European Commission
Rue de la Loi / Wetstraat 170
B-1049 Bruxelles/Brussel
Belgique/België

Topic Group on
Governance of Algorithms

governance@thinktech.ngo

Date
14.06.2020

Statement
on the White Paper on Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final

Dear Madam President of the European Commission, dear Members of the European Commission,

in the following you find the comment of ThinkTech e.V. to your White Paper. ThinkTech is an independent non-governmental organisation of researchers and entrepreneurs, coders, philosophers, lawyers, and artists driven by the goal to understand, discuss, and shape the impact of artificial intelligence and other digital technologies on individuals and society. Aiming for a beneficial and sustainable use of these technologies, we seek to assess and address their chances and challenges: to contribute to academic discourse, to foster an informed public debate, and to provide scientific policy advice.

We appreciate that an open public consultation has been put in place and are happy to contribute our perspective.

Yours faithfully,

Dominik Dahlhaus
Senior Research Fellow

Valentin Schrader
Senior Research Fellow

Svenja Breuer
Associate Fellow

Gabriel Lindner
Associate Fellow

ThinkTech e.V.
c/o Alexander Ladwein
Situlistraße 71b
80939 München
Deutschland

Sitz und Gerichtsstand: München
Registergericht: Amtsgericht München, Nr. VR208542
als gemeinnützig anerkannt durch das Finanzamt München

www.thinktech.ngo

Aufsichtsrat: Dominik Dahlhaus, Vorsitzender; Gunnar König, Chiara Ullstein

Vorstand: Alexander Ladwein, Vorsitzender und Vorstand (einzelvertretungsberechtigt) i.S.d. § 26 BGB; Louis Longin, Julius Morandell, Julia Pfeiffer

Table of Content

1 Introduction

The European Commission's White Paper on Artificial Intelligence (AI)[1] is one of the first comprehensive approaches to create a specific governance framework for this emerging digital technology. It pursues a twofold objective[2]: firstly, to build an "Ecosystem of Excellence" (pp. 5-9), i.e. to enable technological progress where possible to bring the benefits of AI to society and economy by "support[ing] the development and uptake of AI across the EU economy and public administration" (p. 5); secondly, to build an "Ecosystem of Trust" (pp. 9-25), i.e. to control or limit technology where necessary to minimise the risks of potential harm (pp. 2, 10) and, thus, to "build trust among consumers and businesses in AI, and therefore speed up the uptake of the technology" (pp. 9 f.).

In view of both the potential benefits and risks of AI, a comprehensive governance approach is to be welcomed. Nonetheless, some of the Commission's concrete proposals on the ecosystem of excellence (2.) and on the ecosystem of trust (3.) should undergo revision.

2 Ecosystem of Excellence

The Commision states their main policy proposals to further AI technology in the EU in section 4 of the whitepaper under the caption of "An ecosystem of excellence" (p. 6). In the following section we highlight the three areas of technology-specific promotion, environmental sustainability and public engagement where we see the need for improvement.

2.1 Technology-specific Promotion of AI

The Commission proposes to specifically promote AI development and uptake in the EU, favoring it over other technologies. The White Paper runs risk of giving the impression that AI is promoted as an end in itself, which becomes especially evident in the following points:

- "Promoting the adoption of AI by the public sector" (p. 8);
- "Promote the uptake of AI by business and the public sector" (survey form, p. 8);
- "Increase the financing for start-ups innovating in AI" (survey form, p. 8).

We see a strong need to emphasize a sensible governance approach that heeds ethical, legal, and social considerations in its approach to promote specific solutions. This includes being

[1] COM(2020) 65 final, 2020.
[2] On these two main purposes of most technical legislation, i.e. enabling and controlling technology: Kloepfer, M. (2002*), Technik und Recht im wechselseitigen Werden*, p. 86; Boehme-Neßler, V. (2011), *Pictorial Law: Modern Law and the Power of Pictures*, pp. 10 ff.

conscious about how AI-specific technology promotion distorts the market of ideas and solutions by favoring AI solutions over others.

In the following, we describe our understanding of why a sensible approach to promoting ethical and socially responsible AI solutions is important. We also make some suggestions on how promotion programs as part of a sensible governance approach could be designed in order to enable reasonable innovation, including within SMEs, startups and public administration that remains subject to the ethical regulatory framework that is discussed more in detail in Section 3.

2.1.1 Motivation to Further AI

From our understanding, the ethical/societal and economic perspective on AI go hand in hand. The goal should be to enable a flourishing landscape of ethical AI in the EU. To get there, the EU needs a rigorous framework for ethical AI, on the one hand, and strong European businesses to guarantee truly trustworthy AI technology, on the other. The latter is especially important as we believe ethical technology can only be achieved by thinking ethically right from the first line of code. This means that the task of developing AI for the EU should be taken up by businesses and innovation networks within the EU to ensure that European values are placed at the center of AI innovation. Furthermore, it will prevent risks such as privacy leaks with data flows to servers outside the EU. In order to decrease the already existing dependence on third-party actors such as the USA and China, who do not necessarily act in line with European interests and values, the promotion of AI technology in the EU is necessary. This will enable strong European businesses to build ethical technology tailored to European values, enforced by a EU framework for ethical AI. Promotion of AI technology should be done with this motivation in mind, not as an end in itself.

2.1.2 Designing Promotion Programs

We agree with the whitepaper that EU investment funds for AI are necessary. One major reason for the EU's rather low rate of innovation-to-business transfer is less availability of venture capital for start-ups as compared to the USA where private and institutional investors such as pension funds provide large volumes of venture capital or as compared to China with its state-owned funds[3]. EU investment funds specifically for AI are an approach for regaining a level playing field.

