2019).

extent possible – explainable to those directly and indirectly affected”.[40] A key reason is to be building and maintaining users’ trust. In the literature relating to ethics guidelines targeted at AI, it has been argued that *transparency* is not an ethics principle in itself, but rather a “pro-ethical condition”,[41] for enabling or impairing other ethical practices or principles. As argued in a study on the socio-legal relevance of artificial intelligence, there are several contradictory interests that can be linked to the issue of transparency.[42] Consequently, there are other reasons than pure technical complexity for why certain approaches may be of a ‘black box’ nature, not least the corporate interests of keeping commercial secrets and holding intellectual property rights.[43] Furthermore, the Ethics Guidelines contain an assessment list for practical use by companies. During the second half of 2019, over 350 organisations have tested this assessment list and sent feedback. The High-Level Group is (during the spring of 2020) in the process of revising its guidelines in light of this feedback and will finalise this work during 2020.

B. The White Paper on Artificial Intelligence

As mentioned, Commission President Ursula von der Leyen announced in her political Guidelines[44] a coordinated European approach on the human and ethical implications of AI as well as a reflection on the better use of big data for innovation. The White paper on AI from February 2020 could be seen in light of this commitment. In the White Paper, it is expressed that the Commission supports a regulatory and investment oriented approach with what it calls a “twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology”, and that the purpose of the White Paper is to set out policy options on how to achieve these objectives.[45] A key proposal in the White Paper is taking a risk-based, sector-specific approach to regulating AI, where high-risk applications are distinguished from all other applications. Firstly, a high-risk sector is where “significant risks can be expected”, which may

[40] The High-Level Expert Group (2019a), p. 13.

[41] Turilli & Floridi (2009).

[42] Larsson (2019); Larsson & Heintz (2020).

[43] See Pasquale (2015).

[44] Von der Leyen (2019).

[45] White Paper (2020), p. 1.

initially include "healthcare; transport; energy and parts of the public sector".[46] In addition, the application should be used in a sector where "significant risks can be expected", which means a cumulative approach. This proposal is an either/or approach on risk, and more nuanced alternatives have been proposed elsewhere, for example by the German Data Ethics Commission.[47]

There is a clear value-base in the White Paper, with a particular focus on the concept of trust: "Given the major impact that AI can have on our society and the need to build trust, it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection".[48] An expressed aim of the EU's Policy framework is to mobilise resources to achieve an 'ecosystem of excellence' along "the entire value chain". The key elements of a *future* regulatory framework for AI in Europe is to create a "unique ecosystem of trust", which is described as a policy objective in itself. The Commission's hope is that a clear European regulatory framework would "build trust among consumers and businesses in AI, and therefore speed up the uptake of the technology".[49]

The White paper makes a clear address to the human-centric approach based on the Communication on Building Trust in Human-Centric AI, which is also a central part of the Ethics Guidelines discussed above. The White Paper states that the Commission will take into account the input obtained during the piloting phase of the Ethics Guidelines prepared by the High-Level Expert Group on AI. Interestingly enough, the Commission concludes that those regarding *transparency*, *traceability* and *human oversight* are not specifically covered under current legislation in many economic sectors. The lack of transparency, the Commission brings forward, makes it "difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation".[50]

C. Asian comparison

From an Asian socio-legal perspective, the Chinese and the Japanese developments on AI policy and governance are significant but will only briefly be addressed here. The core in China's AI

[46] White Paper (2020) p. 17.

[47] German Data Ethics Commission (2019).

[48] White Paper (2020), p. 2.

[49] *Ibid*., pp. 9-10.

[50] *Ibid*.,, p. 14).

strategy can be found in the New Generation Artificial Intelligence Development Plan (AIDP), issued by China's State Council in July 2017, and the Made in China 2025, released in May 2015.[51] For example, a goal expressed in the AIDP is to, by 2025, establish initial ethical norms, policies and regulations related to AI development in China by 2020, to be further codified by 2025.[52] This includes participation in international standard setting as well as deepening international cooperation in AI laws and regulations. In 2019, a National Governance Committee for the New Generation Artificial Intelligence was established, which published a set of governance principles.[53] In May 2019, the so-called Beijing AI Principles, which is another set, were released by the Beijing Academy of Artificial Intelligence, depicting the core of its AI development as the realization of beneficial AI for humankind and nature. These Principles have been supported by various elite Chinese universities and companies including Baidu, Alibaba and Tencent.[54]

In Japan, an expert group at the Japanese Cabinet Office has elaborated Social Principles of Human-Centric AI (Social Principles), which has been published in March 2019 after public comments were solicited. In a comparison on Japanese and European initiatives, a recent study concludes that common elements of both notions of governance includes that AI should be applied in a manner that is 'human-centric' and should be committed to the fundamental (constitutional) rights of individuals and democracy.[55] A particular difference is however, according to Kozuka, that Japan's Social Principles are more policy-oriented, while the European Ethics Guidelines have a rights-based approach. Interestingly, Kozuka – with references to Lawrence Lessig in the paper – concludes that "the role of the law as a mechanism of implementation will shrink and be substituted by the code as the use of AI becomes widespread".[56] This notion is particularly

[51] For a comprehensive analysis of the implications and direction of the AIDP, see Roberts et al. (2019).

