Part III: Digital future – unclear prospects and unknown challenges

Christiane Trüe[1]

The EU AI Act Regulation – AI use in innovative public administration?

Summary: Public administration may profit from AI in appropriate cases, provided its use respects fundamental rights and abides by the rule of law in general. The EU AI Act may help in developing AI into the right direction, by ensuring that only 'trustworthy' AI is available on the market, and hence to public administration. The AI Act, as a first attempt at regulating AI, follows a risk-based approach, leaving considerable substantive leeway for developing AI further, but imposing onerous, mostly procedural, risk and quality management obligations on providers, and also on deployers, such as public administration. Legally the AI Act appears coherent, however, its effectiveness as a hardly legible legal document gives rise to doubts. Overall it is a first step which should also enable public administration to use AI, with the relevant precautions taken, namely a well-organised human oversight. The AI Act will be kept under review and become subject to amendments in due course, based on experience gathered and documented according the its current provisions.

Keywords: AI Act, public administration, trustworthy AI, fundamental rights, rule of law, risk based approach vs. substantive regulation, risk management, conformity assessment, documentation obligations

A. Introduction

Public administration should serve the citizens, provide good public services, maintain public security, promote sustainable development, and generally set a good framework for people and business to thrive. This requires modern, innovative administration, also using modern technology in public

1 Prof. Dr. Christiane Trüe LL.M. is Professor for Public Law at the University of Applied Sciences Bremen, Germany. She has published monographs, articles and contributions in the area of her denomination, in particular on EU law, constitutional and administrative law and environmental law, often with an international or comparative perspective. At the moment she is working on climate change law and EU internal market law including the AI Act. Contact: Christiane.Truee@HS-Bremen.de

administration. Modern administration may include the use of Artificial Intelligence (AI), which has been used elsewhere already, making public services more efficient, but also with at times disastrous consequences ending in suicides committed due to decision-making based on biased AI-supported risk assessment.[2] In order to curb and control risk, a legal framework is usually one of the means of choice. The AI Act Regulation of the EU[3] ('AI Act' in the following) has been passed after long debate and various amendments, and is, together with more general legislation, such as the General Data Protection Regulation (GDPR)[4], set to provide the legal framework for AI use in the EU in years to come. Its aim is to promote 'trustworthy' AI, and regulate low-risk, medium-risk and high-risk AI whilst prohibiting AI involving a risk deemed 'unacceptable'. This appears to be the way forward to profit from the benefits and avoid the risks of AI. In the words of EU Commission President Ursula von der Leyen,

> *'We all know that Artificial Intelligence can do amazing things. And I think we do not talk enough about what Artificial Intelligence is able to do to improve our daily lives. […] We want citizens to trust the new technology. And technology is always neutral, it depends on what we make with it. And therefore, we want the application of these new technologies to deserve the trust of our citizens. This is why we are promoting a responsible, human centric approach to Artificial Intelligence.'*[5]

And on a later occasion:

> *'AI is a general technology that is accessible, powerful and adaptable for a vast range of uses – both civilian and military. And it is moving faster than even its developers*

2 Netherlands case of alleged child benefit frauds (2021), e.g. reported on politico https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms.

3 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

4 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119/1. A list of EU legislation relevant to High-risk AI, such as product safety regulations and directives, is accumulated in Annex I of the AI Act, referred to in Art. 8 (2) of the AI Act.

5 Press remarks by President von der Leyen on the Commission's new strategy: Shaping Europe's Digital Future, Brussels, 19 February 2020, https://ec.europa.eu/commission/presscorner/api/files/document/print/nl/speech_20_294/SPEECH_20_294_EN.pdf.

> *anticipated. So we have a narrowing window of opportunity to guide this technology responsibly. I believe Europe, together with partners, should lead the way on a new global framework for AI, built on three pillars: guardrails, governance and guiding innovation.*'[6]

Bill Gates appears to confirm this view and also calls for rules:

> *'This new technology can help people everywhere improve their lives. At the same time, the world needs to establish the rules of the road so that any downsides of artificial intelligence are far outweighed by its benefits, and so that everyone can enjoy those benefits no matter where they live or how much money they have.*'[7]

Coming from a civil society side, it is argued that

> *'increasing power imbalances arise from the fact that the ability to deploy and benefit from large AI systems is concentrated in the hands of a couple of (mostly private) organisations. This leads to issues of democratic accountability and the deepening of societal and economic divides.*'[8]

In using AI in its provision of public services, public administration may also have a solution for becoming more effective, and coping with demographic change, in order to provide public services with far fewer civil servants. However, AI may involve a loss of human control and be a step towards government by computers, as depicted in the 'Terminator' movies and the like. From the public administration perspective, the most relevant question is whether the AI Act makes AI sufficiently trustworthy to be used in public administration. In the EU, this must involve an administration governed by the rule of law, i.e. respecting human rights, administrative law, be it general procedural law or administrative law regarding specific areas, such as data protection law, product safety and consumer protection law, planning and building law etc. After all, public administration must not circumvent fundamental rights in providing public services and exercising public authority by using AI, since these are directly binding on the EU in the shape of the Fundamental Rights Charter of the EU (FRC) and on the Member States under their constitutions. At the same time, the EU will need to take care not to stifle innovation by setting too tight a framework

---

6 2023 State of the Union Address by President von der Leyen, Strasbourg, 13 September 2023, https://ec.europa.eu/commission/presscorner/api/files/document/print/ov/speech_23_4426/SPEECH_23_4426_OV.pdf.

7 Bill Gates, The Bill Gates Blog, A new era – The Age of AI has begun, 21st March 2023, last paragraph, https://www.gatesnotes.com/The-Age-of-AI-Has-Begun.

8 Dignum, Virginia, Progressive Post, 15/06/2023: Future-proofing AI: regulation for innovation, human rights and societal progress, https://feps-europe.eu/future-proofing-ai-regulation-for-innovation-human-rights-and-societal-progress, para 4.

for AI use: the AI Act thus does not specifically regulate AI in public administration, but AI provision and use in general.

This contribution will provide a first attempt at understanding the EU AI Act, first, its aims (B.) and second, concept and definition of AI covered by the AI Act (C.). This is followed by a brief outline of potential or actual uses of AI in public administration (D.) and threats and risks as well as benefits and opportunities of AI (E.). Following this the main rules of the AI Act will be presented (F.), attempting a first prognosis as to whether it provides for trustworthy AI which admits AI use in public administration, identifying regulatory gaps and shortcomings and, finally, offering some conclusions (G.).

B. Aims of the AI Act

The legislative process regarding the AI Act began with the Commission proposal[9] presented on 21 April 2021: the Commission's professed aims in presenting its proposal were to 'guarantee the safety and fundamental rights of people and businesses, while strengthening investment and innovation across EU countries'. Relevant fundamental rights under the Fundamental Rights Charter of the EU (FRC) include, as stated in the Commission proposal[10], protection of personal data and privacy (Art. 7/8 FRC), intellectual property rights (Art. 17 (2) FRC), and non-discrimination, in particular on grounds such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation (Art. 21 FRC) or gender equality (Art. 23 FRC).

Opinions of the Economic and Social Committee[11] and the Committee of the Regions[12] as well as of the European Central Bank[13] were sought. Taking these on board the Council came to a Common Position[14], stressing

9 Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts, 21 April 2021 (COM(2021) 206 final).

10 Commission proposal (fn. 9), p. 11/section 3.5. Fundamental rights.

11 22 September 2022, OJ C 517, 22.12.2021, p. 56.

12 2nd December 2022, OJ C 97, 28.2.2022, p. 60.

13 29 December 2022, OJ C 115, 11.3.2022, p. 5

14 Council of the European Union, 29 November 2021, 14278/21, 2021/0106(COD); final decision of the Council of 21 May 2024, https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/.

the will of the Member States to promote secure AI, respecting fundamental rights. This was followed by a Position of the European Parliament[15] with amendment proposals, and, after a trilogue within the legislative procedure, the AI Act was finally adopted on 13 June 2024, to apply generally from 2 August 2026, but Chapters I and II to apply already from 2 February 2025 (General provisions and provisions on prohibited AI systems); Chapter III Section 4, Chapter V, Chapter VII and Chapter XII and Article 78 shall apply from 2 August 2025, with the exception of Article 101, and Article 6(1) and the corresponding obligations in this Regulation shall apply from 2 August 2027 (Art. 113 AI Act). The AI Act's entering into full force is thus staggered according to urgency and practicality e.g. regarding the set-up of authorities.