[3] European Commission, Digital Transformation Monitor (January 2018), *USA-China-EU plans for AI: where do we stand?,* retrieved from https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_AI%20USA-China-EU%20plans%20for%20AI%20v5.pdf (last downloaded June 14, 2020).

However, depending on the concrete design, AI-specific technology promotion distorts the market of ideas and solutions. Badly designed, it could lead to AI being favored over other solutions (such as social, managerial, or other technological innovation) that would possibly be simpler, cheaper, more robust, less opaque or simply more suitable for a given context.

Nonetheless, focusing these investment funds on the widely recognized future key technology that is AI, in a manner compliant with the proposed EU regulation, is a matter of prioritization as public money is not unlimited. Even though promoting innovation through venture capital funds irrespective of the underlying technology would be a sensible way to avoid possible market distortions and inefficiencies, a focus on AI technology does need to be made in order to make any difference on the global playing field of AI innovation whose political significance seems to be ever-growing.

To still avoid negative side-effects as much as possible, there needs to be sufficient forward-looking deliberation[4] on the (de)merit of AI-based solutions for the specific context when promotion decisions are being made. To some extent, this is addressed implicitly in the whitepaper where it states that “it is also essential to make sure that the private sector is fully involved in setting the research and innovation agenda and provides the necessary level of co-investment” (p. 7). Involving the private sector for co-investment should at least partially prevent misallocation through political actors and ensure efficient investment. However, we further endorse the "open and transparent sector dialogues" (p. 8) that the White Paper proposes and suggest that these dialogues center strongly around the needs of the sector's workers and clients and keep a wide enough view on the solution space[5].

Furthermore, the purchasing power of public procurement is a common instrument of industrial policy through steering demand. Promoting the uptake of mature AI technology compliant with a rigorous EU regulation in businesses and the public sector would effectively complement increased venture capital. This goes into the same vein as Kuziemski and Pałka[6] who not only recommend to "incentivise compliance-centred innovation in AI", but who also lay out principles on how to effectively do so.

2.1.3 No “move fast and break things”-Mentality

The White Paper states that "[i]t is essential that public administrations, hospitals, utility and transport services, financial supervisors and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities" (p.8). The notion of "rapid deployment" should be reconsidered, especially in areas of public interest since failures could lead to a decline in public trust towards AI technology and the institutions that deploy them. A

[4] Wilsdon, J./Willis, R. (2004), *See-through science. Why public engagement needs to move upstream*; see also Guston, D.H. (2014), in: *Social Studies of Science* 44 (2), p. 218–242.
[5] Gudowsky, N./Peissl, W. (2016), in: *European Journal of Futures Research* 4 (1), p. 135.
[6] AI Governance Post-GDPR: Lessons Learned and the Road Ahead, in: *Policy Brief* 2019/07, European University Institute.

discourse about the positive and negative effects of the technology involving academics and the public should not be omitted in favour of being particularly fast.

A recent example of how it should not be done is the job centre algorithm in Austria[7]. The Austrian algorithm is used to inform the decision on who should get access to training programs. Scientists and civil rights organisations criticize the discriminatory potential inherent to this algorithm[8]. In such cases, further dialogue with experts from academia and the broader public is necessary to prevent such tools from being discriminatory.

Of course, there are scenarios where the technology in question is already mature enough to justify accelerating the adoption. In such cases, a thorough risk assessment procedure should weigh the cost of waiting any longer against the possible damage the technology can create. But overall, a "move fast and break things" mentality stemming from silicon valley is not the right approach to the deployment of AI systems in areas of public interest.

In most parts of the White Paper the emphasis is on "[...] our values and fundamental rights such as human dignity and privacy protection" (p. 2) which is the right approach to the problems introduced by AI. Looking at the development of AI from a higher vantage point, the overarching goal should be the alignment of technology to our values. Because of this, it is also dangerous to paint the adoption of AI as a race between nations. There is no merit in being faster than other nations in using AI systems that go against our fundamental values. Additionally, furthering the 'AI race' narrative risks reinforcing it and may lead to ethical corner-cutting with potentially dire consequences.[9]

2.2 Environmental Sustainability and AI

Mentioning environmental goals as a side-note here and there throughout the whitepaper is not enough: Environmental sustainability should be much more built into the AI support programmes and there should be at least a plan on how to make sure that AI development and uptake in the EU will put environmental sustainability into practice. There could be different proposals on how to reach this, either from a regulatory perspective or through public promotion programmes. In both cases, a first step to elaborate such implementation plans could be to organize expert committees, similar to the High-Level Expert Group on AI (AIHLEG).

---

[7] See Wimmer, B. (2019), AMS gibt grünes Licht für Bewertung von Arbeitslosen durch Algorithmus, in: *futurzone*, 17.09.2019, retrieved from https://futurezone.at/netzpolitik/ams-gibt-gruenes-licht-fuer-bewertung-von-arbeitslosen-durch-algorithmus/400607894 (last downloaded June 14, 2020).

[8] See Köver, C. (2019), Streit um den AMS-Algorithmus geht in die nächste Runde, in: *Netzpolitik.org*, 10.10.2019, retrieved from https://netzpolitik.org/2019/streit-um-den-ams-algorithmus-geht-in-die-naechste-runde/#spendenleiste (last downloaded June 14, 2020).

