AI Act for the Working Programmer*

Holger Hermanns[1], Anne Lauber-Rönsberg[2],
Philip Meinel[2], Sarah Sterz[1], and Hanwei Zhang[1]

[1] Saarland University, Saarland Informatics Campus, Saarbrücken, Germany
{hermanns, sterz, zhang}@depend.uni-saarland.de
[2] TU Dresden University of Technology, Institute of International Law, Intellectual Property and Technology Law, Dresden, Germany
{anne.lauber-roensberg, philip.meinel}@tu-dresden.de

Abstract. The European AI Act is a new, legally binding instrument that will enforce certain requirements on the development and use of AI technology potentially affecting people in Europe. It can be expected that the stipulations of the Act, in turn, are going to affect the work of many software engineers, software testers, data engineers, and other professionals across the IT sector in Europe and beyond. The 113 articles, 180 recitals, and 13 annexes that make up the Act cover 144 pages. This paper aims at providing an aid for navigating the Act from the perspective of some professional in the software domain, termed "the working programmer", who feels the need to know about the stipulations of the Act.

1 Introduction

After extensive deliberations, the European Union has taken the final step for adopting the AI Act [10]. The AI Act aims to ensure the development and deployment of safe and trustworthy AI by relying on a risk-based approach – the higher the risks to fundamental rights and society, the stricter the legal requirements.[1] However, the demarcations of the regulated areas of AI often seem blurred. The idea of this paper is therefore to provide the "working programmer"[2] with some initial help in navigating the complexities of the AI Act. In doing so, we make three main contributions:

- We provide an overview of the regulated AI technologies and how to distinguish between them. This is essential for the working programmer to determine which legal obligations under the AI Act might apply to their work.
- We map the relevant obligations to help the programmer understand which parts of the AI Act may be relevant to them. This is supported by a flowchart that helps to find the relevant obligations with simple questions and to narrow down the complexities of the AI Act.

* Authors are listed in alphabetic order.

[1] Of course, the AI Act is also not the only law that governs the development and use of AI systems. In addition to the AI Act, other general or sector-specific laws such as the GDPR, the Digital Services Act, anti-discrimination laws, sector-specific legislation such as legislation governing medical devices, to name just a few, must be observed.

[2] In allusion to "ML for the Working Programmer" by Larry C. Paulson [22].

– Finally, we shed light on the question of in how far programmers can make use of ready-made general-purpose AI models, such as large language models, that they want to integrate into their AI system. We do this in an interdisciplinary effort of computer science and law to help the working programmer understand and anticipate legal risks.

It appears important to note at this point that there are still many legal uncertainties, as the AI Act is the first of its kind, developed without a blueprint. Most of the legal requirements need further interpretation, which is beyond the scope of this paper. In addition, many requirements in the Act are intentionally broad in order to be applied to a variety of cases or to be narrowed down by standardisation organisations. In this respect, future work by these organisations is expected to contribute significantly to clarification. Therefore, the goal of this paper can only be to provide the programmer with an overview to enable them to navigate the AI Act effectively.

Organization of the paper. We start off by characterising the primary audience for this paper in Section 2. We then address the types of AI regulated under the Act in Section 3, providing concrete examples to illustrate various distinctions. Section 4 discusses the scope of the AI Act, including its territorial reach and specific exemptions. We then in Section 5 move on to the obligations for providers of general-purpose AI models, before we turn to the various obligations for providers of (specific-purpose) AI systems in Section 6. Section 7, considers the practical implications of building an AI system on top of a GPAI model. Finally, we conclude in Section 8 with a summary of the AI Act's impact on programmers.

2 Addressee of the Paper

The particular relevance of the AI Act for the working programmer arises from the fact that it not only governs the use of AI systems, but primarily sets out requirements *for their development*. The AI Act addresses a variety of stakeholders along the AI value chain, among them "deployer", "provider", "distributor", and "importer". Against this backdrop, the working programmer will most likely be considered (being part of) a "provider" of AI systems or AI models and this role will therefore be the focus of the further analysis. The AI Act defines a provider as any natural or legal person, public authority, agency or other body that develops an AI system or a general purpose AI model or that has an AI system or a general purpose AI model developed and places them on the market or puts the system into service[3] under its own name or trademark, whether for payment or free of charge.[4]

In other words, any entity responsible for the development of a system or model within the scope of the AI Act could be affected by the requirements of the Act if it

[3] Note that 'placing on the market' means the first making available of an AI system or a general purpose AI model on the Union market, Article 3(9) [10]. 'Putting into service' means the supply of an AI system for first use directly to the deployer or for own use in the Union for its intended purpose, Article 3(11) [10].

[4] Article 3(3) [10].

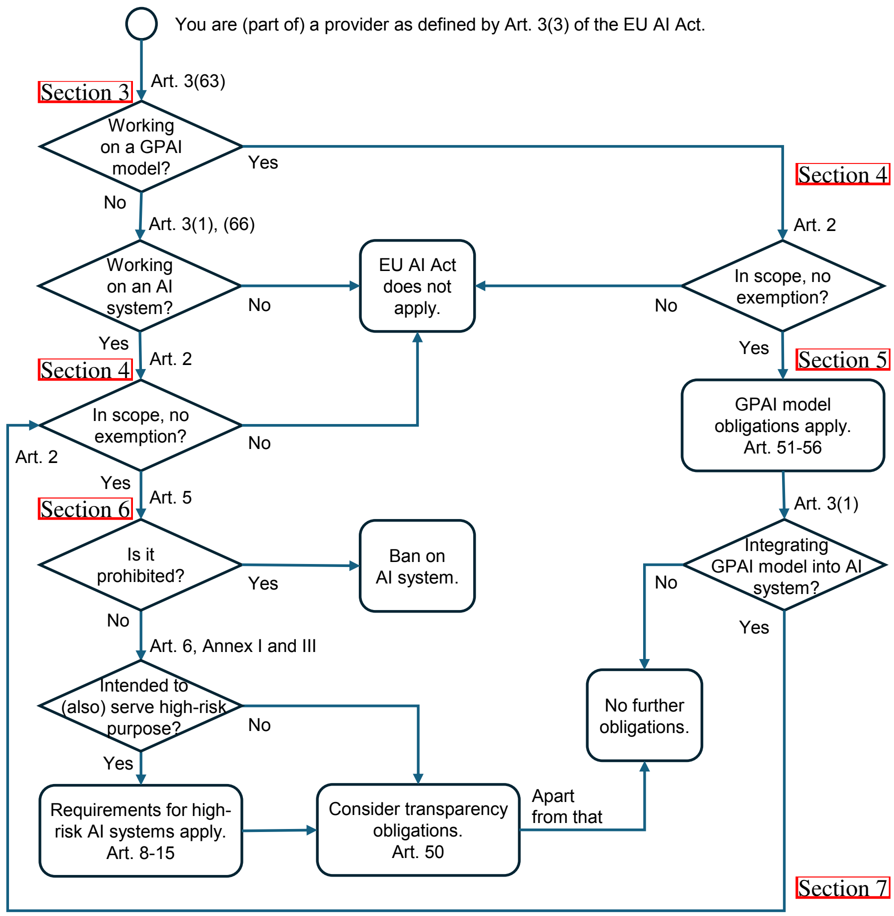

Fig. 1. The working programmer navigating the AI Act

makes the system available to third parties or uses it for its own purposes. However, it is important to note at this point, that not only the development of a new AI system can give rise to the provider's obligations under the AI Act. Rather, it may be sufficient to significantly modify an existing system in order to be considered a provider within the meaning of the AI Act.[5]

The flowchart in Fig. 1 takes the programmer's role as a provider under the AI Act as a starting point. The subsequent branches help to determine the resulting obligations by means of yes/no questions. They also reflect the further structure of this paper and can therefore be used as a reading aid.