[52] On ongoing efforts to develop AI governance theories and technologies from the perspective of China, see Wu, Huang & Gong (2020).

[53] National Governance Committee for the New Generation Artificial Intelligence (2019) Governance principles of the new generation artificial intelligence – developing responsible artificial intelligence.

[54] Daly et al. (2019).

[55] Kozuka (2019), p. 322.

[56] *Ibid.*, p. 329.

meaningful in relation to automated policy-implementation on large-scale digital platforms, shaping both human and institutional behaviour.[57]

IV. Conclusions

This article has put forward the difficulty of defining AI as one of the regulatory challenges that follow from the implementation and development of AI. While the historically visionary and contemporary heterogeneous AI field arguably provides for favourable conditions for research and development, the conceptual fuzziness creates a challenge for regulation and governance. It is perhaps the data dependency of today's machine learning – much critical and recent research show – in combination with a complexity that creates a lack of explainability, that stresses the risks of resulting in societal imbalances not only being reproduced, but also reinforced, at the same time as they evade detection. Furthermore, the article provides with an account on recent boom in ethics guidelines as a tool for governance in the field of AI, but with particular focus on the EU. Finally, three main concluding statements from the perspective of law and society can be made.

A. The temporality issue of technology and law

History teaches us that regulatory balancing is difficult, especially in times of rapid technological change in society. At the same time, legal scholars such as Karl Renner, who analysed property laws of Western Europe's industrialization, also teach that law can be an extremely dynamic and adaptive organism. It is conceivable that central parts of the Ethical Guidelines may be formalized with support for European and national legislation and regulation, focussing on the importance of ("an ecosystem of") trust. The interpretation of *existing legislation* in light of functionalities, possibilities and challenges of AI systems is also a matter of serious concern associated with major challenges. Even though the "legal lag" is more complex than it may seem,[58] the speed of change, in particular, is still repeatedly a difficult challenge in relation to the inertia of traditional

[57] See Katzenback & Ulbricht (2019) on "algorithmic governance"; Larsson (2013); van Dijck, Poell & de Waal (2018).

[58] See Bennet Moses (2011).

regulation.[59] Legislative processes aimed at learning technologies with increasing agency[60] require reflection, critical studies and more knowledge in order to be able to find the desirable societal balances between various interests. Especially *transparency*, *traceability* and *human oversight* are not clearly covered – or understood – under current legislation in many economic sectors. The temporal aspect of the difference between new technologies and well-fitted regulation, in combination with the many-headed balancing of interests, is very likely a significant contributor to why governance in the area is heavily characterised by ethics guidelines at the moment.

B. From principles to process

The White paper signifies an ongoing process of evaluation towards where the principled take on AI governance expressed by a multitude of ethics guidelines can find a balanced formalisation in law. This is also signified by the work conducted by The High-Level Group, as it is revisiting and assessing the Ethics Guidelines jointly with companies and stakeholders during 2019 and 2020. The Member States' supervisory authorities and agencies could be addressed specifically here too, in the sense that they will very likely be the ones to carry out relevant parts of any regulatory approach on AI focusing what The High-Level Expert Group has expressed as a need for "explicability", i.e. auditability and explainability. This particular aspect of transparency stresses the need for both methodological development[61] and likely closer collaboration between relevant supervisory authorities than what is often the case at a Member State level.

The great range of ethics guidelines still displays a core of principal values, but – being ethical guidelines – are relatively poor in procedural arrangements, compared to law. This can be understood as an expression of how quickly the transition in society towards a data-dependent and applied AI has been, where the principle stage is essential. The subsequent procedural stage is necessary, however, both to strengthen the chances of implementing the principal values as well as to formalise in legislation, assessment methodologies and standardisation. If one can regard the growth in ethics guidelines as an expression of the rapidity of the development of AI methods, the procedural stage is an expected second stage. However, if one regards the ethics guidelines as

[59] I.e. Abel (1982).

[60] Hildebrandt (2015).

