Response to the public consultation on the White Paper: On Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final

Jens-Peter Schneider, Professor of Public Law at the University of Freiburg, Germany, jp.schneider@jura.uni-freiburg.de, Reporter of the ELI project 'Artificial Intelligence (AI) and Public Administration – Developing Impact Assessments and Public Participation for Digital Democracy' together with Co-Reporters Marc Clément and Paul Craig

Christiane Wendehorst, Professor of Private Law at the University of Vienna, Austria, Christiane.Wendehorst@univie.ac.at,
European Reporter of the ALI-ELI project 'Principles for a Data Economy'

The present response does not reflect the official view of the European Law Institute or of any other body the authors may be associated with.

---

1. Introduction and General Remarks

The authors are Reporters on two **projects of the European Law Institute (ELI)**.[1] Jens-Peter Schneider, Marc Clément and Paul Craig are the Reporters of an ELI project on Artificial Intelligence (AI) and Public Administration,[2] which recently started to draft model-rules. The aim is to provide a basis for developing European legislation on artificial intelligence in the context of public administration, which will not hinder innovation, while providing solid safeguards to improve citizens' confidence in the use of the technology in this field. They have focussed on the public sector perspective (point 2). Christiane Wendehorst is the European Reporter on a transnational project on the Data Economy, which is conducted together with the American Law Institute (ALI), and has focussed on the private sector perspective (point 3).[3] The authors would like to take the opportunity to respond jointly to the public consultation of the European Commission on the White Paper 'On Artificial Intelligence'.

The authors wish to express their **full support for both an 'ecosystem of excellence' and an 'ecosystem of trust'** as has been proposed by the European Commission, and they welcome the balanced approach of the White Paper taking into account the "twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology" (p. 1). Consequently, the legal framework for AI must serve the enabling function of the law as well as its function to establish safeguards against unjustified intrusions into citizen´s rights.

In their capacity as experts in the field of law, and against the background of the nature of the projects they are conducting, jointly with others, on behalf of the ELI, they will **focus their response on the 'ecosystem of trust'**, and in particular on the legal aspects.

---

[1] https://www.europeanlawinstitute.eu/.
[2] https://europeanlawinstitute.eu/projects-publications/current-projects-feasibility-studies-and-other-activities/current-projects/ai-and-public-administration/
[3] http://www.thealiadviser.org/data-economy/.

2. Regulating public use of AI applications: challenges and options for a legal framework with safeguards and incentives

Processing of data and information is fundamental to the work of public administrations. New technologies, such as artificial intelligence (AI), can play a significant role in the modernisation and overall improvement of the functioning of public administration. On the other hand, it is equally important to guarantee the transparency, correctness and security of the processed data, as well as to ensure the respect of the rule of law and especially the rights of European citizens and businesses.

Public administration is, as a result, confronted with specific opportunities and challenges in the deployment of AI and, more generally, the deployment of automated decision-making systems, whether or not they use specific AI technologies (such as machine learning). The use of these techniques poses specific problems related to the particular requirements associated with the principle of good administration. However, AI might also be an instrument for enhancing the quality of public services, or compensating for constraints of human resources that European administrations face due to demographic developments.

The twin objectives of a European AI strategy raise challenges for the respective legal framework as those objectives can conflict with each other. AI innovations need to be socially acceptable and a solid set of guarantees for individual rights and conflicting public interests is essential for building a trustworthy AI ecosystem. The European debate on Covid-19 tracking or tracing applications indicates how important public trust is for an effective implementation of socially highly important technologies – and how easily this public trust can be undermined. That being said, the 'ecosystem of trust' as identified in the AI White Paper must provide effective legal instruments and procedures to guarantee the protection of individual rights and public values.

2.1. Lack of knowledge about public use of AI tools

A very recent report by an interdisciplinary team of scholars from Stanford University and New York University submitted to the Administrative Conference of the United States (ACUS) on "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies"[4] highlights that US administrative agencies already use a diverse set of AI tools across the full range of government tasks including:

- law enforcement,
- single-case decision-making,
- monitoring and analysing risks to public health and safety or other policy objectives,
- extracting information from the government´s data resources including statements in administrative multi-party consultations, consumer complaints as well as environmental data,
- communicating with citizens and business about their legal rights or obligations as well as about various other matters of interest, for instance by using chatbots, and
- intra-administrative management of resources including procurement and maintenance of public facilities.