This AI Act is the first attempt at setting an AI legal framework worldwide, aimed at setting a legal framework for the EU Internal Market, avoiding fragmentation of the internal market by a joint set of rules rather than individual AI statutes of the Member States, on the legal bases for internal market legislation (Art. 114 TFEU) and data protection (Art. 16).

C. Concept and Definition

First, it is important to define what is to be understood by 'AI'. In fact, we use AI in our everyday lives quite a lot, often without realizing it. In the context of the AI Act this definition is important from a legal perspective, as the AI definition will decide the scope of application of the regulation. The AI Act provides its own definition in its Art. 3:

> *'Article 3 Definitions*
> *For the purposes of this Regulation, the following definitions apply:*
> (1) *'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it re-*

15 European Parliament, P9_TA(2023)0236, Artificial Intelligence Act, Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), 14 June 2023; final position of the European Parliament of 13 March 2024, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html#ref_2_1.

ceives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.
(2) *…'*

At the draft stage of Art. 3 (1)[16], the Commission had proposed a more detailed and specific approach, providing more legal certainty on the one hand, but raising concerns regarding potential narrowness of definition and potential gaps, with an annex I to bring more legal clarity to the definition:

'For the purpose of this Regulation, the following definitions apply:
(1) *'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;*
(2) *…'*

Draft Annex I was not adopted in the final version of the AI Act, but is still helpful in making more specific what exactly AI systems are:

'ANNEX I ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES referred to in Article 3, point 1
(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
(c) Statistical approaches, Bayesian estimation, search and optimization methods.'

During the legislative process, the Council had favoured a definition designed to make a clearer distinction from traditional software:

"artificial intelligence system' (AI system) means a system that is [1] designed to operate with elements of autonomy and that, [2] based on machine and/or human-provided data and inputs, [3] infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and [4] produces system-generated outputs such as content (gener-

16 Commission Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}, 21 April 2021 (COM(2021) 206 final).

> *ative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts".*

For its part, the European Parliament had proposed the following amendment:

> *"artificial intelligence system' (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments.'*

Summarising these various attempts at definition, they gravitate around four AI criteria:

- elements of autonomy,
- being based on data and inputs (human or machine-based),
- inferring how to achieve objectives using Machine Learning or logic- or knowledge-based approaches,
- producing system-generated outputs (such as predictions, content, recommendations, or decisions) which influence their environment.

Given that the annex of the Commission version was omitted for being potentially too narrow it may be assumed that the annex list can still serve as an example of what definitely constitutes AI, whilst the final version remains in the abstract and allows for an extension of the list as technology progresses.

D. AI use in Public Administration

It is generally believed that AI may be, or already is, used in public administration. This is seen as a potential solution in many European states for the shortage of labour, which also produces a shortage of civil servants and other administrative staff. It may allow public administration to not only fulfil its tasks with less staff, but with improved quality, in particular, by a more objective approach excluding the human factor with its potential for mistakes or for bias.

Areas particularly open to AI use in public administration seem to be:

- Predictive policing: prediction of crime hotspots, recidivism
- Search for victims or criminals (which?)
- Welfare, award of benefits – eg checking that the recipient fulfils requirements
- Healthcare

- Immigration, asylum, and border control
- Education
- Personnel selection
- General advice eg on application procedures, benefits available.[17]

In these, and possibly further, areas, AI might help to provide a more effective, efficient, and low-cost public administration and close retirement gaps in administrative staff. In addition, there is the "promise of neutrality" rather than human intuition, allegedly with the potential to overcome cognitive bias and limitations.

E. Threats, Risks and Benefits of AI

However, it is clear that there are dangers to be considered, too, linked both to AI use in general, and its specific use in public administration. Those looking predominantly at the dangers compare the dangers emanating from AI with those of nuclear weapons or refer to Arnold Schwarzenegger's 'Terminator' films, suggesting the danger of extinction of humankind by AI. Another concern is that humanity may be enslaved by AI, as it may be able to control us rather than the other way round if we do not join forces globally to develop it into a beneficial direction and rather seem to compete to develop it fastest at all costs.[18] One might also think of George Orwell's 1984-'big brother's watching you', invoking the spectre of mass-surveillance, thus signaling the end of privacy and forgetting about data protection.

The promises of AI are also queried, with critics suggesting that AI may rather perpetuate bias and automate inequality than provide the basis for more objective decision-making. AI can only work on the data it gets fed, and where there is human influence, be it only in the data collection, these data will reflect what the data collectors thought relevant, or what was relevant at the time of collection.

Others regard this as exaggerated alarmism, pointing to AI as a similarly important innovation to electrification and the printing press, warning that

17 Commission Strategy and Policy Priorities, webpage, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en#eu-and-ai. Similarly the more generally-oriented list of benefits put forward by Bill Gates, The Bill Gates Blog, A new era – The Age of AI has begun, 21st March 2023, https://www.gatesnotes.com/The-Age-of-AI-Has-Begun.

18 Jonas Breng, Spiegel-Interview: „Starhistoriker Yuval Harari: »KI hat das Potenzial, uns zu versklaven«", 23/10/2024, https://www.spiegel.de/geschichte/starhistoriker-yuval-harari-ki-hat-das-potenzial-uns-zu-versklaven-a-70587486-9be1-4c7c-a1ae-a83c743e74ef.

(over-)regulation might constitute an impediment to innovation, with Europe losing out due to old-fashioned administration.

The EU Commission recognizes both sides and points out the benefits coupled with prerequisites for being able to reap these, requiring excellence and trust in AI: *'Trustworthy AI can bring many benefits, such as better healthcare, safer and cleaner transport, more efficient manufacturing, and cheaper and more sustainable energy'* thus positing that *,The EU's approach to AI will give people the confidence to embrace these technologies while encouraging businesses to develop them.'*[19]

From this the Commission has drawn four key policy objectives, namely:

– Setting enabling conditions
– Building strategic leadership in the sectors concerned
– Ensure that AI technologies work for people and
– Make the EU the right place.[20]

Given the experience of BigTech being domiciled in the US, or increasingly in China, these are grand aims, and it remains to be seen whether the EU will be able to achieve them. There may, in particular, be a fear that the AI Act might stifle innovation in Europe. Regarding AI use in public administration, the third aim – ensuring that AI technologies work for people – appears particularly important, as this coincides with the general aim of public administration. The second important aim is the setting of enabling conditions. Public administration forms a framework for technology development and deployment, namely the legal framework of guaranteed freedoms of business and property as well as providing limits to these guided by public interest and the rights of others. These aspects will be looked at in more detail in the following.

19 Fn. 17, Commission Strategy and Policy Priorities 2019 – 2024, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en.

20 Fn. 17, Commission Strategy and Policy Priorities 2019 – 2024, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en.

F. Main Rules of the AI Act – Trustworthy AI: Regulation by the AI Act

I. Legal effect and scope of the AI Act

The legal effect of the ‚AI-Act' depends on which type of EU act with legislative force it is under Art. 288 TFEU. Its full name clarifies that it is a regulation, meaning that it has direct application in every EU Member State (Art. 288 (2) TFEU), comparable to any member state statute. As a regulation, the AI Act may include individual rights and obligations for each person or undertaking operating in the EU Internal Market, which is confirmed and spelt out in Art. 2 (1) of the AI Act. As mentioned above (B.) the AI Act is a piece of Internal Market and Data Protection regulation based on the legal bases for internal market legislation (Art. 114 TFEU) and data protection (Art. 16 TFEU). This raises the question as to whether the EU can prescribe terms and conditions of AI use for public administration. The EU does not have a legislative competence to legislate on public administration of the Member States generally. The latter is, in principle, subject to autonomous legislation of the Member States with only specific inroads for EU aims, e.g., on legal bases regarding data protection, consumer protection, environmental protection etc.[21] As far as data protection is concerned, there is a relevant EU competence under Art. 16 TFEU.