[9] Cave, S./ÓhÉigeartaigh, S.S. (2018), An AI Race for Strategic Advantage, in: Furman, J. et al. (eds.): *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES '18. the 2018 AAAI/ACM Conference. New Orleans, LA, USA, 02.02.2018 - 03.02.2018*, p. 36–40.

An Expert Group on Environmentally Sustainable AI should accompany the creation and implementation of the EU AI governance framework by reviewing proposed measures with respect to their environmental impact. Furthermore, it should develop principles and programmes for regulation and public promotion, where it sees the need. Its goal should be to ensure that AI developed and used under the EU governance framework on AI will be as environmentally sustainable as possible and adequately contributes to the overall EU environmental sustainability goals. The group should be composed of experts coming in equal shares from academia, NGOs and business.

A concrete example for regulatory policy measures could be to introduce mandatory energy consumption information for AI services or products relying on them. In terms of promotion programmes, concrete steps could be to make environmental friendliness a condition for investment of the planned AI startup fund or dedicate a subsection of it to startups focused on environmental sustainability.

2.3 Public Engagement

The White Paper fails to emphasize the level of engagement of the European population with issues associated with emerging AI that would be necessary to make sure that AI will be aligned to society's values and that the public's trust can be established that is necessary for AI uptake. To ensure this, it is important to involve the general public broadly in dialog and deliberation about possible futures with AI and to introduce the technical aspects of AI to the public in a way that enables informed engagement. One possible solution would be to promote citizen training offers as part of the planned digital innovation labs. These could give the wider population a low-barrier possibility to get familiar with AI and motivate them for more engagement in the discussions on how a socially desirable future with AI could look like. Such engagement could be facilitated through citizen-expert juries, participatory technology assessment or other public dialog events, co-creative research agenda setting and innovation.[10] In these formats, it is particularly important to bring a range of stakeholders together (citizens, policy makers, engineers, researchers, workers and managers of affected companies/industries) in order to ensure that a wide range of expertise and concerns are included, as well as the necessary decision making power to make such dialog processes count when it comes to steering how AI will be developed and governed. Enabling the public to be involved early on in these processes can potentially mitigate public backlash[11] and enables citizens to timely and actively build a relation to the new AI technologies[12] that will be enmeshed in their professional and private lives.

[10] Gudowsky, N./Peissl, W. (2016), in: *European Journal of Futures Research* 4 (1), p. 135.
[11] Winickoff, D.E./Pfotenhauer, S.M. (2018), Technology governance and the innovation process, in: *OECD Science, Technology and Innovation Outlook*, p. 6.
[12] Felt, U./Fochler, M. (2010), in: *Minerva* 48 (3), pp. 219–238.

3 Ecosystem of Trust

In the White Paper's second part and under the caption of "An ecosystem of trust" (pp. 9-25), the Commission outlines its proposal for the EU's future AI regulation. This, in itself, is worth noting since it demonstrates a turning from previous ethics-based approaches that were heavily criticised as an attempt of "ethics washing"[13]. The Commission now commits to ensuring compliance not only with principles and values but with European citizens' fundamental rights and other rules of existing EU legislation (pp. 2, 9 ff.). Ethics would, thus, be a basis and not a substitution for regulation and self-regulation hence a governance option instead of a governance framework.

The outlined regulation is likely to have immense implications: For one thing, the Commission puts forward an EU-wide approach, fearing that divergent national regulation[14] could fragment the internal market (p. 2). For another, the EU's prospective legislation can be expected to have extraterritorial effects, as well: *De jure*, protection of European citizens entails safeguarding against extraterritorial and third State violations so that "[i]n the view of the Commission, it is paramount that the requirements are applicable to all relevant economic operators providing AI-enabled products or services in the EU, regardless of whether they are established in the EU or not. Otherwise, the objectives of the legislative intervention [...] could not fully be achieved" (p. 7). Additionally, businesses will, *de facto*, likely seek to minimise the cost of compliance and may, thus, adhere to strict EU requirements globally, given the economic importance of the European Single Market.[15]

In view of the importance of the EU's future AI policy, the European Commission should, however, clarify and partially reconsider the exact scope of the prospective regulatory framework. In particular, there are questions remaining regarding its concrete subject matter (1.), the objects of protection that are to be taken into consideration (2.), as well as the specific protective measures (3.).

3.1 Subject Matter: Regulating "AI"

The White Paper's title already indicates the European Commission's intention to regulate AI in general. This partially confirms concerns that had already been raised in reaction to the European Parliament passing a resolution in 2017 that asked the Commission to submit "a

[13] Wagner, B. (2018), in: Bayamlioglu, E. et al. (eds.), *Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen'*, 2018, pp. 84-87; Metzinger, T. (2019), Ethics washing made in Europe, in: *Der Tagesspiegel*, April 8, 2020, retrieved from https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html (last downloaded June 14, 2020).
[14] For an overview on policy approaches see Law Library of Congress (2019), *Regulation of Artificial Intelligence in Selected Jurisdictions*.
[15] See Young, A.R. (2015), in: *Journal of European Public Policy* 22 (9), pp. 1233-1252.

proposal for a legislative instrument on legal questions related to the development and use of robotics and AI foreseeable in the next 10 to 15 years, combined with non-legislative instruments such as guidelines and codes of conduct"[16] based on "a generally accepted definition of robot and AI"[17]: The White Paper's working assumption is that "the regulatory framework would apply to products and services relying on AI" provided in the EU (p. 16). This marks a turning point in EU technology legislation that is so far predominantly technology-neutral[18] and mostly follows the reasoning that "[r]egulation that is based on specific technology can quickly become outdated, and may lead to inefficient investment by market players"[19].