[4] The study can be downloaded from two websites: https://law.stanford.edu/education/only-at-sls/law-policy-lab/practicums-2018-2019/administering-by-algorithm-artificial-intelligence-in-the-regulatory-state/acus-report-for-administering-by-algorithm-artificial-intelligence-in-the-regulatory-state/; See also: https://www.acus.gov/newsroom/news/acus-stanford-law-school-and-nyu-school-law-announce-report-artificial-intelligence

Unfortunately, no empirical study similar in scope and depth exists with regard to the actual use of AI tools by EU authorities, or by Member States authorities.[5] Nevertheless, there are clear indications that some European authorities have implemented AI tools or are planning to do so. Therefore, Action 6 of the White Paper is a step in the right direction, but in comparison to the situation in the US it is too limited in its reach. The fields of actual or potential application of AI technologies by public authorities is very broad, and should be captured by future EU initiatives although concentration on certain key areas, as experimental sandboxes might be helpful to gain new transversal insights and momentum.

2.2. The risk of widening the public-private technology gap

The ACUS report indicates AI has the potential to widen, not narrow, the public-private technology gap. In their in-depth case studies on top US Federal Agencies Stanford computer scientists rated public AI technologies – outside the military and secret services – far less sophisticated compared to state of the art technologies for private companies. There is good reason to believe that the situation in Europe is not different. For instance, the Covid-19 pandemic has revealed deficiencies in many national administrative systems within the EU, even with regard to basic forms of digitalization and e-government. The recent Digital Economy and Society Index (DESI) 2020 prepared by the European Commission also supports this presumption.[6] Thus, we can imagine that the situation concerning advanced forms of AI technology is not any better.

Interestingly, the ACUS report does not advocate increased contracting-out to private service providers as a promising countermeasure. Instead, the report promotes a nuanced approach combining external contracting with regard to technically complex but standardized administrative tasks on the one hand, with developing internal capacities, or non-commercial collaborations, to provide AI solutions for administrative tasks with high relevance for public values on the other hand. Consequently, Action 5 of the White Paper concerning public private partnerships in AI projects should take into account that this approach might not be suitable in all fields of public AI. In a relevant number of cases, public services cannot rely simply on standard AI products and customize them. While Public-Private-Partnerships should certainly not be excluded, they need to be complemented by enhancing human and technology AI resources within the public service. It is a strategic challenge to European policy makers to identify criteria for rational make-or-buy decisions concerning the development and implementation of AI technologies.

2.3. Impact Assessments as a source of trustworthy public AI

The scholarly debate in the US shows the **need for an organizational and legal framework adapted to new challenges** accompanied with any implementation of public AI tools in order to accomplish trustworthiness of public AI in accordance with the twin objective of the White Paper.

---

[5] Nevertheless there exist some interesting studies with empirical information: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/001-148_AW_EU-ADMreport_2801_2.pdf; https://www.legiscompare.fr/web/Activites-de-la-section-921; http://www.aca-europe.eu/index.php/en/evenements-en/637-the-hague-14-may-2018-colloquium-an-exploration-of-technology-and-the-law (for the detailed material see http://www.aca-europe.eu/index.php/en/colloques-top-en); other excellent legal studies of European scholars focus mainly on various doctrinal aspects of AI but are not aimed to provide comprehensive empirical insights on the use of AI and even less so on such usage by public administrations: Thomas Wischmeyer/Timo Rademacher (eds.) Regulating Artificial Intelligence, Springer 2020; Karen Yeung/Martin Lodge (eds.), Algorithmic Regulation, OUP 2019; Woodrow Barfield/Ugo Pagallo (eds.), Research Handbook on the Law of Artificial Intelligence, Edward Elgar 2018.

[6] https://ec.europa.eu/digital-single-market/en/desi.