[5] This is the case if an AI system is considered to be high-risk after the modifications made by the programmer, cf. Article 25(1) lit. b), c) [10]. For the definition of 'substantial modification' see Article 3(23) [10].

The usual role of the working programmer facing AI in their daily work is that of (being part of) a provider in the sense of the AI Act, and this is what the subsequent discussion presupposes.

3 What types of AI are regulated under the AI Act?

Compliance with the AI Act is of course only an issue if the AI Act applies to the projects and tasks of the working programmer.

3.1 What characterises AI according to the AI Act?

The definition of "AI systems" in Article 3(1) [10], which we will examine in Section 3.2, contains some insights on what technical approaches the AI Act considers as "AI" and which it does not. As a key characteristic of AI systems, the AI Act underlines their capability to infer how to generate its output, which "transcends basic data processing by enabling learning, reasoning or modelling".[6] The techniques that enable such inference "include machine learning approaches that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved.'[7] This seems to be what, according to the AI Act, distinguishes AI from "simpler traditional software systems or programming approaches".[8] The AI Act thus does not apply to "systems that are based on the rules defined solely by natural persons to automatically execute operations".[9] This means that – at least to our understanding – the final version of the AI Act does not cover traditional rule-based systems written by humans, even if they are complex and their deployment is associated with high risks.[10] This approach has however sparked criticism among legal scholars who advocate a broader scope of application of the AI Act.[11]

Example 1. The following list contains some concrete and distinguishing examples of AI based on our interpretation of the criterion of whether it "infers, from the input it receives, how to generate outputs".[12]

[6] Recital 12 [10].
[7] Recital 12 [10].
[8] Recital 12 [10].
[9] Recital 12 [10].
[10] A notable exception to this general observation might be expert systems that are not trained, but draw inferences in an elaborate way from an extensive knowledge base, cf. Recital 12 [10]. However, this will need to be established in future legal interpretation.
[11] E.g. Krönke [16, p. 529-530] argues that it was an open question, to the detriment of legal certainty, whether and from what level more complex rule-based systems also fall under the Regulation's definition of AI. He advocates from a teleological point of view, that the AI Act should also be applicable to such systems, as the use of complex rule-based systems can also lead to risks typical of AI, such as a lack of transparency.
[12] Article 3(1) [10].

1. **Not AI – a complex but traditional system:** A typical compiler for a high-level programming language – such as Java – is not considered to be an AI system, regardless of its (potentially excessive) logical complexity. *This system will fall outside the scope of the AI Act, because it does not infer* how *to decide its outputs.*
2. **AI – a machine-learned system:** A machine-learned system that decides which refugees should be granted asylum is AI, if the method by which it decides these cases is not directly given in a relevant sense but is machine-learned from data. *This system will fall inside the scope of the AI Act, because it uses inference on* how *to decide its outputs.*
3. **AI – a traditional system that infers how to decide its outputs:** A version of the asylum-system that is not machine-learned but purely logic-based (such as an automated reasoner) is still considered AI, namely if it does not directly infer the output from the input but infers *how to decide* whether an asylum seeker should be granted asylum. So, it first needs to infer a way whereby to arrive at an output for the given input, in this case a way by which to decide for a given asylum case (input) whether the asylum seeker should be granted asylum (output). The output is not given directly, but only after the system applied the method it inferred previously to the input. *This system will fall inside the scope of the AI Act, because it infers* how *to decide its outputs.*
4. **Not AI – a traditional system that does *not* infer how decide its outputs:** If instead the asylum-system is a logic-based system that decides asylum cases by directly inferring the decisions with a human-defined set of rules, it will not be AI. *This system will fall outside the scope of the AI Act, because it does not infer* how *to decide its outputs.*

The difference between (3) and (4) may seem like a fine line to distinguish at first sight, but it would make the difference between falling under the AI Act and not doing so. These considerations might not be relevant to most practical applications, though, since to us it appears difficult to imagine that purely logic-based systems of the kind discussed in (3) can exist in reality.[13] Logic-based systems notoriously employ human-made semantics and therefore do not infer how to generate outputs, but instead generate outputs in a direct (though possibly complex) way. Therefore, the system in (3) would be considered to be much like the compiler from (1).

Overall, the inference of *how* to arrive at a conclusion seems to be a crucial capability of an AI system according to Article 3(1) [10] of the AI Act. Nevertheless, it arguably will be a point of legal and technical controversy in the future. This is especially true considering that Recital 12 [10] names "logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved" as an example of approaches that enable the kind of inference needed to constitute an AI system. However, it can broadly be argued that every program ever written uses logic-based approaches that infer from a symbolic representation of the task to be solved. So, obviously, a more narrow reading is needed instead. In our opinion, this should pivot on inferences regarding *how* to generate outputs, as explicitly stipulated in Article 3(1) [10].

[13] We surveyed a non-representative set of logicians for opinions and the reactions were not supporting the idea that system as in (3) can exist.

This discussion indicates that the recipients of the AI Act are left with uncertainty regarding what is and is not AI according to the Act. However, for many practical cases it will arguably be obvious whether something is AI or not, because most systems in question will use some form of learning from data, and will therefore clearly be AI. This is why many working programmers will probably not have to deal with the above discussed demarcation issues in their work.

3.2 What different types of AI does the Act regulate?

Within this general classification, the AI Act distinguishes three addressed forms of AI:

GPAI models "GPAI models" are defined as AI models (including where such AI models are trained with a large amount of data using self-supervision at scale) that display significant generality, are capable of competently performing a wide range of distinct tasks, and can be integrated into a variety of downstream systems or applications.[14] The AI Act regards these GPAI models as a fundamental component for subsequent use cases. Therefore, the legal obligations for these GPAI models arise when they are placed on the market, regardless of how this is done, e.g., through libraries, application programming interfaces (APIs), as a direct download, or as a physical copy.[15] GPT-4, which serves as the basis for ChatGPT as well as a number of downstream applications, is probably the most relevant example of such a GPAI model.

AI systems Secondly, an "AI system" is defined as "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 receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".[16] With other words, AI systems are the systems that use AI and that may be deployed to end users to achieve explicit or implicit objectives, like the asylum decision system in Example 1. Such AI systems can be either built from scratch or on top of a ready-made GPAI model.

GPAI systems In general, AI systems that are built on the basis of ready-made GPAI models, qualify as regular AI systems that serve a set of explicit or implicit objectives. In some cases, however, AI systems that are built on the basis of GPAI models may have the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems. As a special form of AI systems, the AI Act defines such systems as general-purpose AI systems.[17] ChatGPT or Google Gemini might be the most prominent examples of such GPAI systems.

3.3 Relating and explaining the different types of AI

In order to understand the general scope of the AI Act, it is essential to relate the different forms of AI to each other in light of the drafting process.

[14] Article 3(63) [10].
[15] Recital 97 [10].
[16] Article 3(1) [10].
[17] Article 3(63) [10].

The special role of General Purpose AI. The AI Act originally only aimed to regulate AI that is capable of performing a distinct range of tasks for a limited number of purposes, e.g. job recruitment decisions or biometric identification systems. However, due to the (then) surprising advent of multimodal large language models in late 2022, spearheaded by the GPT family of models, the EU felt pressured to adapt this approach. In an attempt to comprehensively regulate AI along the entire value chain, general-purpose AI was added to the scope of regulation. General-purpose AI, as described above, is therefore considered to be a special form of before regulated AI, which were only capable to serve some distinct purposes.