3.1.1 Non-Specificity of AI-Specific Regulation

Producing technology neutral regulation that will likely be applicable in spite of new technological developments without the need to be frequently revised and amended[20] evidently entails challenges for drafting in technology neutral terms[21]. By definition, technology-specific legislation promises to be more specific about the subject matter which it regulates. This is why the White Paper states that it is a "key issue [...] to determine the scope of [this framework's] application. [...] AI should therefore be clearly defined for the purposes of [the] White Paper, as well as any possible future policy-making initiative" (p. 16).

But in the case of AI technology-specific regulation may not be specific at all since the notion "AI" refers to various concepts, technologies, and applications.[22] Consequently, there are widely varying definitions, none of which appears to be commonly accepted.[23] The AIHLEG that refined a previous definition of the Commission[24], therefore, explicitly warns that its attempt at a definition "is a very crude oversimplification of the state of the art" and not intended to "precisely and comprehensively define all AI techniques and capabilities"[25].

A legal nomenclature that covers the wide range of techniques, capabilities, applications, and uses of "AI" in general[26] would, hence, need to be highly inclusive, probably at the expense of its

[16] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), in: *Official Journal of the European Union*, 2018/C 252/249.
[17] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), in: *Official Journal of the European Union*, 2018/C 252/239.
[18] See Reed, C. (2007), in: *SCRIPT-ed* 4 (3), pp. 264 f., with further references.
[19] COM (1999) 539 final, 1999, p. 14.
[20] Koops B.-J. (2006), in: Koops, B.-J. et al. (eds.) *Starting Points for ICT Regulation: deconstructing prevalent policy one-liners*, pp. 77, 83 ff.
[21] Reed, C. (2007), in: *SCRIPT-ed* 4 (3), pp. 279 f.
[22] European Parliament Research Service (2019), *Artificial Intelligence ante portas: Legal & ethical reflections, PE 634.427 – March 2019*, pp. 1 f.
[23] Buiten, M.C. (2019), in: *European Journal of Risk Regulation* 10 (1), p. 43.
[24] COM(2018) 237 final, p. 1.
[25] AIHLEG (2019), A definition of AI, p. I (disclaimer).
[26] European Parliament Research Service (2019), *Artificial Intelligence ante portas: Legal & ethical reflections, PE 634.427 – March 2019*, pp. 1 f.

suitability for differentiation[27] and might, thus, not provide the legal certainty required for legislation. A corresponding regulatory framework based on such a broad definition and intended to address the different domain- and application-specific risks and considerations would, thus, need to be extremely or perhaps even "impossibly wide in scope"[28]. This has led the AI100 Standing Committee and Study Panel to the unambiguous conclusion that any "attempts to regulate 'AI' in general would be misguided. [...] Instead, policymakers should recognize that to varying degrees and over time, various industries will need distinct, appropriate, regulations"[29] . The AIHLEG, correspondingly, states that while the Ethics Guidelines for Trustworthy AI "aim to provide guidance for AI applications in general" their implementation "needs to be adapted to the particular AI-application", given the "context-specificity of AI systems".[30]

3.1.2 Context-Overarching Needs for Regulation

None of this is to say that there are no context-overarching needs for regulation of "AI": On the one hand, there are cross-sectional issues that need to be addressed in the context of a new technology. On the other hand, a regulatory rationale may be directly linked to the specific technology itself.

There are various challenges usually associated with AI, arising from characteristics commonly attributed to AI, such as but not limited to complexity, opacity, openness, autonomy, (un-)predictability, data-dependence, or vulnerability due to constant interaction with outside information.[31] We think that these supposed technology-specific aspects do not justify (technology-specific or any) regulatory interventions on their own. There are, nonetheless, cross-sectional issues arising from fundamental rights or legal principles which may be especially pronounced in the context of AI - due to its above-listed characteristics: Even though humans can act on bias, AI systems can enshrine, exacerbate, and perpetuate these biases while backing up their analysis "with reams of statistics, which give them the studied air of even handed science"[32].[33]

Even though the effectiveness of legislation enacted prior to AI to protect fundamental rights and legal principles and to weigh conflicting interests might be impaired due to said challenges of AI, "these are exactly the kinds of problems faced generally in applying rules, designed for an

---

[27] See European Parliament Research Service (2019), *Artificial Intelligence ante portas: Legal & ethical reflections, PE 634.427 – March 2019*, p. 2.
[28] Reed C. (2018), in: *Philosophical Transactions of the Royal Society A*, p. 377; Buiten, M.C. (2019), in: *European Journal of Risk Regulation* 10 (1), p. 45.
[29] Stone, P. et al. (2016), *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence. Report of the 2015 Study Panel*, p. 48.
[30] AIHLEG (2019), *Ethics Guidelines for Trustworthy AI*, p. 6.
[31] See Expert Group on Liability and New Technologies – New Technologies Formation (2019), *Liability for Artificial Intelligence and other emerging digital technologies*, pp. 32 f.
[32] O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, p. 162.
[33] See Dignum, V. et al. (2020), *First Analysis of the EU Whitepaper on AI*, retrieved from https://allai.nl/first-analysis-of-the-eu-whitepaper-on-ai/ (last downloaded June 14, 2020).

older socio-technical landscape, in a new context."[34] Technology-specific regulation of these cross-sectional issues fails to recognize the imperative of assessing potentially existing competing interests in a concrete and specific manner and may well result in unnecessary legislative fragmentation, duplication, compartmentalisation, and inconsistency[35]. These trade-offs are especially evident in the following scenario tweeted by *Hinton* in response to the EU White Paper: "Suppose you have cancer and you have to choose between a black box A.I. surgeon that cannot explain how it works but has a 90% cure rate and a human surgeon with an 80% cure rate. Do you want the A.I. surgeon to be illegal?"[36] In this example, the principle of non-discrimination must be weighed against the interest in an accurate prediction while fundamental rights and the rule of law may establish a Right to Justification[37].

