Cambridge Forum on AI: Law and Governance (2025), **1**, e1, 1–13
doi:10.1017/cfl.2024.6

RESEARCH ARTICLE

LLMs meet the AI Act: Who's the sorcerer's apprentice?

Ugo Pagallo

Department of Legal Sciences, University of Turin, Turin, Italy
Email: ugo.pagallo@unito.it

(Received 01 August 2024; accepted 24 October 2024)

Abstract

The paper examines the legal regulation and governance of "generative artificial intelligence" (AI), "foundation AI," "large language models" (LLMs), and the "general-purpose" AI models of the AI Act. Attention is drawn to two potential sorcerer's apprentices, namely, in the spirit of J. W. Goethe's poem, people who were unable to control a situation they created. Focus is on developers and producers of technologies, such as LLMs that bring about risks of discrimination and information hazards, malicious uses and environmental harms; furthermore, the analysis dwells on the normative attempt of European Union legislators to govern misuses and overuses of LLMs with the AI Act. Scholars, private companies, and organisations have stressed limits of such normative attempt. In addition to issues of competitiveness and legal certainty, bureaucratic burdens and standard development, the threat is the over-frequent revision of the law to tackle advancements of technology. The paper illustrates this threat since the inception of the AI Act and recommends some ways in which the law has not to be continuously amended to address the challenges of technological innovation.

Keywords: AI Act; artificial intelligence; general-purpose AI model; legal governance; neutrality principle; risk

1. Introduction

Forms of artificial intelligence (AI), such as generative AI (Gen AI), foundation models, and large language models (LLMs), broke the news by the end of 2022 and throughout 2023. Benefits and threats of technology ignited people's imagination and drew the regulatory attention of law makers and institutions, including the Group of Seven (G7) countries under the Japanese Presidency in May 2023. Yet, defining AI and its subfields is no easy task as shown by the efforts of the European Union (EU) legislators since the proposal of the AI Act in April 2021, and with the series of amendments by the European Parliament in 2023, up to the final text of the Act in July 2024, i.e., Reg. (EU) 2024/1689.

Roughly speaking, LLMs can be understood as a subcategory of AI which is foundational and generative. "Foundation AI" refers to models upon which other AI systems and applications are built (Bommasani et al., 2021). Generative AI represents a type of AI that draws on the model design of deep neural networks and the developments of machine learning, finding out patterns from existing data to create new and unique outputs, such as texts, videos, images, audios, or 3D models. In the phrasing of Section 3 of the U.S. Executive Order from October 2023, Gen AI "means the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content" (EO, 2023). LLMs are then a particular subcategory of AI models that hinge on machine learning techniques to grasp and produce outputs on the basis of billions parameters through probabilistic inferences, rather than causal understanding. Among the most popular instances of

LLMs, the breakthroughs of the Generative Pre-trained Transformer, or GPT, are instructive. Trained on a huge corpus of data – as large as the whole web – the model starts from a source input, i.e., the prompt, to create the most likely output, such as sequences of code, words, or other data that are statistically appropriate, given that starting prompt. For better or for worse, the results have been outstanding. LLMs can be adapted to support manifold applications, from creative writing (Hsieh, 2019) to spell-checking (Hu et al., 2021), allowing users to innovate upon such models without any gatekeeping: the emergence of several open-source AI models, such as Meta's Llama 2 and Stable Diffusion, reinforces these trends (Dickson, 2023). However, LLMs also raise concerns, for example, the new generation of fake news and deep fake videos. According to certain scholars (Zellers et al., 2019), AI-generated fake news are more credible to human raters than human-written disinformation. Moreover, texts, images, audios, and videos generated by LLMs can be impossible to distinguish from human creation (Nightingale & Farid, 2022).

In this paper, the focus of the analysis is restricted to this latter side of the coin, i.e., the risks brought forth by LLMs. The intent is not to overlook the benefits of the technology. Instead, the aim is to appreciate the range of potential misuses and overuses of LLMs, and to grasp on this basis how law makers intend to tackle them, therefore striking a balance between pros and cons of technology. Correspondingly, the chapter is divided into four parts. Next, focus is on risks of LLMs related not only to misuses of technology but also to users who overestimate the capabilities of the technology and employ it in dangerous ways. Section 3 inspects the different techniques with which the law may aim to govern the development and uses of this technology. Special attention is drawn to the principle of technological neutrality. Section 4 dwells on the AI Act of EU law because it provides the first overall regulation of AI systems, LLMs, and general-purpose AI (GPAI) models. Section 5 dissects drawbacks and shortcomings of this regulatory effort that should not hinder technological innovation nor require over-frequent revision to tackle advancements of technology. The conclusion of the analysis brings us back to the question posed by the title of this paper. Whereas, in the spirit of J. W. Goethe's poem, a sorcerer's apprentice refers to people who were unable to control a situation they created, we may wonder on whether such sorcerer's apprentices are only developers and producers of LLMs, or they include some law makers that attempt to govern misuses and overuses of technology.