AI systems and AI models. Secondly, the AI Act distinguishes between the technical applications of ”AI models” and ”AI systems,” both of which may also serve general purposes. However, the clear distinction between both is controversial, not least because the legislature has failed to explicitly define "AI model" in the legal text.[18] In principle, the AI model can generally be described as algorithms or statistical models that are designed to perform a variety of tasks without being directly usable by end users. It only becomes usable by integrating the AI model into an AI system, which incorporates the AI model and combines it with the components needed to deploy the AI model, such as a user interface.[19] Thus, AI models can be understood as an essential part of the AI system, and AI systems can be understood as final product that may be deployed to end users. For instance, in summer 2024 the chatbots ChatGPT and Microsoft Copilot would be an AI *systems* wrapping the GPAI *model* GPT-4 of Open AI.

AI Systems based on GPAI Models. AI systems built on the basis of ready-made GPAI models will arguably play an important role in the industry. The EU acknowledges this fact, even though AI systems based on GPAI play a less accentuated role in the AI Act. It is difficult to develop a stringent classification of these systems because the AI Act partly relates to the system's *capability* to perform a variety of purposes and partly to the *objectives* pursued by the system as defined by the provider (which might be narrower than the system's capabilities and differ from the purposes pursued by the deployer). Accordingly, these AI systems may serve one or more specific objectives, and therefore be considered regular ”AI systems” under the AI Act. Possible challenges from this classification will be addressed in Section 7.1. In other cases, AI systems that are built on the basis of GPAI models may have the capability to serve a variety of purposes, and therefore qualify as general-purpose AI systems.[20] The only obligation imposed not only on providers of AI systems, but also explicitly on providers of GPAI systems, is the transparency obligation under Article 50(2) [10] to ensure that synthetic content can be recognized as artificially generated or manipulated. Beyond that, it is not entirely clear under which conditions further stipulations by the AI Act apply to such GPAI systems. This is particularly relevant for GPAI systems that can be used directly

[18] However, AI models that do not serve general purposes are mentioned in various places in the legal text (cf. Recital 97 [10], Article 3(63) [10].

[19] Recital 97 [10].

[20] Article 3(63) [10].

	AI models	AI systems
Specific purpose AI	SPAI model	SPAI system
General purpose AI	GPAI model	GPAI system

Table 1. AI models and systems for specific and general purposes.

by deployers for at least one purpose that is to be classified as high-risk[21] or even a prohibited practice [22]. These considerations will be the topic of Section 7.2.

Introducing the term "SPAI". As described above, the AI Act distinguishes AI along two dimensions: AI model or AI system, and its purpose. Table 1 illustrates these categories. To streamline the wording, we will introduce the term *specific-purpose AI* (or *SPAI* for short). We use this to denote all AI that is not general purpose. The Act itself does not use this term. However, "AI system" is by default meant to encompass both SPAI and GPAI. This is apparent when considering that GPAI systems are defined as a special form of AI systems in Article 3(66) [10]. Moreover, while AI models that do not serve general purposes are mentioned in various places in the legal text (cf. Recital 97 [10] or Article 3(63) [10]), they are not directly subject to obligations under the AI Act. Nevertheless, there may be indirect regulation through the system requirements for high-risk AI systems and the transparency regulation, as they also anticipate development decisions for the underlying models. We therefore consider the term SPAI valuable for understanding the general demarcations of the AI Act.

The AI Act addresses AI systems and models capable of inferring how to generate outputs, including machine learning and logic- or knowledge-based approaches, but excludes traditional rule-based systems. It mainly distinguishes between General Purpose AI (GPAI) models, which perform a wide range of tasks, and AI systems, which may or may not use these models for specific or general purposes.

4 Scope of Application

Other than the regulated forms of AI, the AI Act makes some further stipulations on its scope of application.

4.1 Territorial Scope

The territorial scope of application of the AI Act is very broad. Every AI system and GPAI model that is placed on the market or put into service within the EU has to comply with the AI Act – regardless of whether the provider has its establishment in the EU or

[21] Article 6 [10].
[22] Article 5 [10].

is located there.[23] This so-called market location principle is well known from other legislations, such as the GDPR.

Furthermore, the AI Act also applies when not the AI system as such, but only the output generated by the AI system is used in the EU. The aim of this provision is to prevent the circumvention of the Regulation and to ensure an effective protection of persons located in the EU.[24] For example, when a company established in the EU contracts services in the field of staff recruitment to an operator established in a third country, an AI system deployed during this process would need to comply with the requirements of the AI Act.[25] For providers in third countries, this means that the AI Act must also be observed when developing AI systems whose results are to be used within the EU. However, this presents the challenge that it is necessary to anticipate during the development of the AI system whether the outcome will later be used in the EU. In addition, in many cases it may be difficult to determine whether a particular outcome originates from an AI system or was generated by humans.

Even if a GPAI model is developed outside the EU, the AI Act applies to it if it is put on the market or into service there. Even if an AI system provides its service outside the EU, the AI Act still applies if that service's output is used in the EU.

4.2 Exemptions

For certain high-risk AI systems in fields such as civil aviation and motor vehicles, specific provisions have been set out by other EU legislation. Therefore, the AI Act as such does not apply in these cases, e.g. to autonomous cars, but instead only the specific rules for the case apply (cf. Article 2(2) [10] and Annex I, Section B [10]). There are also other areas in which the AI Act does not apply:

- AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development (Article 2(6) [10]).
- Research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service.[26] However, testing in real world conditions shall not be covered by that exclusion (Article 2(8) [10]).
- AI systems released under free and open-source licences, unless they are placed on the market or put into service as high-risk AI systems or as an AI system that enables prohibited Artificial Intelligence Practices according to Article 5 [10]. Furthermore, the transparency obligations under Article 50 [10] still apply (Article 2(12) [10]).
- Military, defence or national security purposes (Article 2(3) [10]).

[23] Article 2(1)(a) [10].
[24] Article 2(1)(c) [10].
[25] See also Recital 12 [10].
[26] For some reason GPAI models that are used for research, development or prototyping activities before they are placed on the market are also explicitly excluded as according to Article 3(63) [10].

– Public authorities in a third country and international organisations that use AI systems for law enforcement and judicial cooperation, provided that such a third country or international organisation provides adequate safeguards with respect to the protection of fundamental rights and freedoms of individuals (Article 2(4) [10]).

There are some exemptions in Article 2 [10] to the applicability of the AI Act that may be sector-specific or that may relate to the intended uses, such as scientific research and development.

5 Requirements for the Provider of GPAI models

For developers of GPAI models, the AI Act follows a two-tiered approach, which is divided into general requirements and additional requirements for GPAI models with systemic risks. Providers of GPAI models with or without systemic risks should demonstrate compliance with these requirements by applying harmonized standards or – until corresponding standards have been published – by complying with codes of practice.[27] The latter are voluntary codes that are meant to be developed with the help from the AI Office, and can be considered as lighter versions of technical standards.