Furthermore, thinking about such general regulatory problems as technology-specific might relieve regulators from the burden to comprehensively address the problem while, at the same time, creating a "bias against the technological"[38] that can discourage the uptake of the regulated technology[39].

This notwithstanding, in certain cases, a regulatory rationale can be directly linked to a specific technology,[40] e.g. if there is a moral objection to a technology as such or if there are hazards inherent to that technology with the potential to cause adverse effects irrespective of any specific form of use but due to mere exposure to that technology.[41] The White Paper, for example, raises the issue of "changing functionality of AI systems [...] that require frequent software updates or which rely on machine learning", noting that "[t]hese features can give rise to new risks that were not present when the system was placed on the market" which is "not adequately addressed in the existing legislation which predominantly focuses on safety risks present at the time of placing on the market" (p. 14). This, however, does not justify regulation that targets "AI": If anything, it may only constitute grounds for considering technology-specific regulation *stricto sensu*, i.e. of machine learning technology:[42] As a first step to addressing this uncertainty we propose to make use of the following two measures for machine learning:

- regular audits to ensure that development processes in companies comply to certain standards and

[34] Moses, L.B. (2013), in: *Law, Innovation and Technology* 5 (1), p. 16.
[35] See Brenner, S. (2007), *Law in an Era of 'Smart' Technology*; Moses, L.B. (2013), in: *Law, Innovation and Technology* 5 (1), p. 14.
[36] Hinton, G. (2020), in: *twitter,* February 20, 2020, retrieved from https://twitter.com/geoffreyhinton/status/1230592238490615816?s=20 (last downloaded June 14, 2020); see Kahn, J. (2020), The problem with the EU's A.I. strategy, in: *Fortune Newsletters, Eye on AI*, February 25, 2020, retrieved from https://fortune.com/2020/02/25/eu-a-i-whitepaper-eye-on-a-i/ (last downloaded June 14, 2020).
[37] Wischmeyer, T. (2018), in: *Archiv des öffentlichen Rechts* 143 (1), p. 55.
[38] Moses, L.B. (2013), in: *Law, Innovation and Technology* 5 (1), p. 17.
[39] Reed C. (2018), in: *Philosophical Transactions of the Royal Society A*, pp. 274, 278; Moses, L.B. (2013), in: *Law, Innovation and Technology* 5 (1), pp. 15, 17.
[40] Moses, L.B. (2013), in: *Law, Innovation and Technology* 5 (1), p. 15.
[41] See Meyer, S. (2018), in: *Zeitschrift für Rechtspolitik*, p. 234.
[42] See Moses, L.B. (2013), in: *Law, Innovation and Technology* 5 (1), p. 16.

- ex-post compliance benchmarks of machine-learning products and services.

While the first proposal is rather evident, the second needs some explanation. As mentioned above, the changing nature of self-learning systems might lead to pre-market-entry regulation becoming insufficient. Therefore, ex-post compliance benchmarks could be made mandatory in certain cases to ensure post-market-entry compliance. For some software, these benchmarks could maybe even be done in an automated manner, reducing the cost of such regulation.

With regard to context-overarching issues in "AI"-regulation, a comprehensive policy proposal on "AI" is to be welcomed, even though regulation of "AI" is, as shown, not suitable. For the purpose of producing a common understanding of such a proposal's intended scope, the Commission's White Paper should, therefore, be complemented by a definition of its central notion - a definition of AI - without the necessity of it providing the legal certainty that would be required for legislation: Albeit citing the AIHLEG[43], the White Paper simply describes AI as "a collection of technologies that combine data, algorithms and computing power" (p. 2) which is not sufficient or even suitable to distinguish AI from any software[44].

3.2 Objects of Protection: What about Collective Goods?

Regulatory interventions may entail encroachments upon fundamental rights and principles recognized by the Charter of Fundamental Rights of the European Union. In particular and firstly, the freedom of sciences (Art. 13 ChFR) and the economic fundamental rights (Art. 15 ff. ChFR) of those who develop, manufacture, and market AI systems should be taken into account. As much as regulation may result in a delay in innovations that, in turn, could prevent third parties from benefiting from this technology, an infringement of their legal positions comes into question, as well.[45]

Fundamental rights do not constitute unfettered prerogatives and may be restricted, provided that the restrictions correspond to objectives of general interest pursued by the measure in question and that they do not involve, with regard to the objectives pursued, a disproportionate and intolerable interference which infringes upon the very substance of the rights guaranteed.[46] The Commission's proposal lists numerous such legitimate objectives of its outlined regulation:

---

[43] (2019), *A definition of AI*, p. 8.

[44] See Dignum, V. et al. (2020), *First Analysis of the EU Whitepaper on AI*, retrieved from https://allai.nl/first-analysis-of-the-eu-whitepaper-on-ai/ (last downloaded June 14, 2020); of the other opinion are MacCarthy, M./Propp, K. (2020), The EU's White Paper on AI: A Thoughtful and Balanced Way Forward, in: *Lawfareblog*, March 5, 2020, retrieved from https://www.lawfareblog.com/eus-white-paper-ai-thoughtful-and-balanced-way-forward (last downloaded June 14, 2020).