Art. 2 (1) of the AI Act confirms a general application for providers and deployers, importers and exporters, manufacturers etc. of AI participating in the EU internal market, as well as any person affected and located in the EU:

> *Article 2*
> *Scope*
> *1. This Regulation applies to:*
> *(a) providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country;*
> *(b) deployers of AI systems that have their place of establishment or are located within the Union;*

21 Such as the GDPR with its general application including public administration, or inserting the obligation to conduct an Environmental Impact Assessment into relevant administrative procedures in environmental law, based on the specific legal basis for environmental law, Art. 191/192 TFEU.

(c) providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the Union;
(d) importers and distributors of AI systems;
(e) product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark;
(f) authorised representatives of providers, which are not established in the Union;
(g) affected persons that are located in the Union.

Even though public administration is not particularly mentioned in the scope of the AI Act, public entities like a state, a sub-state public entity such as a federal state in Germany, or a municipality, or entities organized or otherwise linked to these, may at least count as an 'affected person', unless they are 'deployers'. A 'deployer' of an AI system is, under Art 3 (4) AI Act 'a natural or legal person, public authority, agency or other body using an AI system under its authority' (with the exception of personal non-professional use). This explicit inclusion of public authority AI use clarifies that the AI Act means to bind public authorities, and that this is another inroad into Member State competence to regulate their public administrations autonomously. Accordingly, the scope of application of the AI Act includes any administrative authority using AI, or being otherwise exposed to AI. In addition, the underlying aim that only 'trustworthy' AI – AI compliant with the AI Act – should be available on the Internal Market of the EU, and legal to use in the EU, means that public administration, too, will not legally be able to obtain any AI solution not complying with the AI Act, and is bound just like any private entity, to use it legally, i.e. according to the AI Act.

II. Overview, Risk Classification

The AI Act is divided into 13 chapters, beginning with General Provisions in Chapter I, proceeding, according to risk, from unacceptable to minimal or no risk and varying risk:

- Chapter II defines and prohibits AI Practices deemed unacceptable,
- Chapter III regulates high-risk AI Systems, which are permitted but subject to compliance and specific requirements and an *ex-ante* conformity assessment,
- Chapter IV introduces transparency obligations for providers and deployers of certain AI systems with limited risk,
- Chapter V deals with general-purpose AI.

The following chapters include a Chapter VII on Governance, whilst the other chapters appear more relevant for business rather than public administration, and will only be considered further where relevant for AI use in public administration.

III. According to Risk: from Prohibition to Transparency Obligations

1. Unacceptable-Risk: Prohibited AI

Art. 5 of the AI Act provides a list of prohibited AI uses. Some are obviously out of bounds for a state governed by the rule of law and respect for human rights. Under Art. 5 (1) of the AI Act prohibitions exist namely for:

- subliminal behaviour manipulation,
- exploitative behaviour manipulation of vulnerable persons or groups,
- governmental (and private) social scoring (out of context or of a certain gravity),
- risk assessment (apart from certain human-surveilled uses regarding criminal involvement of persons),
- facial recognition (only untargeted scraping of facial images),
- recognition of emotions in workplace and education institutions,
- biometric categorization on criteria which form the basis of discrimination.

Accordingly, Art. 5 (1) of the AI Act prohibits

> *'(a) the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm;*
> *(b) the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm;*
> *(c) the placing on the market, the putting into service or the use of AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or*

predicted personal or personality characteristics, with the social score leading to either or both of the following:

(i) detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;

(ii) detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity;

(d) the placing on the market, the putting into service for this specific purpose, or the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity;

(e) the placing on the market, the putting into service for this specific purpose, or the use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;

(f) the placing on the market, the putting into service for this specific purpose, or the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons;

(g) the placing on the market, the putting into service for this specific purpose, or the use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation; this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement;

…'

Each of these types of AI might easily violate human dignity (Art. 1 FRC) above, by making human beings into mere objects of state activity. In addition, and without the space for examining this in more depth, other Charter rights may also be engaged. Article 3 FRC guarantees the right to the physical and mental integrity of any person, Article 7 FRC respect for private and family life, Article 8 FRC, together with the GDPR, the protection of personal data concerning the relevant person. Each of these

fundamental rights might easily be violated by manipulating behavior and by social scoring or mere AI risk assessment.

There is still room for interpretation, and principally beneficial uses may be prohibited, too. For instance, as van de Sloot[22] points out, this could apply to

> *'[t]he case in which Augmented Reality systems are used in smart cities where people are subconsciously manipulated. Although it is true that these manipulations are so far 'for the better' – for example, where Augmented Reality systems are deployed as a means of crowd control to prevent traffic jams or to make people in nightlife areas less aggressive – it cannot be excluded that these techniques could and would be used for more controversial purposes. In addition, it will become clear, presumably through jurisprudence on the matter, the extent to which subconscious manipulation itself amounts to harm, even when it is used 'for the better'.*

Indeed, any behavior manipulation would violate the fundamental right under Article 10 FRC, the freedom of thought, conscience and religion: how should a person, under the influence of such manipulation, 'exercise their freedom to change religion or belief and freedom, either alone or in community with others and in public or in private, to manifest religion or belief, in worship, teaching, practice and observance'[23]? In addition, biometric categorisation that employs criteria excluded as justifiable reasons of discrimination under Art. 21 and 23 FRC would form the basis of exactly such discrimination.

The prohibition of AI use for social scoring was originally specifically addressed to public administration[24], but has been extended to a general prohibition, meaning that private parties, such as religious or other societies will be unable to use social scoring, either. It may appear a violation of rights to religious freedom and freedom of conscience etc. to prohibit such social scoring, given that they are entities upholding moral and ethical values. However, the prohibition is subject to limits itself, as it does not prohibit social scoring altogether and absolutely, but only if the consequences are unrelated or unjustified or disproportionate to the persons' social behaviour or its gravity. There will be a lot of work required on the part of public

22 Van de Sloot, Regulating Synthetic Society, p. 149.

23 Art. 10 FRC.

24 Art. 5 (1) lit. c) of the Commission Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) 21 April 2021 (COM(2021) 206 final), see fn. 16.

administrations and courts to make this prohibition, as well as its limits, sufficiently specific.

There are a number of further prohibitions, coupled with limited exceptions for criminal prosecution. Germany had suggested more prohibitions, such as on the use of 'robo-judges', whilst the assessment of future delinquency and systematic workplace surveillance have been added according to Germany's wishes.

2. High-Risk AI: risk control, upholding of fundamental rights

Chapter III of the AI Act regulates and severely restricts the use and deployment of 'high risk' AI, imposing strict requirements on its use. In the following, the concept of 'high-risk'-AI will be developed, at least as far as it appears relevant to public administration (a). After this the requirements and restrictions on the use of high-risk AI will be explored (b). In addition, there are provisions on 'Notifying authorities and notified bodies' (Chapter III section 4 AI Act) and 'Standards, conformity assessment, certificates, registration' (Chapter III section 5 AI Act) regarding high-risk AI, which will only be touched upon, and only where relevant to AI use in public administration is concerned (not regarding the setting up of AI surveillance authorities and procedures).

a) Types of High-Risk AI Systems

aa) AI as product or safety component of a product covered by EU harmonisation

Under Art, 6 (1) AI Act, an AI system shall be considered to be high-risk where both of the following conditions are fulfilled:

'(a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I;

(b) the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I.'

In the public administration context, relevant AI systems under Art. 6 (1) AI Act, ie systems that are either a product or safety component of a product already subject to an EU conformity assessment, may, for instance, include

- high-risk AI systems in machinery (No. 1)[25], as far as they are used by public entities,
- lifts and safety components for lifts (No. 4)[26], namely in public buildings,
- radio equipment (No. 6)[27], where radio stations are public,
- personal protective equipment (No. 9)[28], e.g. regarding police and rescue operations
- appliances burning gaseous fuels (No. 10)[29],
- medical devices (No 11 of Annex I)[30], namely in public hospitals, such as an AI application for robot-assisted surgery.
- *in vitro* diagnostic medical devices (No 12. of Annex I)[31] in public hospitals,
- civil aviation security (No 13)[32],
- rail systems (No 17)[33], where they are public, etc.