5.1 General Requirements regarding GPAI models

Article 53 [10] primarily contains general obligations regarding the need for documentation of the GPAI model. The aim of these obligations is to simplify the use of GPAI models for downstream AI systems. In the view of the legislature, this requires a good understanding of the models used in order to enable integration and fulfil the downstream provider's obligations under the AI Act and other regulations.[28] This includes:

– draw up and keep up-to-date the technical documentation of the model, including its training and testing process and the results of its evaluation;[29]
– draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems;[30]
– put in place a policy to comply with Union law on copyright and related rights and a possible reservation of rights of the copyright holders;[31]

[27] Article 53(4), 55(2), 56 [10].
[28] Recital 101 [10].
[29] Article 53(1) lit. a) [10], this should contain, at a minimum, the information set out in Annex XI [10]. This technical documentation is intended to be non-public and should only be provided, upon request, to the AI Office.
[30] Article 53(1) lit. b) [10], the information provided should enable the downstream developers to comply with possible obligations of the AI Act, e.g. compliance with high-risk obligations. It should contain, at a minimum the information set out in Annex XII [10].
[31] Article 53(1) lit. c) [10].

– draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.[32] The aim is to enable authors and other right holders to assess whether their rights have been infringed. The summary should e.g. list the main data collections or sets that went into training the model, such as large private or public databases or data archives, and provide a narrative explanation about other data sources used.[33]
– appoint an authorised representative which is established within the Union.[34]

The first two obligations shall not apply to GPAI models that are released under a free and open-source licence if the model isn't associated with systemic risks (see next point).[35] Additionally, information should be made available to those potential downstream providers who intend to use the GPAI model for an AI system. In any case, the provider of the GPAI model should be protected by the recipient's confidentiality obligations.[36]

Article 53 [10] imposes documentation and transparency obligations on general purpose AI model providers in order to facilitate the integration and compliance of downstream AI systems.

5.2 Additional requirements regarding GPAI models with systemic risks

In case the GPAI model is associated with a "systemic risk" the AI Act imposes some additional obligations on the programmer. According to Article 51(1) [10], GPAI models are associated with systemic risk if they either have high-impact capabilities or are considered equivalent by the Commission. Whether a model has high-impact capabilities shall be "evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks", and is presumed when the cumulative amount of computation used for its training measured in floating point operations is greater than 10^{25}.[37] If a GPAI model is associated with systemic risks, the provider is obliged to:

– Perform a model evaluation described in accordance with standardised protocols and tools reflecting the state of the art;[38]
– assess and mitigate possible systemic risks at Union level;[39]
– keep track of, document, and report relevant information about serious incidents and possible corrective measures to address them;[40]

[32] Article 53(1) lit. d) [10].
[33] Recital 107 [10].
[34] Article 54(1) [10].
[35] Article 53(2) [10], this is only the case if the license allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, the information on the model architecture, and the information on model usage, are made publicly available.
[36] Article 53(7), 78 [10].
[37] Article 51(1) a), (2) [10].
[38] Article 55(1) lit. a) [10].
[39] Article 55(1) lit. b) [10].
[40] Article 55(1) lit. c) [10]; "serious incidents" are defined in Article 3(49) [10].

- ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risk and the physical infrastructure of the model.[41]

As soon as a GPAI model meets the technical requirements for systemic risks, the provider must notify the Commission within two weeks at the latest.[42] In its notification, the provider can also put forward arguments that the GPAI model does not present any systemic risks despite the technical circumstances.[43]

If a general-purpose AI model poses "systemic risks", as defined by high-impact capabilities, the AI Act additionally requires standardised evaluations, risk assessments, cybersecurity measures, incident reporting, and timely notification to the Commission.

6 Requirements for the Provider of AI Systems

If the developed AI system falls within the scope of application, three main types of relevant legal ramifications can arise from the AI Act: the system may (1) be prohibited, (2) be considered high-risk, or (3) be none of the two.

6.1 Prohibited systems

The AI Act defines a number of AI systems that are associated with what it considers to be unacceptable risks. These systems are prohibited under Article 5 [10].[44] The catalogue of AI systems listed here encompasses certain use cases:

- Subliminal techniques that have the objective or effect of materially distorting a person or group of persons' behaviour causing them or others significant harm.[45] This could affect the development of recommendation systems used in social media or advertising. The question that will arise here is when a distorting effect on one's own behavior reaches a material level. This also applies to other systems that exploit any vulnerability of natural persons due to their age, disability or a specific social or economic situation;[46]
- Systems for the evaluation or classification of natural persons or groups based on their social behavior or personality characteristic with a social score leading to detrimental or unfavourable treatment.[47] This refers to social evaluation systems that result in detrimental or unfavorable treatment in certain contexts that is either unrelated to the contexts in which the data was generated or disproportionate to the behavior of the individuals;

[41] Article 55(1) lit. d) [10].
[42] Article 52(1) [10].
[43] Article 52(2) [10].
[44] The systems listed there may be already prohibited by data protection law, Union law or the law of the Member States, for Union law explicitly Article 5(8) [10].
[45] Article 5(1) lit. a) [10].
[46] Article 5(1) lit. b) [10].
[47] Article 5(1) lit. c) [10].

– Systems for risk assessment of natural persons to predict the risk of a person committing a criminal offence based solely on profiling or assessing personality traits and characteristics;[48]
– Creating or expanding facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;[49]
– Systems to infer emotions of a natural person in the workplace or education institutions except for medical or safety reasons;[50]
– Biometric categorisation systems that categorise individually natural persons based on biometric data to infer protected attributes or characteristics.[51] This does not apply to the labelling or filtering of lawfully acquired biometric datasets or categorising of biometric data for law enforcement.

The placing on the market, putting into service or use of the aforementioned AI systems is prohibited. In addition, only the *use*, but not the development, of "real-time" biometric identification systems in publicly accessible spaces for law enforcement is generally prohibited, yet subject to broad exceptions.[52]

Article 5 [10] prohibits AI systems with unacceptable risks that cause significant harm by distorting behavior, exploiting vulnerabilities, using social scoring, predicting criminal behavior based solely on profiling, scraping facial images, inferring emotions outside of medical or security contexts, or categorizing individuals based on biometrics, while also restricting real-time biometric identification in public spaces with exceptions.

6.2 High-risk AI systems

If AI systems are not prohibited by the AI Act, it is generally permissible to market such systems. However, the AI Act imposes specific requirements on the development of those AI systems that are deemed to pose high risks. There are two ways in which an AI system can be considered high risk (also cf. Figure 2).

High-risk AI systems covered by Annex I The first option is found in Article 6(1) [10], which states that the AI system is considered high risk if two conditions are met. First, the AI system must serve as a safety component of a product covered by one of the harmonized EU acts listed in Annex I [10] or be such a product itself.[53] Second, the AI system must be subject to conformity assessment by a third party with a view to its placing on the market or putting into service on the basis of those acts listed in Annex I [10].

[48] Article 5(1) lit. d) [10].
[49] Article 5(1) lit. e) [10].
[50] Article 5(1) lit. f) [10].
[51] Article 5(1) lit. g) [10].
[52] Article 5(1) lit. h), (2)-(7) [10].
[53] "Safety components" are defined in Article 3(14) [10] as a component of a product or of an AI system which fulfils a safety function for that product or AI system, or the failure or malfunctioning of which endangers the health and safety of persons or property.

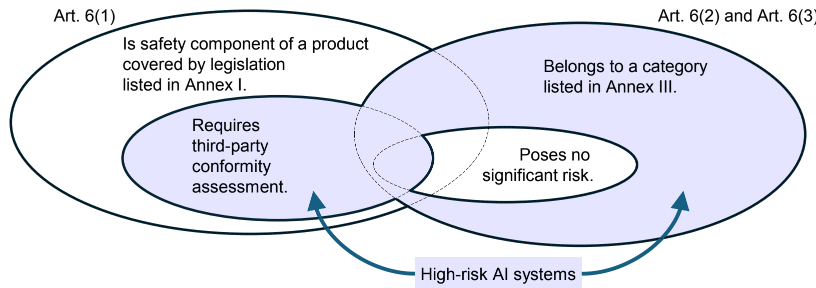

Fig. 2. Two avenues for an AI system to be classified as high-risk, represented as a Venn diagram.