[45] See Meyer, S. (2018), in: *Zeitschrift für Rechtspolitik*, pp. 233 f.

[46] ECJ, ECLI:EU:C:2010:146 (Alassini), 63; ECLI:EU:C:2008:476, 181 (FIAMM); see Trstenjak, V./Breysen, E. (2012), in: *Zeitschrift Europarecht* 3, p. 279.

- the protection of "safety and health of individuals" (p. 10),
- the protection of fundamental rights, "including the rights to freedom of expression, freedom of assembly, human dignity, non-discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, as applicable in certain domains" (p. 11), and
- the "protection of personal data and private life, or the right to an effective judicial remedy and a fair trial, as well as consumer protection" (p. 11).

This variety of objects of protection need not necessarily raise the concern[47] that the prospective legislation might be too vague and indeterminate. It rather suggests that the Commission is aware of the plethora of possible impacts of AI on fundamental rights.

Such an enumeration of specific objects of protection entails the risk of protection gaps, though. In particular, the list of potentially impacted individual rights indicates that the Commission completely overlooked the necessity to safeguard collective goods, as well:[48] Even though it is acknowledged that "the impact of AI systems should be considered not only from an individual perspective but also from the perspective of society as a whole" (p. 2)[49], the White Paper only vaguely states that a regulatory framework also "must ensure socially, environmentally and economically optimal outcomes" (p. 10). While the policy proposal only insufficiently considers the environmental implications of AI, as pointed out in section 1, it does not address or even mention the various specific challenges for democracy[50] and the rule of law [51], prosperity and social justice[52].

For example, often discussed risks for democracy range from cyberattacks on elections that use AI for targeting victims and defeating cyber defences[53] – an AI-enhanced threat to cyber security, in general – over AI-enabled distortion campaigns by micro-targeting and emotionally influencing "what should be a deliberative, private, and thoughtful choice"[54] to disinformation campaigns based on fake news which are made to "seem more realistic or relevant" by AI[55]. The principles of democracy and the rule of law are also challenged by the formative power of a system's architecture: Code is law[56]. In this regard, factual standard-setting by ethics and or industry-wide standards might bypass formalised legislative procedures with specific

[47] See Borutta, Y. et al. (2020), in: *MMR-Aktuell*, p. 427809.
[48] See Borutta, Y. et al. (2020), in: *MMR-Aktuell*, p. 427809.
[49] See Joint Research Centre (2018), *Artificial Intelligence. A European Perspective*, p. 56; Stahl, B.C./Timmermans, J./Flick, C. (2017), in: *Science and Public Policy* 44 (3), pp. 373 ff.
[50] See Manheim, K.M./Kaplan, L. (2019), in: *The Yale Journal of Law & Technology* 21, pp. 133 ff.
[51] See Bayamlioglu, E./Leenes, R. (2018), in: *Law, Innovation and Technology* 10 (2), pp. 295-313.
[52] See Agrawal, A./Gans, J./Goldfarb, A. (2019), *The Economics of Artificial Intelligence. An Agenda*.
[53] Townsend, K. (2016), How Machine Learning Will Help Attackers, in: *Security Week*, November 29, 2016, retrieved from https://www.securityweek.com/how-machine-learning-will-help-attackers (last downloaded June 14, 2020).
[54] Manheim, K.M./Kaplan, L. (2019), in: *The Yale Journal of Law & Technology* 21, p. 138.
[55] Manheim, K.M./Kaplan, L. (2019), in: *The Yale Journal of Law & Technology* 21, p. 145.
[56] Lessing, L. (1999), in: *Harvard Law Review* 113, pp. 501 ff.; see Wischmeyer, T. (2018), in: *Archiv des öffentlichen Rechts* 143 (1), pp. 20 f.

requirements for (democratic) decision-making such as, but not limited to, overcoming conflicts and disputes or ensuring the possibility for participation in public deliberations.

In this regard it is also important to mention that the whitepaper does not address implications for collective security[57], in general, and explicitly excludes "the development and use of AI for military purposes" (p. 1), in particular, without giving reasons for this major restriction. However, the potential of AI to threaten fundamental rights and collective goods is most direct and obvious in military applications such as autonomous lethal weapon systems. It might be politically difficult for the European Commission to make suggestions on this topics due to diverging interests among EU member states, but the attempt of the commision to provide a "coordinated European approach on the human and ethical implications of AI" (p. 1) must remain largely incomplete without considering military applications. For example, the AIHLEG recommends to "[m]onitor and restrict the development of automated lethal weapons, considering not only actual weapons, but also cyber attack tools that can have lethal consequences if deployed"[58].

3.3 Protective Measures: High-Risk AI Applications only

In contrast to other AI strategies, the White Paper does not only acknowledge potentially detrimental impacts of AI but also proposes specific protective measures. The Commission is of the opinion that these would be applicable to "high-risk" AI, only (p. 17). Since risk-based regulatory approaches are less likely to stifle innovation than those solely focusing on a technology's hazards, i.e. on its inherent potential to cause adverse effects,[59] they are common in EU technology legislation.

In view of the Commission, an AI application would be considered "high-risk" if it meets two cumulative criteria:

- employment of the given AI application in "a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur" and
- use of that AI application "in the sector in question [...] in such a manner that significant risks are likely to arise" (p. 17).