A **major component** of such a legal framework could be a **rigorous as well as proactive AI Impact Assessment**. Therefore, the ELI has set up a group of European experts to draft model rules establishing such a regulatory instrument. An AI impact assessment would expand the more focused impact assessment provided by Article 35 GDPR to new dimensions of assessment. Another source of inspiration is the Canadian Algorithmic Impact Assessment.[7] It is important that the EU legal framework for an AI impact assessment cover preconditions for a successful implementation of AI tools. Consequently, the ELI model rules will provide safeguards for a comprehensive assessment of factors for successful implementation of public AI tools, as well as for identifying and managing real risks for public interests or individual rights accompanied with AI usage by public administrations. Up to now the debates in public, political and scholarly domains remain rather abstract and tend to focus on – or even overrate – either risks or promises of public purpose AI. A legal framework for Impact Assessments of public AI needs to take these discussions into account. However, it should also provide incentives (1.) to conduct a rigorous while balanced assessment of real and concrete risks connected with a concrete AI tool in question and (2.) to search proactively for options to exploit gains for public services by usage of AI while mitigating the risks associated with such AI usage.

Obvious as well as less obvious – but not less important – **issues to be evaluated** in an impact assessment of public AI include,

- questions of meaningful transparency,
- compliance with concrete hearing rights and accountability mechanisms applicable in the concrete legislative context of implementing AI or options for adapting these traditional administrative law instruments to the new challenges of public AI,
- the avoidance of gaming by regulatory subjects who might be able to hire computer scientists who can reverse-engineer an agency´s AI model in order to manipulate administrative decision-making,
- various appearances of discrimination or bias,
- potential over-reliance of human decision-makers on output of AI systems, as well as under-reliance by human field experts who – sometimes falsely – do not trust the accuracy and efficiency gains to be delivered through AI systems.

In order to establish a transparent and reliable assessment regime, the ELI project team seeks to organize the **standards for evaluating AI** tools more consistently than recent proposals. For instance, the principles and key requirements of the High-Level Expert Group on AI, which are a basis also for the White Paper, are a very good starting point of discussion. Nevertheless, some of them are not very precise, definitions of some requirements include aspects not covered by the literal meaning of the requirement, in some cases requirements are overlapping, and some conflict with each other. While such features might be suitable or acceptable in broader ethical deliberations or policy papers, a legal text should avoid them as far as possible, or organize transparent procedural and organisational solutions for conflict resolution.

The authors support the analysis in the White Paper concerning risks for fundamental rights. In line with Article 35 GDPR and the White Paper, the drafters of the ELI Model Rules on AI Impact Assessment are contemplating a **risk-based approach**. Nevertheless, impact assessments and other features of a future EU legal framework for AI should not only serve a risk management objective. Instead, such legal instruments should **equally foster learning and the exploitation of advantages** for the public and the individuals by public use of AI tools. Among those advantages are,

- the exoneration of administrative staff from time consuming and exhausting standard tasks,
- optimized case allocation and information gathering,
- safeguards against simple errors of overloaded human decision-makers,

[7] https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

- provisions for more equal treatment.

Consequently, the impact assessment should provide insights not only for risk minimization, but for an optimization of efficiency and effective implementation of the public function served by an AI tool as well.

In order to do so further research is needed concerning mechanisms that provide in addition to privacy by design also **"accountability by design"**. The latter phrase has been emphasized in the above-mentioned ACUS report[8] and covers new (additional) structural instruments incorporated into AI systems. For instance, collaborative forms of governance, or agency internal mechanisms of accountability, may compensate growing dysfunctions of traditional accountability mechanisms based on individual (privacy) rights. Both mechanisms would complement, not substitute each other. Interesting options discussed in the mentioned ACUS report enclose AI experimental usage in "supervised sandboxes", or continuous "benchmarking" of AI tools by reserving a random hold-out sample of cases for human decision, thus providing critical information when an algorithm has gone astray, or "automation bias" has led human decision-makers to excessively defer to an algorithm.

Important procedural questions discussed by the ELI drafting team concern,

- the body to be responsible for conducting the impact assessment either internally or externally of the public body implementing an AI tool,
- the integration of expert knowledge into the assessment process,
- options for public participation,
- protection of business, private or administrative secrets,
- need of iterative assessments due to the evolving nature of self-learning AI technologies,
- the exact relationship with the data protection assessment according to Article 35 GDPR, which covers many aspects of a comprehensive AI impact assessments, but also leaves ample gaps in proactive steering of implementing AI technology – especially with regard to effective and efficient fulfilment of public tasks and duties.