Annex I [10] contains a large collection of EU legislation related to product safety law, e.g. the Medical Devices Directive, the Machinery Regulation and the Toys Directive. It is important to emphasise, however, that the AI system is not automatically considered high risk if the relevant product falls under one of these legal acts. This is because a third-party conformity assessment is only required in certain cases within the respective legal acts. For instance, the above-mentioned Machinery Directive regulates the safety of machinery placed on the market. Its scope is therefore quite broad. However, only some, but not all of the machinery mentioned need to undergo a third party assessment.[54] For this reason, the final classification of AI systems in these cases can only be determined on a case-by-case basis, which will likely require specialized legal expertise.

Under Article 6(1) [10], an AI system is considered high-risk if it serves as a safety component of a product under harmonized EU legislation listed in Annex I [10] and requires third-party conformity assessment for its market placement or putting into service, though this has to be determined on a case-by-case basis.

High Risk Systems covered by Annex III Secondly, pursuant to Article 6(2) [10] an AI system is considered to be high-risk if it is listed in Annex III [10]. The classification of these systems is more accessible than the list in Annex I [10] because general areas of application are outlined (points 1-8) and within each of these points specific use cases are listed exhaustively. Accordingly, AI systems used for the following use cases are generally considered high risk:

– Biometrics, including AI systems used for remote biometric identification, biometric categorisation according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics, or emotion recognition;

[54] Included are, for example, vehicle servicing lifts or, more relevant, safety components with fully or partially self-evolving behaviour using machine learning approaches to ensure safety functions, cf. Art. 25(1), (2) and Annex I Part A (3), (5) of the Machinery Regulation [9].

- AI systems for the use 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, including systems to determine access to institutions, to evaluate learning outcomes, assessing appropriate levels of education, or for monitoring students during tests;
- Recruitment tools or systems used to make decisions affecting terms of employment, like promotion or termination of contractual relationships, allocation of tasks, or monitoring;
- Access to and enjoyment of essential private or public services and benefits, like healthcare, credits, insurance, or assistance in emergencies;
- Law enforcement, e.g. systems to be used to assess the risk of natural persons becoming victims of criminal offences, to evaluate the reliability of evidence, assess the risk of a person (re-)offending, or detect, investigate or prosecute criminal offences;
- Migration, asylum and border control management, e.g. systems to assist authorities for the examination of applications, or for border control management;
- Administration of justice and democratic processes, including to assist a judicial authority in researching and interpreting facts and the law or applying the law, or systems to influence the outcome of an election or referendum.

There is, however, an exception to this general classification of high-risk AI systems, which was added during the finalisation of the text of the AI Act: even if the system in question falls under one of these categories it shall not be considered high-risk if it doesn't perform profiling of natural persons,[55] and doesn't pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including when the system does not materially influence the outcome of a decision making process.[56] This applies if the system is intended to perform a narrow procedural task, to improve the results of prior human activity, to detect decision-making patterns or deviations therefrom without replacing human judgement, or to perform a preparatory task to an assessment related to the use cases in Annex III [10]. This could become an essential way to avoid the strict obligations that the AI Act defines for high-risk systems. However, the remaining legal uncertainties associated with the wording of these exceptions will need to be reduced in the future by the deciding courts.

According to Article 6(2) [10], an AI system is high-risk if it is listed in Annex III [10] and thus used in sensitive areas such as biometrics, critical infrastructure, education, recruitment, essential services, law enforcement, migration or justice, although some exceptions apply in the absence of significant risk.

Obligations: Safety requirements for high-risk AI systems Compliance with the following safety requirements contained in Articles 9 to 15 [10] must be ensured when placing high-risk systems on the market or putting them into service:

[55] Article 6(3) subpara. 3 [10].
[56] Article 6(3) subpara. 1 [10].

- Establishment, application, documentation and maintenance of a risk management system, Article 9 [10];
- Development and training of high-risk systems with training and validation data according to certain quality criteria, data governance, Article 10 [10];
- Technical documentation, Article 11 [10];
- Automatic recording of processes and events (logging) according to recognized standards, Article 12 [10];
- Transparency and provision of information to the deployer, Article 13 [10];
- Enabling human oversight during the use of the system, Article 14 [10];
- Accuracy, robustness and cybersecurity in relation to the intended purpose of the specific system, Article 15 [10].

In general, the system requirements are very broad, which can lead to difficulties for programmers during implementation. As mentioned above, technical standards being developed by European standardisation organisations are foreseen to make the obligations specific and to give clear guidance to providers. However, it remains to be seen whether these standards will be sufficient to provide a comprehensive and workable set of rules.

Compliance with the safety requirements must be demonstrated by the provider through a conformity assessment procedure.[57] In most cases this procedure will take place as part of an internal control. This is particularly the case if a system is considered a high-risk system within the meaning of Annex III [10] and available technical standards have been applied completely.[58] All other high-risk systems must undergo an external assessment procedure under the above-mentioned harmonized acts, which will in future also verify compliance with the safety requirements of the AI Act.[59]

Furthermore, the provider's responsibility for a high-risk system does not end with the completion of the assessment procedure. Rather, ongoing compliance of the system must be ensured. This includes adjustments to the system if they have reason to consider that the AI system no longer meets the safety requirements.[60] In addition, a post-market monitoring system must be set up and maintained.[61] Finally, the assessment procedure must be repeated if substantial modifications have been made to the system.[62]

For high-risk AI systems, compliance with Articles 9 to 15 [10] mandates a comprehensive risk management system, adherence to quality data standards, technical documentation, automatic logging, transparency, human oversight, and cybersecurity, with initial and ongoing conformity assessments required to ensure safety before and after market entry.

[57] Article 16 lit. f) [10].
[58] Article 43(1), (2) [10], Annex VI [10].
[59] So-called "notified bodies", Article 43(3) [10].
[60] Article 20(1) [10].
[61] Article 16 lit. h), 72 [10].
[62] Article 43(4) [10]; note that self-learning systems do not have to go through a new assessment procedure for changes that were predetermined by the provider.

Obligations resulting indirectly from deployers' obligations? Lastly, it appears possible that further obligations of the provider result indirectly from obligations imposed on the deployer. An example for this phenomenon is rooted in the right to explanation, stipulated in Article 86 [10]. Deployers of high-risk systems within the meaning of Annex III [10] must provide clear and meaningful explanations about the role of the AI system in the decision-making process to affected persons.[63] Although the provision is aimed at deployers and its scope is controversial, it may also indirectly impose obligations on the provider. For example, this could result in an obligation to at least enable an explanation of the decisions of an AI system. This does not seem far-fetched, as the AI Act requires general transparency and explainability of high-risk systems in various instances.[64] However, there is currently very little clarity on what Article 86 [10] will entail for developers.

Deployers of stand-alone high-risk AI systems must clearly explain the AI's decision-making role to affected persons, which may indirectly impose obligations on developers.