2. Looking for the sorcerer's apprentice

In addition to the benefits of LLMs, scholars have stressed their threats. In Weidinger et al. (2021), for example, the ethical challenges associated with LLMs are divided into six risk fields. They regard, (i) discrimination, for LLMs can perpetuate social stereotypes and biases, triggering representational and allocational harms; (ii) information hazards that may impinge on people's privacy by leaking personal information and inferring sensitive data; (iii) misinformation hazards that can affect trust, or lead to less well-informed users; (iv) malicious use by people with criminal intent, for example, large-scale frauds or personalised scams; (v) human–computer interaction harms that depend on users overvaluing the capabilities of LLMs and using them in unsafe ways; and (vi) environmental harms that hinge on the energy power which is necessary to train and operate the technology (Pagallo et al., 2022).

Further normative issues posed by LLMs regard the legal domain. Arguably, one of the most sensitive fields has to do with the protection of intellectual property rights, in particular, copyright. Authors of copyrighted materials, e.g., books, often claim their texts have been ingested and used to train such LLMs, as ChatGPT, without their consent. Lawsuits have been consequently lodged (also but not only) in the U.S. under the provisions of the Copyright Act (17 U.S.C. § 501) and the Digital Millennium Copyright Act (17 U.S.C. § 1202). Other controversial cases concern personal data similarly ingested in the training datasets of the LLMs. In EU law, the opinion of courts and regulators suggests that web scraping involving the collection of personal data is unlawful if it does not comply with principles and provisions of the general data protection regulation, the GDPR (Pagallo and Ciani 2023). In March

2023, the Italian privacy authority ordered a ban of ChatGPT over alleged data protection violations. The service was reactivated a month later, once OpenAI addressed some of the issues raised by the Italian authority, such as lack of adequate information to its users and of suitable legal basis for personal data processing (Bertomeu et al., 2023).

Against this backdrop of ethical troubles and open legal issues, scholars have stressed time and again that the self-regulatory policies of private companies in all their variants – from "pure" unenforced forms of self-regulation to "approved" self-regulation (Marsden, 2011) – fall short in coping with the challenges of AI technologies (Pagallo et al., 2019). It is remarkable that most law makers across the board have been very active over the past two years. China adopted interim measures for Gen AI in 2023; the White House released its Executive Order also but not only on Gen AI in October of that year; and the EU finally adopted Reg. (EU) 2024/1689 in July 2024. The race to AI regulation can be understood as a response of legislators to certain opinions in the business sector, and among technological savvies and aficionados that gained track over the same short span of months and years in which the race to the regulation of AI started. A few weeks after ChatGPT reached an amazing 100 million monthly active users within just two months of its launch, a group of business leaders and experts, including the boss of ChatGPT, Sam Altman, warned in May 2023 that humanity faces the risk of "extinction" following the rise of AI.[1] Combating AI-related risks should be "a global priority on par with other society-wide risks, such as pandemics and nuclear war," as the signatories of the manifesto claimed on the website of the Center for AI Safety, a U.S.-based non-profit organisation.[2] To be fair, such concerns are not new, since another organisation, i.e., the Future of Life Institute released a similar open letter in 2015 in which the likes of Bill Gates, Elon Musk, Stephen Hawking, and others warned against threats of AI and robotics (Pagallo, 2017). Some of these risks are real and yet, it is the source of the warning that is at times unconvincing. It reminds us of the sorcerer's apprentice who blames its own fate. Consider Sam Altman: on one hand, the CEO of OpenAI warns against risks of human extinction; on the other, his company goes on making business as usual.

The assumption of this paper is that other sorcerer's apprentices are out there in the field of AI. Which means that sorcerer's apprentices may not only be some darlings of the market or of technological innovation but also public guardians that should protect citizens against such apprentices. The next parts of the paper intend to explore this disturbing possibility.

3. Legal remedies

There are multiple ways in which the law may tackle the normative challenges of LLMs, which includes the production and dissemination of disreputable language, e.g., sexist or racist comments (Kirk et al., 2021), or answers that contain factual errors that can contribute to the erosion of social trust (Evans et al., 2021). Some scholars propose to tackle such issues through technical solutions, for example, (i) the preprocessing of training data; (ii) the fine-tuning of LLMs on data with desired properties, such as predefined ethical principles; or (iii) procedures to test LLMs before their deployment (Avin et al., 2021; Perez et al., 2022). Others suggest guaranteeing LLMs transparency through the watermarking of the model's outputs (Kirchenbauer et al., 2023); datasheets (Gebru et al., 2021); system cards (MetaAI, 2023); or model cards (Derczynski et al., 2023). Further strategies concern the social aspects of technological development. The challenges of LLMs should be addressed in this case with the set-up of more diverse developer teams and human-in-the-loop protocols (Wang et al., 2021); or structured access protocols (Shevlane, 2022).