3. Regulating private use of AI applications: Squaring the circle of a high level of protection that avoids too much red tape

The ambit of the White Paper goes far beyond the public sector and extends to private use of AI applications. In its quest to square the circle of achieving a high level of protection while avoiding uncertainty and a climate that is hostile towards innovation, the White Paper (p. 17) suggests restricting the scope of the future regulatory framework for AI. Whether an AI application is included should, in the first place, be determined by the 'sector' that is concerned. The 'sectors', which should be exhaustively listed, could be areas such as healthcare, transport or energy. In the second place, and within the listed sectors, the risk posed by a particular application needs to exceed a particular threshold. Only if both criteria are cumulatively met will an AI application normally be subject to the new regulatory framework.

While the authors fully share the concern that an overreaching regulatory approach might undermine Europe's position as the place for R&D and an 'ecosystem of excellence' they are **not convinced the sectoral approach as it is currently described in the White Paper will achieve the desired effects**.

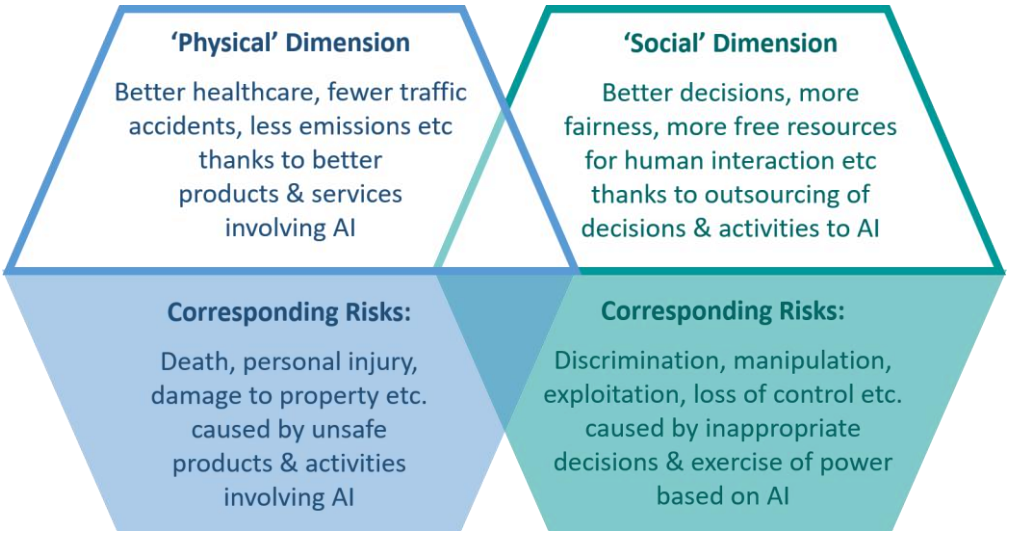
[8] The ACUS report quotes in this respect a wording of Margot Kaminsky, Binary Governance: Lesson's from GDPR's Approach to Algorithmic Accountability, 92 S. Cal. L. Rev. (2019), 1529, 1557 note 118.

3.1. The 'physical' and the 'social' dimension of AI applications

The challenges posed by AI and modern digital ecosystems in general – such as opacity ('black box-effect'), complexity, and partially 'autonomous' and unpredictable behaviour – are similar, irrespective of where and how AI is deployed. However, at a somewhat lower level of abstraction, and closer to what regulators might actually wish to address, the potential risks associated with AI appear as normally falling into either of two dimensions:

A. 'Physical' risks, i.e. death, personal injury, damage to property etc. caused by unsafe products and activities involving AI; and

B. 'Social' risks, i.e. discrimination, total surveillance, manipulation, exploitation etc. and general loss of control caused by inappropriate decisions made with the help of AI or otherwise inappropriate deployment of AI.



The first of the two dimensions (**'physical'**) includes concerns such as the safety of self-driving vehicles, drones or medical and care robots and of liability where damage has been caused by the operation of such devices. While the opacity, complexity, unpredictability etc. that comes with AI may certainly add to the risk created by such devices, the risk as such is of a very traditional type. In other words, other technical developments could have a similar effect on 'physical' risk, and the effect on a victim may be the same, irrespective of whether the traffic accident that killed them was caused by AI, or by any other vehicle component.