6.3 Further transparency obligations for certain AI systems

Lastly, certain AI systems must meet a number of transparency requirements according to Article 50 [10]. Transparency in this context means that programmers must provide specific information to the end user of their system. This information must be provided to the end user in a clear and perspicuous manner no later than at the time of the first interaction or exposure.[65]

On the one hand, this applies to systems that are designed to interact directly with end users.[66] These systems must be designed and developed to inform end users that they are interacting with an AI system if this is not obvious to a reasonably well-informed, observant, and circumspect person.[67] In other words, chatbots, for example, must be designed and developed in such a way that they make it clear to their users that they are bots ("Bot-Disclosure").[68]

On the other hand, generative AI must make clear the artificial origin of its output.[69] This equally applies to AI systems and GPAI models that generate synthetic audio, image, video or text content. The providers of these models must ensure that the output of their systems is labelled in a machine-readable format and can be identified as artifi-

[63] Article 86(1) [10]; "deployers" are defined as any natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity, Article 3(4) [10].
[64] Article 13, 14 [10].
[65] Article 50(5) [10].
[66] Article 50(1) [10].
[67] Note the exception in Article 50(1) [10] for AI systems authorised by law to detect, prevent, investigate or prosecute criminal offences.
[68] Critique to this approach, cf. [27, p. 106].
[69] Article 50(2) [10].

cially generated or manipulated.[70] This obligation does not apply if the system or model makes minor changes to the input data. However, it is unclear at what point a change to an input is considered to be minor, so providers would be better advised to introduce a general label for all outputs.

Certain AI systems must meet transparency requirements according to Article 50 [10], necessitating that systems interacting with users clearly disclose their AI nature and that generative AI outputs are labelled as artificially created or manipulated.

7 Building an AI System on top of a GPAI model

We now aim at reflecting some practical consequences of what we discussed thus far. We consider the question of how to navigate the AI Act if aiming to build an AI system that uses a GPAI model as one of its core functional components. These downstream AI systems can take two forms under the AI Act. They can either be considered SPAI systems that are built on top of a GPAI model if they have specific purposes. Or they can be considered a GPAI system if the final system itself has the capability to serve a variety of purposes.[71]

7.1 Obligations for SPAI systems built on top of a GPAI model

When integrating a GPAI model into an SPAI system that is categorised as high-risk, some compliance issues may arise. To illustrate these issues, let us elaborate a use case. Specifically, we consider the following hypothetical auto-ranking system used by a university admissions committee to decide which applicant should be granted admission to a particular degree program:

Example 2. The AI system "AutoRa" produces a (partial) order of applicants by aggregating pairwise decisions of the form "candidate X is no less suited than candidate Y". The resulting partial order is presented to an admission board, which uses it to decide where to put the bar for acceptance, by basically cutting the order in two parts: one below the acceptance threshold (that are rejected), and one above (that are accepted). The individual decisions "candidate X is no less suited than candidate Y" are relegated to a GPAI model "SmartLLM", which AutoRa is built ontop. Among its multitude of competences, SmartLLM has the ability to directly process the complete application material of candidate X and of Y (each presented as a single pdf consisting of CV, motivation letter, reference letters, English language certificate, transcript of records, and

[70] The AI Act also names some examples for appropriate techniques in its Recital 133 [10], such as watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints.

[71] Defined in Article 3(66) [10]; Recital 100 [10] makes this clear when stating that "when a general-purpose AI model is integrated into or forms part of an AI system, this system should be considered to be general-purpose AI system when, due to this integration, this system has the capability to serve a variety of purposes".

other certificates). It thereby judges who, if any of the candidates, is better suited for the study program. Prompt engineering has been used to make sure SmartLLM will answer "Yes" or "No" according to its judgement to the question "Is candidate X is less suited than candidate Y?" Furthermore, the GPAI model has been fine tuned by retraining it based on the decisions taken over the past five years by the board.[72] using state-of-the-art data curation techniques. The AI system AutoRa internally uses standard rules to minimize the number of calls to SmartLLM when assembling the partial order.

Legal classification under the AI Act. AutoRa, being an AI system that decides over the access to higher education, clearly falls under the high-risk systems listed in Annex III 3 (a) [10]. And thus AutoRa needs to fulfil the requirements for high-risk systems. This includes compliance with the quality requirements for training data under Article 10 [10], for instance with respect to possible biases that are likely to lead to discrimination.[73] a risk that is notoriously high for the university admission situation in our focus. In this regard, Article 10 [10] requires that the training, validation and testing data sets shall be subject to data governance and management practices appropriate for the intended purpose of the system.[74] and for this intended purpose it does not make a difference if the underlying technology uses a GPAI model or not.

Compliance through the training of the underlying GPAI model. First, one could wonder if the original training of SmartLLM can fulfil these requirements and whether compliance could be reasonably ensured by the provider of AutoRa. We doubt that the training of GPAI models is sufficient for high-risk use cases, as they are usually trained on very large datasets and therefore lack the necessary curation. In any case, however, the providers of downstream AI systems probably won't have the necessary means to verify whether or not the training of the GPAI model suits the intended purposes of their AI system. As outlined above, the provider of GPAI models, in our case SmartLLM, has certain information obligations for the benefit of downstream providers who build an AI system with a specific purpose based on the GPAI model.[75] The provision of this information also is intended to make it easier for providers of downstream systems to comply with the obligations of the AI Act.[76] which in our setting is the provider of AutoRa. However, the information obligations shall apply without prejudice to the provider's intellectual property and trade secrets.[77] Accordingly, the majority of providers of GPAI models will reduce the provided information to a minimum, thereby hampering verification of compliance through downstream providers. We therefore doubt that the information requirements for GPAI model providers are sufficient to enable downstream providers to verify compliance. This is particularly true with regard to the obligations concerning training data, as according to our reading the AI Act

[72] We assume that the applications of earlier candidates where graded on a 5-point scale from "strong reject" to "strong accept" by the selection committee, and that the fine tuning is done according to these ratings.
[73] Article 10(2) f) [10].
[74] Article 10(2) [10].
[75] Article 53(1) lit. b) [10].
[76] Article 53(1) b) (i) [10].
[77] Article 53(1) b) [10].

stipulates that the provider of AutoRa as a high-risk AI system would have to duly examine the training data in view of possible biases that are likely to lead to prohibited discriminatory output and to take appropriate countermeasures.[78]

Compliance through fine-tuning. Another option to ensure compliance would be to adapt an existing GPAI model sufficiently to the purpose of the AI system built on top of it through fine-tuning, like it was done in the case of AutoRa. The AI Act mentions the modification of GPAI models through fine-tuning only in passing: According to Recital 109 [10], fine-tuning results in the responsible party entering into the obligations of GPAI model providers with respect to the changes. However, the question of how fine-tuning affects compliance with the requirements for high-risk systems remains unanswered. This being said, pre-training a general model and then fine-tuning it on specific downstream tasks has proven to be highly effective, particularly in natural language processing [12]. The advancement of LLMs, such as ChatGPT, exemplifies this success [35,33]. Fine-tuning algorithms can be categorized into additive, selective, reparameterized, and hybrid fine-tuning based on their operations [12]. *Additive fine-tuning* maintains the pre-trained model unchanged and introduces minimal trainable parameters, such as an adapter layer [30,23,14] or a soft prompt [29,17,19]. Instead of adding parameters, *Selective fine-tunings* select a subset of the existing parameters to adapt, enhancing model performance over downstream tasks [28,8,11]. Reparameterization involves equivalently transforming a model's architecture from one form to another by transforming its parameters. Research [1] shows that common pre-trained models exhibit exceptionally low intrinsic dimensionality, making low-dimensional reparameterization effective for fine-tuning [1,32,20,26]. *Hybrid fine-tunings* combine the advantages of diverse fine-tuning approaches or analyze the similarities among them to establish a unified perspective [15,4,21]. However, one could doubt that fine-tuning makes the AI system compliant with the requirements for high-risk systems. For instance, catastrophic forgetting in fine-tuning, which is directly caused by the inaccessibility of historical data, is a research problem that has been extensively studied by numerous researchers [13,31,5,7]. Determining what information from the pre-training data is retained or forgotten after fine-tuning is challenging. For reasons of privacy and fairness, techniques for deliberately forgetting specific information during fine-tuning have been proposed [6,2]. Thus, even if a fine-tuned GPAI model meets the requirements of the AI Act, we cannot directly demonstrate that a high-risk SPAI system using it will satisfy the data quality requirements for such systems under the AI Act. Additional techniques may address these issues, but they must be tested specifically. Finally, fine-tuning a GPAI model for a specific purpose could also make it more difficult to meet the other system requirements, in particular for accuracy, robustness and cybersecurity under Article 15 [10]. Though fine-tuning is effective for downstream tasks in terms of accuracy, its security and robustness are not guaranteed. Research [24] indicates that simply fine-tuning with benign and commonly used datasets can inadvertently degrade the safety alignment of LLMs. Their experimental results demonstrate that fine-tuning LLMs makes it easier to bypass their safety guardrails, causing the model to respond