---

25 Under Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC, OJ L 157, 9.6.2006, p. 24.

26 Directive 2014/33/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to lifts and safety components for lifts, OJ L 96, 29.3.2014, p. 251.

27 Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC, OJ L 153, 22.5.2014, p. 62.

28 Regulation (EU) 2016/425 of the European Parliament and of the Council of 9 March 2016 on personal protective equipment and repealing Council Directive 89/686/EEC, OJ L 81, 31.3.2016, p. 51.

29 Regulation (EU) 2016/426 of the European Parliament and of the Council of 9 March 2016 on appliances burning gaseous fuels and repealing Directive 2009/142/EC, OJ L 81, 31.3.2016, p. 99.

30 Harmonised by Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, OJ L 117, 5.5.2017, p. 1.

31 Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU, OJ L 117, 5.5.2017, p. 176.

32 Regulation (EC) No 300/2008 of the European Parliament and of the Council of 11 March 2008 on common rules in the field of civil aviation security and repealing Regulation (EC) No 2320/2002, OJ L 97, 9.4.2008, p. 72.

33 Directive (EU) 2016/797 of the European Parliament and of the Council of 11 May 2016 on the interoperability of the rail system within the European Union, OJ L 138, 26.5.2016, p. 44.

Equally, further products which are subject to harmonisation are listed in Annex I, and may be used by public entities.

bb) Further AI Systems listed as high-risk

Art. 6 (2) of the AI Act adds the reference to AI systems listed in Annex III (which provides a regularly evaluated and amendable list[34]). Such Annex-III-listed systems shall equally be considered high-risk and include, with relevance to public administration:

- *Biometrics* (Annex III No. 1) are regarded as highly risky (if not prohibited, see above), namely remote *biometric identification* systems unless intended to be used for the sole purpose to confirm that a specific natural person is the person he or she claims to be, as well as AI systems intended to be used for *biometric categorisation*, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics, and AI systems intended to be used for *emotion recognition*.
- *Critical infrastructure* (Annex III No. 2) includes facilities, typically provided by public administration, which everyone needs to use, leaving no choice as to whether to expose oneself to AI or not. Here high risk AI systems are those intended to be used as *safety components* in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity.
- *Education and vocational training* (Annex III No. 3) are equally very sensitive areas, and provided by the state, or under the supervision of the state. Accordingly, AI systems intended to be used to determine *access or admission* or to assign natural persons to educational and vocational training institutions fall into the 'high risk' category, as well as AI systems intended to be used to *evaluate learning outcomes*; or to be used for the purpose of *assessing the appropriate level* of education that an individual will receive or will be able to access. Equally, AI systems intended to be used for *monitoring and detecting prohibited behaviour* of students during tests in the context of or within educational and vocational training institutions are high risk.
- The same applies in the – public or private – *employment* sphere, namely employment, workers' management and access to self-employment (Annex III No. 4): AI systems in *recruitment or selection* of natural persons, to

34 For amendment see Art. 112 of the AI Act.

analyse and filter job applications, and to evaluate candidates are classed as high risk; this also applies to public service employment. Equally, AI use within the employment relationship is classed as high risk, namely if intended to be used to make *decisions affecting terms, promotion or termination* of work-related relationships, or to allocate tasks based on individual behaviour or personal traits or characteristics or to monitor and evaluate the performance and behaviour of persons in such relationships.

- The relevance for public administration is particularly obvious with regard to *access to and enjoyment of essential public services and benefits*, which is also classed as high-risk (Annex III No. 5, together with essential private services): This includes AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the *eligibility* of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services (a), or AI systems intended to evaluate and classify emergency calls by natural persons or to be used to dispatch, or to establish priority in the dispatching of, emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems (d).
- *Migration, asylum and border control management* is a specific, highly sensitive area of public administration, where public debate at times appears to neglect that illegal migrants, too, remain holders of fundamental rights. Where AI use is accepted by law in this area, it is thus classed as high risk use (Annex III No. 7), namely if AI systems are intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies as polygraphs or similar tools (a), or to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered into the territory of a Member State (b), or to assist competent public authorities for the examination of applications for asylum, visa or residence permits and for associated complaints with regard to the eligibility of the natural persons applying for a status, including related assessments of the reliability of evidence (c), or, finally, for the purpose of detecting, recognising or identifying natural persons, with the exception of the verification of travel documents (d).

Following the intricate classification of high-risk AI in Annex III, Art. 6 (3) AI Act provides for an across-the-board *de minimis* exception to which the classification is subjected. An Annex-III-AI system shall not be considered to be high-risk where '*it does not pose a significant risk of harm to the health, safety*

or fundamental rights of natural persons, including by not materially influencing the outcome of decision making'. This exception is made more specific by a list of four criteria, which share the feature of a decisive involvement of human control, namely if the AI system is intended to

- perform a narrow procedural task (a),
- improve the result of a previously completed human activity (b),
- detect decision-making patterns or deviations from prior decision-making patterns, if it is not meant to replace or influence the previously completed human assessment, without proper human review (c), or
- perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III, such as biometrics as far as permitted, safety components in the management and operation of critical infrastructure or in education and vocational training (d).

Art. 6 (3) AI Act then provides for a counter-exception, overriding the *de minimis* exceptions where the AI system performs profiling of natural persons.

Art. 6 AI Act, with all its complexity, including exception and counter-exception, and references to articles and annexes rather than content, does not make it easy to apply in administrative practice. Still, the principal practices classified as 'high risk' appear ones obviously risky for human rights and data protection, or indeed human dignity, requiring safeguards to ensure that human beings are not treated as mere objects. The clarity is to some extent obscured by the across-the-board *de minimis* rule, which is then narrowed down again by the counter-exception regarding AI profiling. In order to secure respect for human rights it may thus be advisable to interpret *de minimis* narrowly, and rather apply the requirements for high risk AI use in public administration wherever there may be an effect, even if small, on human rights.

b) Requirements for High-Risk AI systems

aa) Overview

High-risk AI, regardless of whether it is a stand-alone system, or whether it is part of a product, must meet specific requirements under the AI Act, in order to control the risk attached to them, aiming both at protection by design *a priori*, before market presentation, and after deployment. The risk may pertain namely to health or safety or to the impairment of further fundamental rights (privacy, data protection, freedom of occupation and property, non-discrimination etc.).

Requirements for high-risk AI systems are laid down, first, in Section 2 of Chapter III of the AI Act (Art. 8 et seq.). Article 8 (1) of the AI Act demands that high-risk AI systems shall comply with the requirements laid down Section 2, taking into account their intended purpose as well as the generally acknowledged state of the art on AI and AI-related technologies. Article 8 (2) of the AI Act deals with requirements regarding products containing an AI system, stating that the requirements of the AI Act as well as requirements of other EU harmonisation legislation[35] apply; providers of such products shall be responsible for ensuring that they are fully compliant. In order to ensure consistency, avoid duplication and minimise additional burdens, providers are given a choice of integrating the necessary AI-related testing and reporting processes, information and documentation into the documentation and procedures required under other EU harmonisation legislation. Section 3 of Chapter III states further obligations of providers and deployers of high-risk AI systems and other parties. The manufacturer and the provider of the product are each responsible for compliance with these rules, which should be understood, *inter alia*, as safeguards against self-learning AI systems that go rogue.[36]

bb) Risk management system

The first specific requirement for high-risk systems is having a risk management system (Article 9 of the AI Act), which shall be established, implemented, documented and maintained (Art. 9 (1) AI Act). It shall be understood as a continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating and comprising an identification and analysis of known and foreseeable risks (Art. 9 (1) AI Act). The risk management system shall comprise four steps, namely:

- *identification* and analysis of the known and the reasonably foreseeable risks to health, safety or fundamental rights when in regular use (a);
- the estimation and *evaluation* of the risks that may emerge when the high-risk AI system is used in accordance with its *intended purpose*, and under conditions of reasonably foreseeable misuse (b);

35 Listed in Section A of Annex I.

36 Cf. Preamble No 48 of the Commission Draft and No 73 of the final version of the AI Act.

- the *evaluation of other risks* possibly arising, based on the analysis of data gathered from the post-market monitoring system referred to in Article 72 (c) AI Act;
- the *adoption of* appropriate and targeted risk management *measures* designed to address the risks identified (d).