The White Paper makes a set of proposals for extensive regulation which such high-risk AI systems would face, as shortly described in the following:

- Training data shall be broad enough "to avoid dangerous situations" (p. 20) and sufficiently representative to avoid prohibited discrimination. Also, privacy and personal

[57] Allen, G./Chan, T. (2017), *Artificial Intelligence and National Security. Belfer Center Paper*.
[58] AIHLEG (2019), *Policy and Investment Recommendations for Trustworthy AI*, p. 40.
[59] See Boyd, I.L. (2019), in: van der Linden, S./Löfstedt, R.E. (eds.), *Risk and Uncertainty in a Post-Truth Society*, p. 69.

data shall be "adequately protected during the use[60] of AI-enabled products and services" (p. 20).

- Records shall be kept of the main characteristics of a training dataset and "in certain justified cases, the data sets themselves" (p. 20). Documentation shall be prescribed on methods used to build, test and validate the AI systems "including where relevant in respect of safety and avoiding bias that could lead to prohibited discrimination" (p. 20).
- Information shall be provisioned on the purpose, operating conditions and accuracy of the AI system. "Citizens should be clearly informed when they are interacting with an AI system and not a human being" (p.21).
- AI systems must be sufficiently robust, accurate, reproducible in their outcomes, and resistant to manipulation attempts.
- The appropriate type and degree of human oversight shall depend on intended use and possible effects of such use on citizens and legal entities.
- Specific requirements exist already for remote biometric identification, e.g. face recognition in public space, with the GDPR and the Charter of Fundamental Rights. "AI can only be used for remote biometric identification purposes where such use is duly justified, proportionate and subject to adequate safeguards" (p. 23), but the Commission announces to launch a debate on when its use is justifiable.

In contrast, AI applications that do not qualify as high-risk would not be subject to these mandatory requirements. The Commission proposes to establish a voluntary labelling scheme, instead (p. 24). This, in the view of two members of the AIHLEG, "deliberately simplistic"[61] distinction between high-risk and low-risk applications based on an enumeration of supposed high-risk sectors threatens to jeopardise the proposed policy's objective: While some fear that "moderately risky AI systems will end up falling into the high-risk bucket and being subjected to onerous and disproportionate requirements", resulting in developers ending up "struggling to innovate [...] or eschewing the EU market altogether"[62], it seems at least equally likely that certain high-risk applications will not cumulatively meet the above mentioned criteria, ending up to be deemed low-risk.

[60] The formulation in the White Paper emphasizes "during the use". This raises the question whether individuals' personal rights will also be adequately protected when their data is compiled for a dataset used for training an AI-system before bringing it on the market, i.e. before the actual use of the system by a customer. The individual whose data is used for training the AI-system can be someone else than the customer using the AI-system afterwards.

[61] Metzinger, T./Coeckelbergh, M. (2020), Und was ist mit der Ethik? Warum das EU-Weißbuch zur Künstlichen Intelligenz enttäuscht, in: *tagesspiegel.de,* April 14, 2020, retrieved from https://www.tagesspiegel.de/politik/und-was-ist-mit-der-ethik-warum-das-eu-weissbuch-zur-kuenstlichen-intelligenz-enttaeuscht/25739396.html (last downloaded June 14, 2020).

[62] Crumpler, W. (2020), Europe's Strategy for AI Regulation, in: *Center for Strategic & International Studies. Technology Policy Blog,* April 21, 2020, retrieved from https://www.csis.org/blogs/technology-policy-blog/europes-strategy-ai-regulation (last downloaded June 14, 2020).

Firstly, attempting to "specifically and exhaustively" list the sectors covered (p. 17) will likely result in not covering risky AI applications in other sectors.[63] Admittedly, the proposal acknowledges that the list "should be periodically reviewed and amended where necessary in function of relevant developments in practice" (p. 17) and that "there may also be exceptional instances where, due to the risks at stake the use of AI applications for certain purposes is to be considered as high-risk as such – that is, irrespective of the sector concerned" (p. 18). Considering the nature of emerging technologies, however, kinds and levels of risk may change as well as the perception of what is to be considered high-risk. Hence, the exception may, very well, become the rule which would compromise the supposed certainty of an enumeration of sectors.

In any case, the criteria of what constitutes a high-risk sector remain unclear. In the absence of any such an explanation and justification, the explicit naming of healthcare, transport, energy and parts of the public sector such as asylum, migration, border controls and judiciary, social security and employment services (p. 17) creates the "general impression that the authors of the White Paper wanted to be tough on some public sectors, while leaving the private sector alone"[64]. We fear that the promised clarity and certainty for providers of AI turns into a "competition of the best lobbying efforts"[65] in the run-up to the adoption of the regulatory framework.

Secondly, the Commission's suggestion that "the level of risk of a given use [of AI systems] could be based on the impact on the affected parties" (p. 17), indicates again that the White Paper is exclusively concerned with issues that could have an impact at individual level while not sufficiently taking into account those that could have an impact at societal level[66].