[78] Article 10(2)(f) and (g) [10]. See, however, regarding the obligations of the provider of a GPAI model in case of fine-tuning Recital 109 [10].

to nearly any harmful instructions [24]. Additionally, the security and robustness of general-purpose AI models, such as LLMs, remain unresolved challenges requiring further research [25,18,34,3]. These issues make it difficult to guarantee the safety of using GPAI models for high-risk applications.

Upshot of example. The example of AutoRa illustrates the challenges associated with downstream uses of GPAI models in high-risk AI systems. To our understanding, the AI Act obliges the downstream system providers to ensure that the high-risk requirements are met, including with respect to the underlying GPAI model. This may pose a major hurdle for the downstream provider when using a third party's GPAI model. The legal approach chosen by the AI Act thus arguably makes it difficult in practice to integrate GPAI models into SPAI systems , thereby jeopardising economic efficiency and threatening to stifle innovation. If, on the other hand, European legislation had limited the legal obligations of downstream providers to the parts of the AI system they have developed themselves, this would not have met the objective of ensuring a high standard of protection for societal and individual interests. This shows the trade-offs faced by the European legislature when attempting to regulate GPAI models and to create an effective regulation along the AI value chain.

> *Integrating a GPAI model into a high-risk SPAI system necessitates meeting strict system requirements, particularly concerning the quality of training data. Our case study demonstrated the significant challenges providers face in fulfilling these requirements.*

7.2 Obligations for GPAI systems

GPAI systems (as opposed to GPAI models), on the other hand, are only briefly mentioned in the AI Act. For example, generative AI based on GPAI systems is subject to a special transparency obligation under Article 50(2) [10]. Beyond that there are no specific requirements for GPAI systems. However, it is unclear which other obligations may also apply.

Applying high-risk requirements to GPAI systems. In our opinion, the applicable rules depend on the purposes of the GPAI system as designated by the provider.[79] When a GPAI system is designated to serve a variety of purposes, at least one of which are considered to be, e.g., high-risk, arguably the additional obligations for high-risk AI systems apply. However, when a GPAI system is determined to serve a variety of purposes none of which is categorized as high risk, even though a high-risk use might be factually possible, then additional obligations for high-risk AI systems do not apply. Article 25(1)(c) [10] seems to support this reading as it implies that GPAI systems can also be classified as high-risk AI systems.

[79] Article 3(12) [10].

Repurposing GPAI systems. However, this approach can be criticised since it gives the providers the power to determine the applicable rules by defining the purposes of the GPAI system. Because the deployer could then use the GPAI system for different purposes anyway, this could arguably lead to circumventing the requirements for high-risk systems. The AI Act tries to address this issue in Article 25(1)(c) [10] which requires that a deployer who modifies the intended purpose of a GPAI system which has not been classified as high-risk in such a way that the system becomes a high-risk AI system. In this case, the deployer turns into the responsible provider and has to make sure that the relevant system requirements are met.[80] We refer to this new provider as a "deployer-provider". However, even though the AI Act obliges the initial provider of the initially non-high-risk GPAI system to cooperate with the deployer-provider,[81] it is difficult to imagine how the deployer-provider can fulfill all these obligations, e.g. with regard to the quality of training data.[82] These issues are reflected in our example in Section 7.1, which refers to the related case of using GPAI models inside a high-risk AI system. Additionally, the obligation to cooperate with the downstream provider does not apply if the original provider of the GPAI system has clearly specified that the system is not to be changed into a high-risk AI system.[83]

Implications for the Provider. For the initial provider, this means that they need to consider carefully what purposes they want to set for their GPAI system. In our view, they have three options: Either they commit to complying with the relevant system requirements, thereby enabling end users to use their systems for high-risk purposes. Or, second, they can define the purposes of the GPAI system without matching them with high-risk use cases, and then try to provide due support to downstream users. Third, they could define the purposes without high-risk use cases and specify in the terms of use that the system may not be used for high-risk purposes, thereby avoiding the necessary support. It should be noted, however, that the provider could reduce the attractiveness of its GPAI systems if it excludes high-risk uses in order to circumvent the legal requirements, as this will affect the overall usefulness of its product.

When it comes to classifying a GPAI system as high-risk (or not), the AI Act does not focus on the system's capacities, but on its purposes as designated explicitly by the provider . A deployer using that system for a different purpose than designated becomes a provider as well and has to meet the thus induced legal obligations. Furthermore, obligations for GPAI systems include specific transparency requirements for generative AI.

8 Conclusion

This paper has navigated the AI Act from the perspective of the working programmer and has demonstrated that the AI Act contains numerous obligations of relevance for

[80] Article 25(1)(c) [10] and Recital 84 [10].

[81] Article 25(2) [10] and Recital 85 [10]. The initial provider "shall make available the necessary information and provide the reasonably expected technical access and other assistance".

[82] Article 10 [10].

[83] Article 25(2) [10], Recital 86 [10].

their daily work. Beyond that, other provisions of the AI Act may have an indirect effect on programmers and their working environment. For example, so-called regulatory sandboxes are to be established by the Member States[84] that are meant to enable the testing of AI systems in a controlled environment. Concerning programmers' personal skill set, the AI Act stipulates that providers and deployers have to ensure a sufficient level of "AI literacy" of their staff and other persons dealing with the operation and use of AI systems on their behalf.[85] AI literacy is defined as the skills, knowledge and understanding that allow providers, deployers and affected persons to make an informed deployment of the AI system, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause[86] and thus goes well beyond a purely technical expertise. So arguably, there will be an obligation for providers and deployers to organise training for staff members within the months to come.

The AI Act was published in the Official Journal of the EU on 12 July 2024 and thus entered into force on 1 August 2024. However, most parts of the Regulation will only become binding two years after. This is meant to give enough time to adapt to the new legal situation. However, some provisions will become applicable earlier, such as the obligation to ensure that staff have sufficient AI literacy (Article 4 [10]) as well as bans on certain AI practices such as manipulative AI deploying subliminal techniques, social scoring and real-time biometric information systems (Article 5 [10]). Within 12 months, the provisions requiring Member States to set up authorities as well as procedures on notification of general-purpose AI models presenting systemic risks become applicable (Article 51 [10]).[87]

Our paper has also shown that there is still a significant degree of legal uncertainties. It is therefore to be hoped that the transition phase of two years will be used by standardisation organisations to develop technical standards as well as by the AI Office to come up with Codes of Conduct[88] to mitigate insecurities about the legal framework as far as possible. At the same time, this also opens up the opportunity for some working programmers to contribute to the shaping of the future legal framework by submitting comments and engaging in discussions within professional associations, conferences, etc.