The second of the two dimensions (**'social'**) is much **more AI-specific**. This is about AI used for purposes such as HR decisions, personalised pricing, personalised news feeds, or predictive policing. While discrimination, manipulation or exploitation by humans as well as loss of control is as old as mankind itself, the situation with AI is unprecedented in various respects. Human brains are confronted with 'super-human' computing capacities, resulting in a potential inequality of arms and general loss of control. Also, injustice or incompetence is hidden in ostensibly 'neutral' and hardly explainable code, which is copied and built upon many times, meaning we may lose the capacity of detecting and correcting the problems. On the other hand, we may, for the first time, be in a situation to actually analyse, control and optimise decision making procedures, as controlling and optimising machines does not get into conflict with human rights as might the attempt to fully control and 'optimise' human decision-makers.

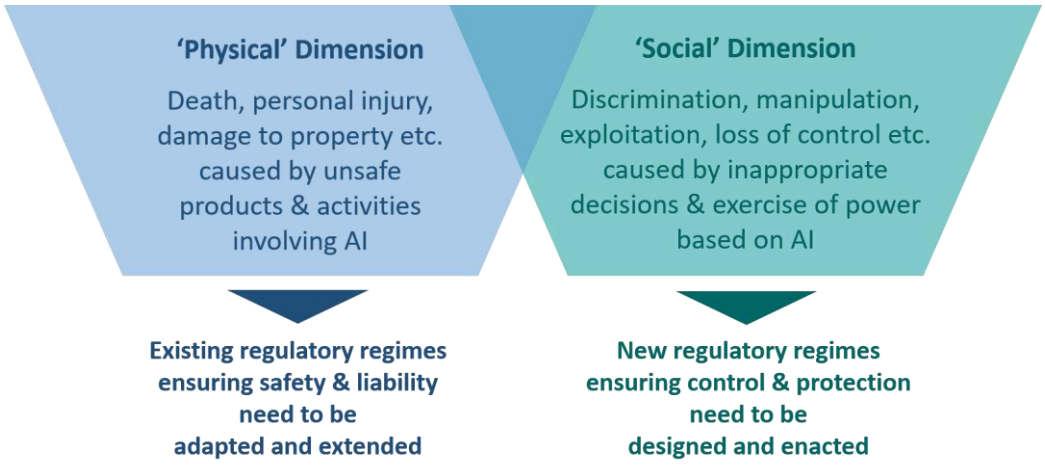
AI-driven devices usually pose challenges in both the 'physical' dimension and the 'social' dimension. For instance, medical decision support systems pose both a traditional risk of causing personal injury

(and are thus included in the Medical Device Regulation), but by their very nature they may also change the doctor-patient-relationship. AI used in the workplace may lead to better decisions, but the employees may feel deprived of their functions and suffer mental health problems. Also, the delineation between the two dimensions themselves is not always clear-cut, and there is a degree of overlap. However, for **most AI applications the clear focus is on one of the two dimensions.**

3.2. A targeted regulatory approach to the 'physical' dimension

This insight has important implications for a regulatory response to AI that is targeted, i.e. addresses problems as directly as possible and avoids the extension of measures to cases where the measures are not necessary and thus disproportionate.

The 'physical' dimension of risks is best addressed by way of a 'digital fitness check' of existing regulatory regimes[9], such as the Product Safety Directive, the Product Liability Directive, and sectoral instruments as well as national law, and not by a new regulatory framework for AI as it seems to transpire from the White Paper (p. 18 ff).



This 'digital fitness check' must not be restricted to considering AI, but should more generally consider digital ecosystems, including software as such, the development towards 'everything as a service', and interconnectedness in the IoT. There will certainly be some few AI-specific measures, such as:

- Development of technical standards for the safety and assessment of self-learning systems;
- Extension of the scope of strict liability regimes, or introduction of new strict liability regimes, where AI makes the risk involved rise above the critical threshold (e.g. slow vehicles, such as big cleaning robots in public spaces, may suddenly pose a risk comparable to the risk that used to be posed only by fast vehicles);
- Re-considering of elements clearly connected with human agency (e.g. extension of vicarious liability for human auxiliaries to harm caused by AI that replaces human auxiliaries)

However, the majority of measures addressing the 'physical dimension' are likely to become necessary because of **digital developments other than AI**.[10] Conversely, it may be **disproportionate to extend**

[9] For details see the Report of the Expert Group on Liability and New Technologies – New Technologies Formation, Liability for Artificial Intelligence and other emerging digital technologies, 2019, https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608

[10] For example, the fact that modern 'products' may rather be software, or digital services, and that they receive updates and depend on data feeds, means that the scope of product liability law may need to be extended, and that the relevant point in time can no longer be the one point in time when the product was put into circulation.