[61] See Larsson (2018) on the "algorithmic governance" of data-driven markets.

industry's reluctance to accept regulation of its activities, as a soft version of legislation that is intended to be toothless, then the procedural stage will meet considerable resistance. Perhaps the lack of expressing power structures of contemporary data-driven platform markets – emphasized by critics – is a sign of the regulatory struggles to come in the leap from principles to process.

C. The multidisciplinary AI research need

Contemporary data-dependent AI should not be developed in technological isolation without continuous assessments from the perspective of ethics, cultures and law.[62] Furthermore, given the applied status of AI, it is imperative that humanistic and social scientific AI research is stimulated jointly with technological research and development. Given aspects of learning in data-dependent AI, there is an interaction at hand where human values and societal structures constitute the training data. This means that social values and informal norms may be reproduced or even amplified – sometimes with terrible outcomes. From an empirical approach, one could conclude that it is often human values and skewed social structures that lead to automated failures. In applied AI, learning simply arises not only from good and balanced examples, but also from the less proud sides of humanity: racism, xenophobia, gender inequalities and institutionalized injustices.[63] Challenges here will thus be to sort normatively among the underlying data, or alternatively to take a normative view on the importance of the automation and scalability of self-learning technologies so that the reproductive and amplifying tendencies become better and more balanced than their underlying material. There is, therefore, a multidisciplinary need for research in this field, that require collaboration between the mathematically informed computer scientific disciplines that have deep insights into how AI systems are built and operate and the humanities and social science-oriented disciplines that can theorize and understand their interaction with cultures, norms, values, attitudes or the meanings and consequences for power-relations, states and regulation.

In conclusion, the AI development issue has come to take a value-based and ethics-focused development within the European administration with a focus on trustworthiness and human-centric design. It is an answer to the question of how to look at AI and its qualities which here is

[62] See Bennet Moses (2011); Koulu (2020); Larsson (2019); Veale (2020); Yeung & Lodge, eds. (2019).

[63] Discussed by Larsson (2019) in terms of a "mirror for social structures".

found to be commendable: the precision of self-learning and autonomous technologies needs to be assessed in its interaction with the values of society. It is a normative definition with bearing on future development lines – a good AI is a socially entrenched and trustworthy one.

References

AIDP (2017) New Generation Artificial Intelligence Development Plan (AIDP), China's State Council.

Abel, R.L. (1981) "Law as Lag: Inertia as a Social Theory of Law." *Michigan Law Review*, 80: 785-809.

Beijing Academy of Artificial Intelligence. (2019) *Beijing AI principles*. Retrieved from: http://www.baai.ac.cn/blog/beijing-ai-principles [last visited 17 May 2020].

Benkler, Y. (2019) "Don't let industry write the rules for AI." *Nature*, 569(7755): 161-161.

Bennett Moses, L. (2011) "Agents of Change." Griffith Law Review, 20(4): 763-794.

Coeckelbergh, M. (2019) "Artificial Intelligence: Some ethical issues and regulatory challenges." *Technology and Regulation*, 31-34.

Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., Wang, W. & Witteborn, S. (2019) "Artificial Intelligence, Governance and Ethics: Global Perspectives." *The Chinese University of Hong Kong Faculty of Law Research Paper*, (2019-15).

Dignum, V. (2019) *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer International Publishing.

van Dijck, J., Poell, T., & de Waal, M. (2018) *The Platform Society: Public values in a connective world*. Oxford University Press.

EU Commission (2020) *White Paper on Artificial Intelligence. A European approach to excellence and trust*. COM(2020) 65 final.

EU Commission (2018) *Artificial Intelligence for Europe*. Brussels, 25.4.2018 COM(2018) 237 final. 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.

Fast, E., & Horvitz, E. (2017). "Long-term trends in the public perception of artificial intelligence." In *Thirty-First AAAI Conference on Artificial Intelligence*.

Gasser, U., &, V. A. Almeida (2017) A Layered Model for AI Governance, *IEEE Internet Computing*, 21(6): 58-62.

German Data Ethics Commission (2019) "Opinion of the Data Ethics Commission." https://datenethikkommission.de/en/ [last visited 15 May 2020].

Hagendorff, T. (2020). "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds and Machines*, 1-22.

The High-Level Expert Group on Artificial Intelligence (2019a) *Ethics Guidelines for Trustworthy AI*.

The High-Level Expert Group on Artificial Intelligence (2019b) *A Definition of AI: Main Capabilities and Disciplines. Definition developed for the purpose of the AI HLEG's deliverables*.

The High-Level Expert Group on Artificial Intelligence (2019c) *Policy and Investment Recommendations for Trustworthy Artificial Intelligence*.