Either way, due to its lack of nuance, the Commission's proposal raises doubts regarding the proportionality principle. Further, it presupposes that existing legislation is sufficient to protect from most applications, i.e. from all but those that are deemed high-risk. Last but not least, it assumes that even the highest risks can be mitigated since nothing in the White Paper states that there are AI applications that are incompatible with EU values and fundamental rights.[67]

[63] See Borutta, Y et al. (2020), in: *MMR-Aktuell*, p. 427809.
[64] Metzinger, T./Coeckelbergh, M. (2020), Und was ist mit der Ethik? Warum das EU-Weißbuch zur Künstlichen Intelligenz enttäuscht, in: *tagesspiegel.de,* April 14, 2020, retrieved from https://www.tagesspiegel.de/politik/und-was-ist-mit-der-ethik-warum-das-eu-weissbuch-zur-kuenstlichen-intelligenz-enttaeuscht/25739396.html (last downloaded June 14, 2020).
[65] Borutta, Y et al. (2020), in: *MMR-Aktuell*, p. 427809.
[66] See Stahl, B.C./Timmermans, J./Flick, C. (2017), in: *Science and Public Policy* 44 (3), pp. 373 ff.; Joint Research Centre (2018), *Artificial Intelligence. A European Perspective*, p. 56.
[67] See Cath-Speth, C./Kaltheuner, F. (2020), Risking everything: where the EU's white paper on AI falls short, in: *New Statesman*, March 3, 2020, retrieved from https://tech.newstatesman.com/guest-opinion/eu-white-paper-on-artificial-intelligence-falls-short (last downloaded June 14, 2020).

In view of this, we endorse the recommendation of the German Data Ethics Commission (GDEC) to adopt a "risk-adapted regulatory approach"[68]. It distinguishes between five levels of criticality and proposes proportionate measures for each level:[69]

- It is unnecessary to carry out special oversight of or impose special requirements on applications associated with "zero or negligible" potential for harm (Level 1).
- Applications with "some" (Level 2), "regular or significant" (Level 3) or "serious" (Level 4) potential for harm should be subjected to increasingly stringent requirements and more far-reaching interventions by means of regulatory instruments.
- A complete or partial ban should be imposed on applications with an "untenable" potential for harm (Level 5).

In agreement with the GDEC we are of the opinion that the system's potential for harm should be determined on the basis of the likelihood that harm will occur and the severity of that harm.[70] The severity of the harm that could potentially be sustained would, in turn, depend on

- the significance of the legally protected rights and interests affected,
- the level of potential harm suffered by individuals,
- the number of individuals affected,
- the total figure of the harm potentially sustained and
- the overall harm sustained by society as a whole.[71]

Since the GDEC's approach requires in-depth analysis of these determinants, comprehensive AI legislation based on the risk-adapted regulatory approach will take time to be adopted. To not delay regulation where most necessary, the Commission's schematic approach to identifying high-risk AI applications could help prioritise policy-making initiatives. During this transition period, the proposed voluntary labelling for no-high risk AI applications (p. 24) could prove beneficial as an interim solution until a comprehensive regulatory framework is in place.

[68] Datenethikkommission (2019), *Opinion of the Data Ethics Commission*, pp. 173 ff.
[69] Datenethikkommission (2019), *Opinion of the Data Ethics Commission*, pp. 177 ff.
[70] Datenethikkommission (2019), *Opinion of the Data Ethics Commission*, p. 173.
[71] Datenethikkommission (2019), *Opinion of the Data Ethics Commission*, p. 174.

4 Conclusion

In view of the potentially immense impacts of AI on individuals and society, a comprehensive and EU-wide governance framework is necessary and we welcome the Commission's effort in bringing it about.

Considering the policy proposals for promotion of AI within the EU in order to create an "Ecosystem of Excellence", we make several suggestions for improvement regarding the design of promotion programmes, environmental sustainability and public engagement.

Whereas we agree with the Commission that promotion programmes for AI technology in the EU are necessary, we believe it is important to stress that this should only be done with the goal of ethical AI in mind and not as an end to itself. In our view, the only way to ensure that AI services and products are designed to be compliant with European values and EU AI ethics regulation is to have strong European businesses turning innovations in the field of AI into market offerings. the design of promotion programmes should be accompanied by open and transparent sector dialogues which take into consideration the (de-)merit of AI for each context. In general, European values should not be sacrificed in order to speed up the adoption of AI.

Environmental sustainability should be much more built into the proposed governance framework: as a criterion for promoting AI, e.g. in investment funds, and mandatory information for the public, e.g. via energy consumption labels. Furthermore, an expert committee should provide reviews of the environmental impact of governance measures and develop principles and programmes for regulation and public promotion.

The white Paper fails to adequately stress the importance of learning and empowerment of the public at large. Public engagement with AI is necessary to make sure that AI will be aligned to society's values and helps to establish the public's trust that is necessary for AI uptake. We recommend promotion of public engagement with and a public dialog on AI technology, e.g. via citizen training offers in the planned digital innovation labs and public dialog events that bring important stakeholders together.

Considering the detrimental potential of AI, a regulatory framework is needed to protect individual rights and collective interests while providing legal certainty for providers and users.

Legislation should, however, not attempt technology-specific regulation of "AI": Firstly, the notion refers to various techniques, capabilities, applications, and uses; any legal nomenclature would, hence, need to be highly inclusive, probably at the expense of its suitability for differentiation and legal certainty. Secondly, technology-specific legislation, e.g. of machine-learning technology, is suitable only if the regulatory rationale is closely tied to the technology itself; otherwise, it can result in duplicative and compartmentalised laws.

The policy proposal and any future regulation should take into account the necessity to not only protect individual rights but also to safeguard collective goods. An exhaustive enumeration of specific objectives of protections is unsuitable since it entails the risk of protection gaps.

The diverse AI systems and applications pose different kinds and levels of risk for individuals and society. We, thus, agree with the GDEC that comprehensive AI regulation requires differentiated rules and increasingly far-reaching interventions, depending on the application's potential for harm and ranging from no special requirements to a complete or partial ban. The Commission should, therefore, reconsider its binary approach to only subject certain high-risk applications to its outlined regulation.