measures addressing the 'social dimension' (such as information duties, rights to receive explanations, full inclusiveness of training data etc.) to AI applications that largely pose very traditional 'physical' risks. Put bluntly: We do not primarily need manufacturers of autonomous cars to explain to us in a transparent manner whom they prefer to kill, and to prove that the ratio of individuals with light skin colour killed matches exactly the ratio of those individuals in the overall population. We primarily need them to develop autonomous cars that are safe and that will not kill anybody – full stop (and when checking the safety in type approval procedures we will, of course, routinely check whether the sensors work reliably without regard to skin colour etc. and will deny approval if they do not). Even if we rightly point to the fact that sensitive choices are made when designing AI, the same holds true for many other design components (cf. the heavy gender and age bias that may be inherent in the size of car seats and the positioning of air bags).

It follows from these considerations that, as far as 'physical' risks in sectors such as healthcare, transport, and energy (White Paper p. 17) are concerned, they should rather not be subject to the new regulatory framework for AI, but to more traditional frameworks. These frameworks need to be fully adapted to the challenges posed by digital ecosystems, including AI.

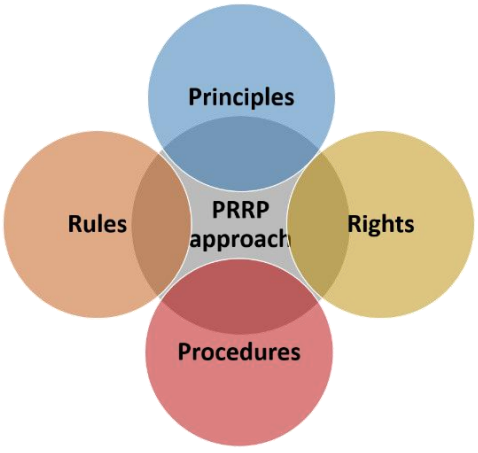
3.3. A targeted regulatory approach to the 'social' dimension

The 'social' dimension of risks is much more AI-specific, and much more difficult to address. This is where all the AI-specific regulatory components, such as ensuring inclusiveness of training data, ensuring that decisions are explainable, information duties, impact assessment, human oversight etc. are fully justified. The dilemma faced by a regulator is that between ensuring a sufficient level of protection across the board, without any significant gaps or loopholes, and avoiding too much uncertainty and/or red tape. In the White Paper, it is proposed to achieve this by reducing the scope of (most of) the new regulatory framework to particular 'sectors' and applications.

A reduction of the scope to an exhaustive list of sectors and applications is only one out of several **regulatory techniques** that serve to avoid too much uncertainty and red tape. Other such regulatory techniques include:

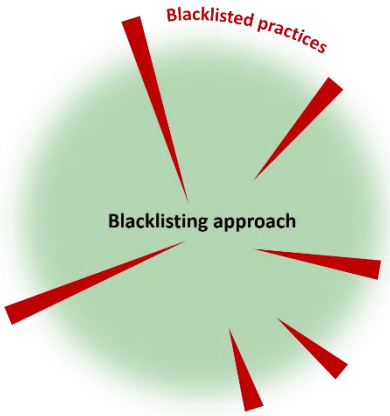
- A smart mix of **regulation, co-regulation and self-regulation**; and
- **Blacklisting** in lieu of mandatory requirements

While the first of the two techniques mentioned is well known, has successfully been applied by the European legislator for quite some time and is referred to also in the White Paper (e.g. voluntary labelling of trustworthy AI), the second technique may require some explanation.



When regulating a particular area the European legislator may formulate mandatory requirements that need to be fulfilled for some activity to be lawful. Normally, these mandatory requirements take the form of principles (e.g. the principles for the processing of personal data in Art 5 GDPR), rules (e.g. the requirement of a legal ground under Art 6 GDPR), rights (e.g. the data subject's rights), and procedures (e.g. a data protection impact assessment, documentation). This **'PRRP approach'**, which has been taken, for example, by the GDPR, has a number of benefits, including: (i) it sends a positive message and is of high symbolic value; and (ii) it is relatively straightforward to

formulate (e.g. it does not take much ingenuity to add information or documentation duties on top of existing lists). The major downside, however, is that it tends to put up a lot of red tape and to mean a lot of extra bureaucracy and costs, potentially being to the detriment of SMEs and enhancing the competitive advantages and market power of the big players.