Hildebrandt, M. (2015). *Smart Technologies and the End (s) of Law: novel entanglements of law and technology*. Edward Elgar Publishing.

Jobin, A., Ienca, M., & Vayena, E. (2019) "The global landscape of AI ethics guidelines." *Nature Machine Intelligence*, 1(9): 389-399.

Katzenbach, C., & Ulbricht, L. (2019). "Algorithmic governance." *Internet Policy Review*, 8(4).

Koulu, R. (2020). "Human control over automation: EU Policy and AI Ethics." *European Journal of Legal Studies*, 12(1): 9-46.

Kozuka, S. (2019) "A governance framework for the development and use of artificial intelligence: lessons from the comparison of Japanese and European initiatives." *Uniform Law Review*, 24(2): 315-329.

Larsson, S. & Heintz, F. (2020) "Transparency in Artificial Intelligence." *Internet Policy Review*, 9(2).

Larsson, S. (2019) "The Socio-Legal Relevance of Artificial Intelligence." *Droit et Société*, 103(3): 573-593. Special issue "Le droit à l'épreuve des algorithmes", ed. by Dubois C. & Schoenaers F.

Larsson (2014) "Karl Renner and (Intellectual) Property – How Cognitive Theory Can Enrich a Sociolegal Analysis of Contemporary Copyright." *Law & Society Review*, 48(1): 3-33. ISSN 1540- 5893.

Larsson (2013) “Sociology of Law in a Digital Society – A Tweet from Global Bukowina.” *Societas/Communitas,* 15(1): 281-295.

Legg, S., & Hutter, M. (2007) “A Collection of Definitions of Intelligence”, in Goertzel, B., & Wang, P. (Eds.). *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms*. Proceedings of the AGI Workshop 2006 (Vol. 157). IOS press, 17-24.

von der Leyen, U. (2019) *A Union that Strives for More. My agenda for Europe.* Political Guidelines for the Next European Commission 2019-2024.

Mandel, G.N. (2009) “Regulating Emerging Technologies.” *Law, Innovation and Technology*, 1(1): 75-92, DOI: 10.1080/17579961.2009.11428365

Martinez, R. (2019). “Artificial Intelligence: Distinguishing Between Types & Definitions.” *Nevada Law Journal*, 19(3): 2015-1042.

Mejias, U. A. & Couldry, N. (2019). “Datafication.” *Internet Policy Review*, 8(4). DOI: 10.14763/2019.4.1428

Metzinger, T. 8 (2019) “EU Guidelines. Ethics Washing Made in Europe”. Der Tagesspeigel, 8 April. https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html [last visited 17 May 2020].

Mittelstadt, B. (2019) “Principles Alone Cannot Guarantee Ethical AI.” *Nature Machine Intelligence* 1:501–507. https://doi.org/10.1038/s42256-019-0114-4

Monett, D., Lewis, C. W., & Thórisson, K. R. (2020). “Introduction to the JAGI Special Issue ‘On Defining Artificial Intelligence’ – Commentaries and Author’s Response.” *Journal of Artificial General Intelligence*, 11(2): 1-100.

National Governance Committee for the New Generation Artificial Intelligence (2019) *Governance principles of the new generation artificial intelligence – developing responsible artificial intelligence*.

Nelken D. (2007) “An E-mail from Global Bukowina.” *International Journal of Law in Context* (3): 189-202.

Pasquale, F. (2015) *The Black Box Society*. *The Secret Algorithms That Control Money and Information*. Harvard University Press.

Poell, T. & Nieborg, D. & van Dijck, J. (2019). ”Platformisation.” *Internet Policy Review*, 8(4). DOI: 10.14763/2019.4.1425

Pound, R. (1910) “Law in Books and Law in Action.” *American Law Review* (44): 12-36.

Renner, K. (2010/1949) *The Institutions of Private Law and Their Social Functions*. New Brunswick, USA, and London, UK: Transaction Publishers.

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2019). "The Chinese Approach to Artificial Intelligence: an Analysis of Policy and Regulation." Available at SSRN 3469784. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3469784

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., et al. (2016) *Artificial intelligence and life in 2030*. Stanford University.

Turilli, M. & Floridi, L. (2009) "The Ethics of Information Transparency," *Ethics and Information Technology* 11(2): 105–112.

Veale, M. (2020). "A critical take on the policy recommendations of the EU high-level expert group on artificial intelligence." *European Journal of Risk Regulation*.

Wu, W., Huang, T., & Gong, K. (2020). "Ethical Principles and Governance Technology Development of AI in China." *Engineering*, 6(3): 302-309.

Yeung, K., & Lodge, M. eds. (2019). *Algorithmic Regulation*. Oxford University Press.