These four steps – risk identification, intended-purpose-use evaluation, other risk evaluation, adoption of measures – appear relatively far-reaching at a first glance. This may raise concerns that the risk management obligation is so onerous as to constitute a competitive disadvantage for AI use in the EU. The risk management system requirements are, however, limited in the following paragraph, Art. 9 (3) AI Act, to those risks 'which may be reasonably mitigated or eliminated through the development or design of the high-risk AI system, or the provision of adequate technical information.' If a serious risk cannot be mitigated, however, this should mean the AI system is prohibited altogether, presumably falling under Art. 5 AI Act (the scope of which may need to be extended). The further paragraphs then specify some requirements regarding the risk management system (Art. 9 (4) et seq. AI Act), requiring consideration of the AI Act requirements, further restricting the elimination or reduction of risks to doing what is 'technically feasible through adequate design and development' of the high-risk AI system, considering risk elimination and reduction by providing information and training to deployers, and reducing the residual risk to what is deemed 'acceptable' (without specifying what that is supposed to mean).

cc) Data and Data Governance

Secondly, Article 10 AI Act sets out requirements with regard to data and data governance. The data used for training or testing AI shall be developed on the basis of data sets that meet the quality criteria referred to in the following paragraphs (2 to 5), subjecting data sets to data governance and management practices 'appropriate for the intended purpose of the high-risk AI system.' Those include measures to make sure data meet certain scientific quality criteria e.g. regarding their origin and validity. What is particularly important regarding non-discrimination is the examination of data 'in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations' (Art. 10 (2) lit. h and g).

dd) Documentation and Recording, Transparency obligations

Third, documentation prior to and during the lifetime of a high-risk system is required. Article 11 AI Act imposes technical documentation duties regarding high-risk AI systems, to be drawn up before the system is placed on the market or put into service and kept up-to date. When training models, data should be kept on the design choices, data collection and relevant data preparation, which should include an examination in view of possible biases and the identification of any possible data gaps or shortcomings. The technical documentation is to prove compliance with the requirements of Art. 8 et seq. AI Act, namely to national competent authorities and notified bodies, with the minimum specified in Annex IV. There are simplified documentation requirements for SMEs, including start-ups – not for public administration. During use, Article 12 requires High-risk AI systems to technically allow for the automatic recording of events (logs) over the lifetime of the system, in order to be able to monitor the operation of high-risk AI systems referred to in Art. 26 (5) AI Act. In addition, Art. 13 AI Act imposes duties of transparency and requires the provision of information by providers to deployers of high-risk AI systems, who should be enabled 'to interpret a system's output and use it appropriately', 'with a view to achieving compliance with the relevant obligations of the provider and deployer'.

ee) Human oversight

Fourth, and catering for fears of a 'Terminator' scenario, Art. 14 demands effective human oversight by natural persons during the period in which high-risk AI systems are in use. Human oversight is meant to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse. The following paragraphs detail the oversight measures, which shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system, namely 'measures identified and built, when technically feasible, into the high-risk AI system' by the provider before it is placed on the market or put into service (a) or measures that are appropriate to be implemented by the deployer (b). This in particular should constitute a safeguard against self-learning AI systems, preventing them from going rogue.[37] Still, all will

37 Cf. Preamble No 48 of the Commission Draft and No 73 of the final version of the AI Act.

depend on the choice and control of persons to oversee the system, and whether or not they perform their tasks, or profit or take pleasure from experimenting with its getting out of control for their own purposes and possibly with a hybris that they themselves may still be able to exercise control. Civil servants may appear particularly trustworthy here, but, e.g., the organisation of reliable oversight with more than one person appears crucial in civil service, too.

ff) Accuracy, Robustness and Cybersecurity

Fifth, Article 15 requires accuracy, robustness and cybersecurity of high-risk AI systems throughout their lifecycle. This is to be achieved by involving expertise of relevant stakeholders and organisations such as metrology and benchmarking authorities, encouraging the development of benchmarks and measurement methodologies, declared in the accompanying instructions of use. Paragraphs (4) and (5) of Art. 15 deal with resilience regarding errors, faults or inconsistencies, calling for technical and organisational measures, for instance technical redundancy solutions, which may include backup or fail-safe plans. The risk of possibly biased outputs influencing input for future operations (feedback loops) in learning AI systems shall be avoided by developing AI in such a way as to eliminate or reduce this risk as far as possible and as to ensure that any such feedback loops are duly addressed with appropriate mitigation measures. In addition, high-risk AI systems need to be resilient against attempts by unauthorised third parties to alter their use, outputs or performance by exploiting system vulnerabilities. The technical solutions aiming to ensure the cybersecurity of high-risk AI systems shall be 'appropriate to the relevant circumstances and the risks,' whatever that is to mean exactly, and shall include measures to prevent, detect, respond to, resolve and control attacks trying to manipulate the training data set (data poisoning), or pre-trained components used in training (model poisoning), inputs designed to cause the AI model to make a mistake (adversarial examples or model evasion), confidentiality attacks or model flaws. The use of such wide legal concepts affects legal certainty and requires fleshing out by public administration and case law.

gg) Further Obligations of Providers

Providers have to ensure that their high-risk AI systems are compliant with the various requirements, have a quality management system in place, draw up the technical documentation, keep logs, do a conformity assessment, take the necessary corrective actions when a breach is established, and inform the supervisory authorities thereof. Section 3/Art. 16 et seq. of the AI Act state these obligations of providers of high-risk AI systems and other parties, namely to

- ensure compliance with section 2 (Art. 16 (a)) AI Act,
- indicate their name, registered trade name or registered trade mark, the address at which they can be contacted Art. 16 (b) AI Act,
- have a quality management system – on top of the risk management system – in place (Art. 16 (c) and Art. 17 AI Act (specifying the details of such quality management),
- keep the documentation, in particular to allow for checks and surveillance by the national authorities, namely technical documentation and the documentation concerning the quality management system et al. (Art. 16 (d) and Art. 18 AI Act),
- keep the logs automatically generated by their high-risk AI systems (Art. 16 (e) and Article 19 AI Act),
- ensure that the high-risk AI system undergoes the relevant conformity assessment procedure as referred to in Article 43, prior to its being placed on the market or put into service, Art. 16 (f) AI Act,
- comply with the registration obligations referred to in Art. 49(1), Art. 16 (i) AI Act,
- take the necessary corrective actions and provide information as required in Article 20, Art. 16 (j) AI Act,
- demonstrate the conformity of the high-risk AI system with the requirements set out in Section 2 upon request of the national competent authority, Art. 16 (k), 21 AI Act, and
- ensure that the high-risk AI system complies with accessibility requirements laid down in EU law[38], Art. 16 (l) AI Act.

These obligations seem to offer a dense package of safeguards for high-risk systems, applying at the beginning of the supply chain, with providers. What is remarkable here is that researchers, engaged in system development, and producers do not appear to be addressed, meaning that, in cases where

38 Directives (EU) 2016/2102 and (EU) 2019/882.

an AI system is not provided to potential deployers, or is not imported from outside the EU (see hh)), it does not need to meet the requirements of these sections. This may support innovation, making sure that conceptual research can be conducted freely to start with, and may be a wise move in regulation of such a new and heretofore unregulated area where updates to the legal framework will be required in any case, potentially requiring restrictions at source, too. In addition, since regulating research may not fall under the EU internal market legislative competence, this will be an area left to Member States' Regulation.

hh) Obligations of Importers, Distributors etc.