An alternative to this PRRP approach is **blacklisting**, which is often combined with a general clause. This second regulatory technique has successfully been applied, e.g., for unfair contract terms control in consumer contracts (Directive 93/13/EC) and unfair commercial practices (Directive 2005/29/EC). Blacklisting means the regulator mainly restricts itself to stating what should definitely NOT be done, and possibly combining this with a fall-back option for similar cases in order to prevent circumvention and obvious gaps. The main drawbacks of this approach are: (i) it tends to send a 'negative message', i.e. is much more difficult to defend from a PR perspective; and (ii) it is relatively difficult to get the formulation of the blacklisted practices right, in particular for sensitive and 'fuzzy' areas, such as discrimination or manipulation. A major advantage of this approach, however, is that it hits in a much more targeted manner precisely what we all agree we want to avoid because of its inconsistency with fundamental European values, and leaves full freedom otherwise. It is also much easier to adapt to changing developments. This approach may be much more beneficial for innovation, in particular by SMEs.

The European legislator may thus wish to consider whether – at least to a certain extent – an **'unfair AI practices approach'** might be better for an innovation friendly environment in Europe than a comprehensive PRRP framework.

At the end of the day, what might be the most beneficial solution is a **combination** of the following:

1. A set of fully horizontal principles whose main function would, however, be to confirm basic European values and to guide the interpretation and application of the whole framework (see, e.g., the 7 key requirements formulated by the HLEG on AI[11] or the 9 general standards identified by the German Data Ethics Commission[12])

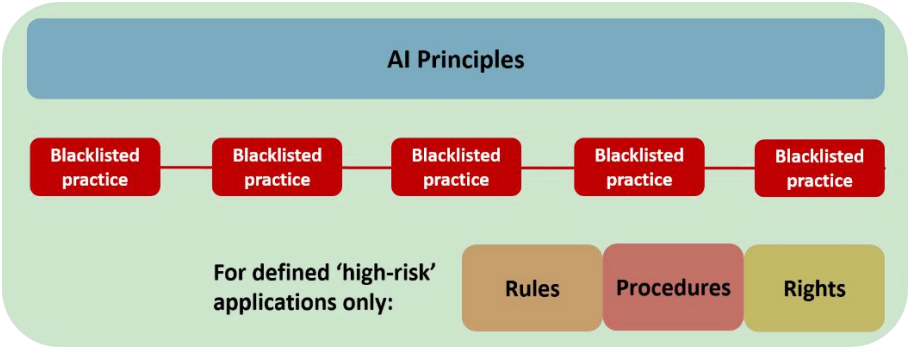
2. A list of blacklisted AI practices, targeting effects we want to avoid in any case, such as discrimination (taking into account that AI may not discriminate along the same criteria as humans do), exploitation of vulnerabilities, profiling and scoring with regard to most intimate human characteristics, total surveillance, manipulation, etc.

3. For defined high-risk applications only: a comprehensive regulatory framework that includes rules, procedures and rights of the types mentioned in the White Paper (p. 18 ff.).

[11] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 2019, p. 14 ff., https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[12] Opinion of the German Data Ethics Commission, 2019, p. 163 ff. https://datenethikkommission.de/.

This could be visualised as follows:



Contrary to what the White Paper seems to suggest, however, the 'high-risk' applications in the private field should not be identified on the basis of the 'sector' concerned (such as healthcare, transport, and energy), but **exclusively on the basis of the nature of the application and its inherent risks**. Assessment of this risk should be guided by a set of criteria as has been formulated, for instance, by the German Data Ethics Commission[13]. The 'high-risk' applications would include applications such as

- Human resources management (recruitment, evaluation of employees etc)
- Allocation of education places
- Credit scoring
- Personal pricing
- …

Measures taken with regard to such applications would have to follow the proportionality principle, i.e. be adjusted to the level of risk posed by the relevant application at hand. As long as the proportionality principle is observed it is not decisive whether this is done by way of reference to different 'risk levels', as is the case under the Medical Device Regulation and has been proposed by the German Data Ethics Commission[14], or on a more flexible basis.

[13] German Data Ethics Commission (n. 12) p. 173 ff.
[14] German Data Ethics Commission (n. 12) p. 177.