The AI Act has high relevance for the working programmer. This paper constitutes an attempt to pinpont the main facets. However, there is still considerable legal uncertainty.

Acknowledgements This work has received financial support by the DFG under grant No. 389792660 as part of TRR 248 – CPEC, by VolkswagenStiftung as part of Grant AZ 98514 – EIS, and by the Federal Ministry of Education and Research of Germany and Sächsisches Staatsministerium für Wissenschaft Kultur und Tourismus in the program Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig" (ScaDS.AI).

[84] Article 57 [10].
[85] Article 4 [10].
[86] Article 3(56) [10].
[87] See Article 113 [10].
[88] Article 95(2) lit. c) [10].

References

1. Aghajanyan, A., Zettlemoyer, L., Gupta, S.: Intrinsic dimensionality explains the effectiveness of language model fine-tuning. arXiv preprint arXiv:2012.13255 (2020)
2. Baik, S., Hong, S., Lee, K.M.: Learning to forget for meta-learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2379–2387 (2020)
3. Chen, C., Shu, K.: Combating misinformation in the age of llms: Opportunities and challenges. arXiv preprint arXiv:2311.05656 (2023)
4. Chen, J., Zhang, A., Shi, X., Li, M., Smola, A., Yang, D.: Parameter-efficient fine-tuning design spaces. arXiv preprint arXiv:2301.01821 (2023)
5. Chen, S., Hou, Y., Cui, Y., Che, W., Liu, T., Yu, X.: Recall and learn: Fine-tuning deep pretrained language models with less forgetting. arXiv preprint arXiv:2004.12651 (2020)
6. Chundawat, V.S., Tarun, A.K., Mandal, M., Kankanhalli, M.: Can bad teaching induce forgetting? unlearning in deep networks using an incompetent teacher. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 7210–7217 (2023)
7. Cong, Y., Zhao, M., Li, J., Wang, S., Carin, L.: Gan memory with no forgetting. Advances in Neural Information Processing Systems **33**, 16481–16494 (2020)
8. Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C.M., Chen, W., et al.: Parameter-efficient fine-tuning of large-scale pre-trained language models. Nature Machine Intelligence **5**(3), 220–235 (2023)
9. European Parliament and Council of the EU: Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC (2023), https://eur-lex.europa.eu/eli/reg/2023/1230/oj
10. European Parliament and Council of the EU: 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) (2024), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689
11. Guo, D., Rush, A.M., Kim, Y.: Parameter-efficient transfer learning with diff pruning. arXiv preprint arXiv:2012.07463 (2020)
12. Han, Z., Gao, C., Liu, J., Zhang, S.Q., et al.: Parameter-efficient fine-tuning for large models: A comprehensive survey. arXiv preprint arXiv:2403.14608 (2024)
13. Hayes, T.L., Kafle, K., Shrestha, R., Acharya, M., Kanan, C.: Remind your neural network to prevent catastrophic forgetting. In: European Conference on Computer Vision. pp. 466–483. Springer (2020)
14. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
15. Hu, Z., Wang, L., Lan, Y., Xu, W., Lim, E.P., Bing, L., Xu, X., Poria, S., Lee, R.K.W.: Llm-adapters: An adapter family for parameter-efficient fine-tuning of large language models. arXiv preprint arXiv:2304.01933 (2023)
16. Krönke, C.: Das europäische KI-Gesetz : Eine Verordnung mit Licht und Schatten. Neue Zeitschrift für Verwaltungsrecht **43**(8), 529–534 (2024)
17. Lester, B., Al-Rfou, R., Constant, N.: The power of scale for parameter-efficient prompt tuning. arXiv preprint arXiv:2104.08691 (2021)
18. Li, N., Pan, A., Gopal, A., Yue, S., Berrios, D., Gatti, A., Li, J.D., Dombrowski, A.K., Goel, S., Phan, L., et al.: The wmdp benchmark: Measuring and reducing malicious use with unlearning. arXiv preprint arXiv:2403.03218 (2024)

19. Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. arXiv preprint arXiv:2101.00190 (2021)
20. Liu, Q., Wu, X., Zhao, X., Zhu, Y., Xu, D., Tian, F., Zheng, Y.: Moelora: An moe-based parameter efficient fine-tuning method for multi-task medical applications. arXiv preprint arXiv:2310.18339 (2023)
21. Mao, Y., Mathias, L., Hou, R., Almahairi, A., Ma, H., Han, J., Yih, W.t., Khabsa, M.: Unipelt: A unified framework for parameter-efficient language model tuning. arXiv preprint arXiv:2110.07577 (2021)
22. Paulson, L.C.: ML for the Working Programmer (2. ed.). Cambridge University Press (1996). https://doi.org/10.1017/CBO9780511811326
23. Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., Gurevych, I.: Adapterfusion: Non-destructive task composition for transfer learning. arXiv preprint arXiv:2005.00247 (2020)
24. Qi, X., Zeng, Y., Xie, T., Chen, P.Y., Jia, R., Mittal, P., Henderson, P.: Fine-tuning aligned language models compromises safety, even when users do not intend to! arXiv preprint arXiv:2310.03693 (2023)
25. Sun, L., Huang, Y., Wang, H., Wu, S., Zhang, Q., Gao, C., Huang, Y., Lyu, W., Zhang, Y., Li, X., et al.: Trustllm: Trustworthiness in large language models. arXiv preprint arXiv:2401.05561 (2024)
26. Valipour, M., Rezagholizadeh, M., Kobyzev, I., Ghodsi, A.: Dylora: Parameter efficient tuning of pre-trained models using dynamic search-free low-rank adaptation. arXiv preprint arXiv:2210.07558 (2022)
27. Veale, M., Borgesius, F.Z.: Demystifying the draft eu artificial intelligence act — analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International **22**(4), 97–112 (2021). https://doi.org/doi:10.9785/cri-2021-220402
28. Vucetic, D., Tayaranian, M., Ziaeefard, M., Clark, J.J., Meyer, B.H., Gross, W.J.: Efficient fine-tuning of bert models on the edge. In: 2022 IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1838–1842. IEEE (2022)
29. Wang, Q., Mao, Y., Wang, J., Yu, H., Nie, S., Wang, S., Feng, F., Huang, L., Quan, X., Xu, Z., et al.: Aprompt: Attention prompt tuning for efficient adaptation of pre-trained language models. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. pp. 9147–9160 (2023)
30. Wang, Y., Mukherjee, S., Liu, X., Gao, J., Awadallah, A.H., Gao, J.: Adamix: Mixture-of-adapter for parameter-efficient tuning of large language models. arXiv preprint arXiv:2205.12410 **1**(2), 4 (2022)
31. Wang, Y.X., Ramanan, D., Hebert, M.: Growing a brain: Fine-tuning by increasing model capacity. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2471–2480 (2017)
32. Yang, A.X., Robeyns, M., Wang, X., Aitchison, L.: Bayesian low-rank adaptation for large language models. arXiv preprint arXiv:2308.13111 (2023)
33. Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., Hu, X.: Harnessing the power of llms in practice: A survey on chatgpt and beyond. ACM Transactions on Knowledge Discovery from Data (2023)
34. Zeng, Y., Lin, H., Zhang, J., Yang, D., Jia, R., Shi, W.: How johnny can persuade llms to jailbreak them: Rethinking persuasion to challenge ai safety by humanizing llms. arXiv preprint arXiv:2401.06373 (2024)
35. Zhong, Q., Ding, L., Liu, J., Du, B., Tao, D.: Can chatgpt understand too? a comparative study on chatgpt and fine-tuned bert. arXiv preprint arXiv:2302.10198 (2023)