Apart from AI providers, other persons and undertakings involved with AI are included among those obliged to ensure that AI be 'trustworthy'. Along the value chain this first applies to importers, who are subject to various verification duties under Art. 23 (1) AI Act, to ensure that the system is in conformity with the AI Act. This mainly amounts to verifying that the provider has fulfilled all obligations under the AI Act, such as verifying that the relevant conformity assessment procedure[39] has been carried out by the provider of the high-risk AI system (a), the provider has drawn up the technical documentation in accordance with Article 11 and Annex IV(b), etc. (see gg)). In addition, there are also obligations imposed on importers to notify any suspected non-conformity or falsification, giving the importer's name, registered trade name or registered trade mark, plus their address, and to keep documentation and present it to the authorities upon request. Similar obligations apply to distributors under Art. 24 AI Act.

In order to exclude any gaps in the obligations of persons or undertakings handling AI, Art. 25 AI Act imposes further responsibilities along the AI value chain, essentially considering any distributor, importer, deployer or other third-party as a provider of a high-risk AI system, and subject to the obligations of the provider under Art. 16, in any circumstances potentially limiting the responsibility of another under para (2), for instance, if they put their name or trademark on a high-risk AI system already placed on the market or put into service (Art. 25 (1)(a) AI Act) or make modifications (b, c).

39 Referred to in Article 43.

ii) Obligations of Deployers of High-Risk AI Systems – with respect to Public Administration

Obligations of deployers of high-risk AI systems are particularly relevant here, as public administration authorities will probably mainly appear in the role of a deployer of AI, not as a provider (although interesting questions may arise if public funding goes into research producing AI systems). Deployer obligations are imposed by Art. 26 AI Act. These duties rest on public administration in the role of a deployer, as well as on any other, but there are some specifically relevant for public bodies.

First, deployers of high-risk AI systems shall take appropriate technical and organisational measures to ensure they use such systems in accordance with the instructions for use accompanying the systems. Second, addressing fears of a 'Terminator' scenario, there is a strict requirement of human oversight: deployers shall assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support (Art. 26 (2) AI Act).

Third, Art. 26 (3) AI Act addresses any concurrent obligations, as well as deployer's freedom of self-organisation, stating that the duties of ensuring use according to instructions and human oversight are without prejudice to other deployer obligations under EU or national law on the one hand and to the deployer's freedom to organise its own resources and activities for the purpose of implementing the human oversight measures indicated by the provider on the other. The latter is particularly relevant for public administration where there is a constitutional autonomy of public entities, as in Germany - namely the federal states' and the municipalities' autonomy, which might be affected if this exception did not exist.

There are further duties on high-risk AI deployers, namely regarding the quality of input data (para (4)) and the monitoring of the operation of the system, coupled with duties of information, namely to inform providers or to pass on documentation. Where high-risk AI deployers have reason to consider that the use of the system in line with the instructions may result in that AI system presenting a risk to persons' health or safety, or to fundamental rights, they shall inform the provider or distributor and the relevant market surveillance authority, and shall suspend the use of that system. Where deployers have identified a serious incident, they shall also immediately inform the provider, and then the importer or distributor and the relevant market surveillance authorities of that incident (para (5)). In addition to the general documentation duties, deployers of high-risk AI systems shall keep the logs automatically generated by that high-risk AI system

to the extent such logs are under their control, for at least six months, unless provided otherwise in particular in EU data protection law, namely the GDPR[40] (para. (6)). Regarding the civil service, if deploying high-risk AI, public administration entities, too, must inform their civil servants and other workforce, like other employers, before putting a high-risk AI system into service or using it at the workplace, both workers' representatives and the affected workers themselves (para (7)).

Public authorities also have additional obligations under Art. 26 (8) AI Act if they deploy high-risk AI systems[41]: deployers that are public authorities, or Union institutions, bodies, offices or agencies shall comply with registration obligations[42], meaning that they must register their selected system's use in the EU database[43] before putting it into service or using it. They should only use (and register their use) if the system is already registered by the provider: Art. 26 (8) demands that, when public deployers find that the high-risk AI system that they envisage using has not been registered in the EU database[44], they shall not use that system and shall inform the provider or the distributor.

Another obligation particularly relevant to public entities under the rule of law and human dignity protection is data protection. This is already catered for by the GDPR as mentioned above[45], but the obligation to carry out a data protection assessment is specifically mentioned for deployers of high-risk AI systems in Art. 26 (9) AI Act: deployers of high-risk AI systems shall use the information provided under the transparency obligation on AI providers under Art. 13 AI Act in order to comply with their obligation to carry out a data protection impact assessment under the GDPR[46].

jj) Fundamental Rights Impact Assessment

Regarding AI with a high risk for human rights protection, there are further specific obligations for public administration, whether bodies governed by

40 GDPR, see fn. 4.

41 I.e. those listed in Annex III, with the exception of high-risk AI systems listed in point 2 of Annex III, Art. 49 (3) AI Act.

42 Referred to in Article 49.

43 Database referred to in Article 71.

44 Database referred to in Article 71.

45 See above A., F.IV.

46 GDPR see above fn. 4. The same applies under Article 27 of Directive (EU) 2016/680 on data protection in criminal prosecution.

public law or entities organized under private law but performing public services. This expressly includes an impact assessment regarding the human rights impact:

> *Article 27*
> *Fundamental rights impact assessment for high-risk AI systems*
> *1. Prior to deploying a high-risk AI system referred to in Article 6(2), with the exception of high-risk AI systems intended to be used in the area listed in point 2 of Annex III, deployers that are bodies governed by public law, or are private entities providing public services, and deployers of high-risk AI systems referred to in points 5 (b) and (c) of Annex III, shall perform an assessment of the impact on fundamental rights that the use of such system may produce. For that purpose, deployers shall perform an assessment consisting of:*
>
> > *(a) a description of the deployer's processes in which the high-risk AI system will be used in line with its intended purpose;*
> > *(b) a description of the period of time within which, and the frequency with which, each high-risk AI system is intended to be used;*
> > *(c) the categories of natural persons and groups likely to be affected by its use in the specific context;*
> > *(d) the specific risks of harm likely to have an impact on the categories of natural persons or groups of persons identified pursuant to point (c) of this paragraph, taking into account the information given by the provider pursuant to Article 13;*
> > *(e) a description of the implementation of human oversight measures, according to the instructions for use;*
> > *(f) the measures to be taken in the case of the materialisation of those risks, including the arrangements for internal governance and complaint mechanisms.*
>
> *2. The obligation laid down in paragraph 1 applies to the first use of the high-risk AI system. The deployer may, in similar cases, rely on previously conducted fundamental rights impact assessments or existing impact assessments carried out by provider. If, during the use of the high-risk AI system, the deployer considers that any of the elements listed in paragraph 1 has changed or is no longer up to date, the deployer shall take the necessary steps to update the information.*
> *3. Once the assessment referred to in paragraph 1 of this Article has been performed, the deployer shall notify the market surveillance authority of its results, submitting the filled-out template referred to in paragraph 5 of this Article as part of the notification. In the case referred to in Article 46(1), deployers may be exempt from that obligation to notify.*

> *4. If any of the obligations laid down in this Article is already met through the data protection impact assessment conducted pursuant to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, the fundamental rights impact assessment referred to in paragraph 1 of this Article shall complement that data protection impact assessment.*
> *5. The AI Office shall develop a template for a questionnaire, including through an automated tool, to facilitate deployers in complying with their obligations under this Article in a simplified manner.*

This provision marks an attempt to leave no stone unturned before public entities deploy AI, whilst facilitating its use after the first time of its being put into service. Certainly this raises the hurdles for using AI at all in the public service, but, at the same time, it may help to maintain public awareness of risks attached to AI use, in particular to human rights protection central to liberal democracies..

kk) Conformity Assessment prior to Use

A conformity assessment prior to use is prescribed in Art. 43 AI Act, among the standards, applying to high-risk AI. Depending on the type of AI system, the conformity assessment may be undertaken by providers of the AI system themselves (Art. 43 (1)(a)/Annex VI); otherwise participation of a notified body is required (Art. 43 (1) (b)/Annex VII). Importers of high-risk AI systems have to ensure that the appropriate conformity assessment procedure has been carried out by the provider of that AI system, that the provider has drawn up the required technical documentation and that the system bears the required conformity marking etc. Equally, distributors have to verify that the system bears the required conformity markings, that it is accompanied by the requisite documentation and instruction of use, and that the provider and the importer of the system have complied with their obligations under sections 2 and 3.

Conformity refers to the AI Act and, where applicable, also to EU legislation other than the AI Act. The conformity assessment aims to demonstrate that the AI system meets the mandatory requirements for trustworthy AI (in terms of data quality, documentation and record keeping, transparency and provision of information to users, human oversight, robustness, accuracy and cybersecurity (Chapter III Section 2, Art. 9 et seq.). The relevant process involves four steps:

Step 1: Development of AI-System

Step 2: Conformity Assessment
Step 3: Registration
Step 4: Declaration of conformity, CE-label.

Only after these four steps are concluded a high-risk AI system may be placed on the market, and only then will it be available – and lawful to use – for bodies of public administration.

ll) Assessment of High-risk AI requirements

The requirements for provision and deployment of high-risk AI presented in the above seem very far-reaching, as befits the potential risk to human rights protection, human dignity in particular, and liberal democracy. At the same time, research even with high risk potential is left free from restrictions as such, which leaves the potential for innovation untouched.

The obligations under this risk-based approach appear more procedural than substantive. In spite of allowing for free research, one also needs to consider who will be driving innovation, and whether they can cope with the burdens under the risk-based approach. As van de Sloot points out:

> *It may stifle innovation. Because of the administrative and bureaucratic costs, it will most likely be larger corporations rather than innovative start-ups that can uphold these rules and because these bigger parties prefer safety. Risk-based regulation is open and vague; parties do not get clear guidance from the regulator on what is and what is not allowed. Strict rules often boost rather than stifle innovation, because companies know within which boundaries they can operate and can thus make stable long-term investments.*[47]

Still, given that technological progress in the field of AI occurs every day it is very difficult, if not impossible, to fix on specific substantive requirements, and thus appears wise to keep them rather general and focus on procedural ones. This leaves a lot of responsibility with providers and deployers of systems, but also leaves technologies largely free to evolve. Future regulation may be based on the results of documentation and recording, enabling legislation to be more specific. It is still a step forward as opposed to leaving the direction and risk control entirely to providers and deployers. As van de Sloot states:

47 Van de Sloot, Regulating Synthetic Society, p. 153.

> *'The evaluation of potential risks is initially left up to private organisations, only later to be assessed by a regulatory [authority] or court. It is known that this generally means that the 'good guys' will err on the safe side, while it will be the 'bad guys' who look for loopholes in the measures to exploit in a business-friendly way. More generally, the goal of impact assessments, under the AI Act, the GDPR and other EU data instruments, is that risks will be identified and that if these risks cannot be adequately mitigated, the project is cancelled. In practice, however, such decisions are seldom made; certain applications that are in essence problematic (from a substantive perspective) are commonly legitimised by adopting additional procedural safeguards.'*[48]

Whilst this appears plausible per se it may store up problems for later on, if a subsequent first attempt at substantive legislation makes the wrong choices in substantive requirements. Accordingly, to avoid the risk of stifling innovation at that stage, it may be better from the start to involve the innovating side – preferably the 'good guys' – in developing the substantive requirements in due course, rather than doing either nothing or the wrong thing. In this regard, some more substantive rules, such as permissible defenses for wrong results, could arguably have been included, combining the risk-based procedural and the substantive approach.[49]

3. General-purpose AI (GPAI)

'General-purpose AI model' means, according to Art. 2 (63) AI Act, an AI model that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market. A GPAI model can be integrated into a variety of downstream systems or applications. Such an AI model may be trained with a large amount of data using self-supervision at scale. A 'general-purpose AI system' means an AI system which is based on such a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems (Art. 2 No 66 AI Act). A well-known example is ChatGPT.

GPAI is regulated in a separate chapter, Chapter V, of the AI Act (Art. 51 et seq.). General-purpose AI models, regardless of risk level, must offer technical documentation, usage guidelines, adhere to copyright laws, and disclose summaries of their training data. Providers of free and open license GPAI have fewer requirements unless they present systemic risks. All GPAI providers with systemic risks, whether open or closed, must conduct model

48 Van de Sloot, Regulating Synthetic Society, p. 153.

49 As suggested by van de Sloot, Regulating Synthetic Society, p. 154.

assessments, adversarial testing, report serious incidents, and ensure cybersecurity measures (Art. 53 et seq. AI Act). In addition, GPAI must fulfil the additional requirements depending on their risk level.

GPAI models that are used for research, development or prototyping activities before they are placed on the market are not covered by the AI Act.

4. Limited-Risk AI: Reveal Yourself - Transparency

The AI Act acknowledges that not all AI necessarily poses a risk, let alone a high risk. Therefore, it largely exempts AI systems which do not lead to a significant risk of harm to the legal interests protected in the AI Act, namely because they do not materially influence the substance or outcome of decision-making or do not harm protected interests substantially. Under Preamble No 53 this is namely the case where the AI system

- is intended to perform a narrow procedural task, such as an AI system that transforms unstructured data into structured data, an AI system that classifies incoming documents into categories or an AI system that is used to detect duplicates among a large number of applications,
- provides only an additional layer to a human activity, such as AI systems that are intended to improve the language used in previously drafted documents,
- is intended to detect decision-making patterns or deviations from prior decision-making patterns, following a previously completed human assessment which it is not meant to replace or influence, without proper human review, such as smart solutions for file handling, which include various functions from indexing, searching, text and speech processing or linking data to other data sources,
- is used for translation of initial documents.

There are several categories of limited-risk AI which accordingly not subject to the main compliance provisions of the AI Act, but merely to transparency rules (Art. 50 AI Act): in short such AI must still reveal itself. „Chatbots", recognition systems or biometric categorisation systems are thus not per se illegal or subject to onerous use conditions, but it must be obvious to the users that they are dealing with AI. For this, the information of AI use shall be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure (Art. 50 (5) AI Act). Namely this applies to limited-risk AI

> which is intended to interact directly with natural persons (Art. 50 (1) AI Act). This means, for instance, that should humanoid robots become truly indistinguishable from humans, they should let natural persons

with whom they interact know that they are non-human, for example, when used for hospitality purposes.[50] The transparency obligations to interacting users serve the purpose that users may consciously decide whether they want to continue using the application. Such a disclosure obligation also applies to AI use by public administration. Here the question arises as to whether public administration may make it obligatory to use AI in order to eg submit an application for a public service, or for citizens to perform a duty to the public. One issue will be inclusion, as it cannot be expected of all users that they are able to use AI systems so far, but a person rejecting AI use remains a citizen who must be able to interact with public administration. This may limit the options for public administration to introduce AI in its processes.

Regarding AI systems that generate or manipulate content, the origin of the content must be disclosed (Art. 50 (2) AI Act):

'Providers of AI systems generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards. …'

This provision has raised concerns as to whether it is not so wide as to be impracticable:

The provision seems to cover all manipulated content; however, the problem is that, depending on the definition of manipulation, any communication technology distorts reality: video services have built-in tools that smoothen skin tones, audio services automatically filter out treble and so forth. Some experts estimate that, in five years, more than 90 per cent of all online content will be manipulated in some form. Importantly, the AI Act also refers to manipulated content about objects, places and events. Would this provision consequently also apply to a smiley-faced sun?[51]

It will remain to be seen whether this concern proves correct, and legal application may have to cater for the issue by a narrow interpretation of concepts. In addition, given that the AI Act is a first attempt at regulating AI, it seems likely that there will be amendments soon, as new issues will arise which have not been covered, and any provisions proving impracticable may be reviewed. In order to win public trust in 'trustworthy AI', as the AI Act aims to do, doing more may appear better than doing too little.

50 Van de Sloot, Regulating Synthetic Society, p. 151.

51 Van de Sloot, Regulating Synthetic Society, p. 152.

The aim of winning trust may also offer the answer to the question as to whom the disclosure obligation would be addressed, the general public, the person depicted or the platform on which it was posted?[52] Creating trust would require as much transparency as possible, so the addressee should be the general public: knowledge appears better than a vague feeling that something may be manipulated. Here, however, the transparency obligation appears to fall short of what is needed for trust: people may wish to not only know, or able to find out, that there was manipulation, but what exactly has been manipulated. Such information would involve further issues, namely of data protection, freedom of art, business secrets etc., which have not been addressed here.[53]

The same applies to AI in emotion recognition and biometric categorization (Art. 50 (3) AI Act):

> *'Deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system, and shall process the personal data in accordance with EU data protection law, namely the GDPR*[54] *(Art. 50 (2) AI Act). This obligation shall not apply to AI systems used for biometric categorisation and emotion recognition, which are permitted by law to detect, prevent or investigate criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, and in accordance with Union law.'*

Transparency is also obligatory regarding deep fakes, etc., unless exceptions apply (Art. 50 (4) AI Act):

> *'Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. …*
> *Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offences or where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content.'*

Apart from criminal prosecution, the last category will be most relevant to public administration. Here the exception of human review of the AI-generated content will often apply, however, in mass procedures without any

52 Van de Sloot, Regulating Synthetic Society, p. 152.
53 Cf. Van de Sloot, Regulating Synthetic Society, p. 152.
54 GDPR, fn. 4, and Regulation (EU) 2018/1725 and Directive (EU) 2016/680.

discretion to be exercised text-generating AI may well be used – with the transparency requirement.

4. Minimal-Risk AI: Permitted with no restrictions

Minimal-Risk AI appears in every-day-use of IT and encompasses the large majority of AI systems. It includes namely AI-supported video games or spam filters. Usually the use of such minimal-risk AI is free of charge. It poses only a minimal or no risk to citizens' rights or public security, and thus requires no licensing, but only obligatory disclosure. Namely spam filters will also be used in public administration.

V. Sanctions

Sanctions are provided in Art. 99 et seq. AI Act. Given the limited criminal prosecution competences of the EU, as well as the limited resources for any relevant prosecution authorities, these sanctions are administered by Member States, and may be administered against civil service staff in cases of violation as well. Civil service violations of AI Act rules may often, due to the danger of public administration violations and the role of public administration as a model, be regarded as grave. Containing risks stemming from AI will depend on measures taken by liberal democracies themselves, in order to preserve their very nature.

G. Regulatory Gaps and Shortcomings: Some Conclusions

The AI Act is treading on largely unexplored legislative ground, although the risk-based approach has been tried and tested before. There still remains a lot to explore, namely the fleshing out of indeterminate concepts, such as 'physical or psychological harm', 'materially distorting a person's behaviour' or 'unfavourable treatment'.[55] The annexes will certainly be helpful to reduce legal uncertainty there, as well as case law from other areas. Still, AI-specific terminology and concepts remain in flux, and the challenge of making sense of them will be on legal practice and public administration now.

Similarly, it remains to be seen whether concerns on the AI Act being too restrictive, imposing too far-reaching duties of transparency,[56] too one-

55 Examples from van de Sloot, Regulating Synthetic Society, p. 149.
56 Above IV.4.

rous burdens of quality and risk management,[57] prove correct. For the time being, legal applications may have to cater for the issue by a wider or narrower interpretation of concepts in line with the aims of the AI Act. In addition, given that the AI Act is a first attempt at regulating AI, it seems likely that there will be amendments soon, as new issues will arise which have not been covered, and any provisions proving impracticable may be reviewed. In order to win public trust in 'trustworthy AI', as the AI Act aims to do, doing more may appear better than doing too little. At the current stage it is still too early to offer a considered view as to whether the AI Act is too restrictive for innovation or too lenient regarding fundamental rights.

The Commission's professed aims in presenting its proposal were, in line with the overall aim of offering 'trustworthy AI', to ‚guarantee the safety and fundamental rights of people and businesses, while strengthening investment and innovation across EU countries.'[58] As stated above, relevant fundamental rights include protection of personal data and privacy (Art. 7/8 FRC), intellectual property rights (Art. 17 (2) FRC), and non-discrimination, in particular on grounds such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation (Art. 21 FRC). Apart from the prohibition of unacceptably risky AI, the AI Act, with its risk-based approach, rather invokes procedural than substantive measures, which amounts to setting guardrails on AI use, in particular in public administration. Still, Member States, being fully and directly bound by fundamental rights, may use the leeway given in this respect by the silence of the AI Act in order to add to the requirements on AI use in public administration. The enforcement of fundamental rights by substantive guardrails coming from a firm legal framework is still largely missing.

Another shortcoming of the AI Act may be the lack of (open?) alignment with the Council of Europe Framework Convention on Artificial Intelligence[59], which, however, carries a later date than the AI Act. The

57 Above III.2.b).

58 European Commission webpage, Shaping Europe's digital future/AI Act, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

59 The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law was adopted on 17 May 2024 by the Committee of Ministers of the Council of Europe at its 133th Session held in Strasbourg, and was opened for signature on the occasion of the Conference of Ministers of Justice in Vilnius (Lithuania) on 5 September 2024, see the Council of Europe's Committee on Artificial Intelligence (CAI) website at https://www.coe.int/en/web/artificial-intelligence/cai.

Council of Europe Convention is not even mentioned in the AI Act although it was elaborated in parallel, and although all EU Member States are also members of the Council of Europe. It appears to share the risk based approach, being 'based on the severity and likelihood of a negative impact on human rights, democracy and the rule of law by AI systems'. How far the Convention may supplement or require amendments of the EU AI Act must be left to a later piece of research.

In terms of transparency and accessibility of the law, the AI Act leaves a lot to be wished for. The Annexes may be helpful, however, at the same time the technique of referencing to other provisions and annexes, used instead of reiterating content, has been carried very far from a user's perspective, namely from the perspective of lawyers in practice and administration to apply the AI Act. Parts of it, in particular the chapter on high-risk AI,[60] are hardly readable or understandable to anyone but legal experts with technical knowledge. Legally the referrals may make the Act coherent by avoiding any inconsistencies. Still, more legibility might improve its transparency and effectiveness. After all, it is meant to be applied, and to shape public perception of AI in the EU. From this perspective less legal perfection and more accessibility to the wider public, be they deployers or subjected to AI use, would have been preferable.

Lastly, the AI Act's risk-based approach results in an outsourcing of monitoring the fulfilment of AI requirements. The assessment of the results of the precautions based only on a risk-based regulation will to a considerable extent be left such private players and their documentation and communication. Will such private players, often including large, non-EU companies, act responsibly, or are they going to use their power just because they can, neglecting human rights, rule of law and democracy if they happen to stand in their way? Notwithstanding this, it is arguably still the EU and the states, rather than profit-driven companies wishing to market AI, that are better equipped to balance the various interests of freedom of business and of research in the possibilities of AI, against fundamental rights, democracy and the rule of law, at least as soon as the technology has evolved more.

List of References

Breng, Jonas, Spiegel-Interview: „Starhistoriker Yuval Harari: »KI hat das Potenzial, uns zu versklaven«", 23/10/2024, https://www.spiegel.de/geschichte/starhistoriker-yuval-harari-ki-hat-das-potenzial-uns-zu-versklaven-a-70587486-9be1-4c7c-a1ae-a83c743e74ef.

60 Above III.2.

Council of the European Union, from: Presidency to: Delegations, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Presidency compromise text, Brussels, 29 November 2021, No. Cion doc.: 8115/20.

Dignum, Virginia, Progressive Post, 15/06/2023: Future-proofing AI: regulation for innovation, human rights and societal progress, https://feps-europe.eu/future-proofing-ai-regulation-for-innovation-human-rights-and-societal-progress.

European Commission, Commission Strategy and Policy Priorities 2019 – 2024, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en.

European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, Brussels, 21.4.2021, COM(2021) 206 final.

European Commission webpage, Shaping Europe's digital future/AI Act, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

European Parliament 2019-2024, Artificial Intelligence Act, Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), (Ordinary legislative procedure: first reading) P9_TA(2023)0236.

Gates, Bill, The Bill Gates Blog, A new era – The Age of AI has begun, 21st March 2023, https://www.gatesnotes.com/The-Age-of-AI-Has-Begun.

Van de Sloot, Bart, Regulating the Synthetic Society - Generative AI, Legal Questions and Societal Challenges, Oxford et al. 2024 (available from https://library.oapen.org/handle/20.500.12657/88178).