Some of these proposals have been endorsed by the 2023 U.S. Executive Order mentioned above in the introduction (EO, 2023). Section 4.1 of the Order concerns the development of guidelines, standards, and best practices for AI safety and security; in particular, the development of "a companion resource to the AI Risk Management Framework, NIST AI 100-1, for generative AI" (4.1(i)(A)); and

[1] See https://edition.cnn.com/2023/10/31/tech/sam-altman-ai-risk-taker/index.html.
[2] See https://www.safe.ai/statement-on-ai-risk.

"a companion resource to the Secure Software Development Framework to incorporate secure development practices for generative AI and for dual-use foundation models" (lett. B). In accordance with a sectorial approach, further specific safeguards regard:

(i) The assessment of biosecurity risks, e.g., "generative AI models trained on biological data" (4.4(ii)(A));
(ii) The prevention of child sexual abuse or "non-consensual intimate imagery of real individuals" (4.5(a)(iv));
(iii) Issues related to generative AI and inventorship of patentable subject matter under the guidance of the U.S. Patent and Trademark Office (5.2(c)(i));
(iv) The "development, maintenance, and use of predictive and generative AI-enabled technologies in healthcare delivery and financing – including quality measurement, performance improvement, program integrity, benefits administration, and patient experience – taking into account considerations such as appropriate human oversight of the application of AI-generated output" (8(b)(i)(A));
(v) Recommendation to federal agencies, in consultation with the Secretary of Commerce, the Secretary of Homeland Security, and the heads of other appropriate agencies, as regards the "external testing for AI, including AI red-teaming for generative AI, to be developed in coordination with the Cybersecurity and Infrastructure Security Agency" (10.1(b)(viii)(A); the "testing and safeguards against discriminatory, misleading, inflammatory, unsafe, or deceptive outputs, as well as against producing child sexual abuse material and against producing non-consensual intimate imagery of real individuals … for generative AI" (lett. B); and "reasonable steps to watermark or otherwise label output from generative AI" (lett. C);
(vi) Further measures to "advance the responsible and secure use of generative AI in the Federal Government" (10.1(d) and (f)). In particular,

> agencies are discouraged from imposing broad general bans or blocks on agency use of generative AI. Agencies should instead limit access, as necessary, to specific generative AI services based on specific risk assessments; establish guidelines and limitations on the appropriate use of generative AI; and, with appropriate safeguards in place, provide their personnel and programs with access to secure and reliable generative AI capabilities, at least for the purposes of experimentation and routine tasks that carry a low risk of impacting Americans' rights. (lett. (f)(i)).

Contrary to the U.S. sectorial approach – which applies not only to AI but also to many other fields of legal regulation, such as personal data protection, or data privacy – the approach of the EU intends to be "horizontal" rather than "vertical." Although within the limits of EU law, which mostly regard issues related to national security, public order and the military, the aim of the EU law makers is to govern the entire life cycle of AI systems, starting from the collection of input data to the final use of such technologies. Theoretically speaking, there are three options for this horizontal approach: "the Switch, the Ladder, and the Matrix" (Mökander et al., 2023). The Switch refers to a binary method, according to which technology is regulated as such, so that law makers shall determine which systems are or are not considered AI with their variables: Gen AI, LLMs, etc. The Ladder refers to a risk-based approach; therefore, AI systems are not considered as such, but rather, clustered into different categories in accordance with the level of risk they pose: no risk, low, medium, high, etc. Finally, the Matrix refers to a multidimensional approach, in which manifold aspects of the technology, such as data input, decision-model type, or the social context and purpose of the AI system must be considered to specify the level of risk. This is the approach recommended by some scholars (Wilson et al., 2020) and institutions (OECD 2022).

Against this tripartition, what is then the horizontal approach adopted in the EU with the 2021 proposal of the European Commission?

4. The principle of technological neutrality

To determine whether the horizontal approach of the AI Act has endorsed the Switch approach, the Ladder approach, or the Matrix, attention should be drawn to a crucial challenge of technological regulation, namely, the future-proofing of the law. In a nutshell, the law should not be often revised to keep the pace of technological innovation, nor hinder such innovation with its own provisions. Against this twofold constraint, a further challenge of technological regulation has to do with the fact that the law is not the only regulatory system out there. Further regulatory systems include ethics and the forces of the market, social customs and technology as a regulatory system of its own. Social regulatory systems may reinforce each other: for example, the provisions of the 1996 Treaties of the World Intellectual Property Organization on the use of "digital rights management" illustrate the convergence between powerful economic interests, technological solutions, and the law. However, the claims of such regulatory systems may also clash and even render the normative claim of another regulatory system irrelevant (Pagallo & Durante, 2016). The EU e-money directive 46 from 2000 reminds us of how the regulatory claims of the law may fail vis-à-vis digital innovation. Soon after the implementation of the directive, which aimed at expanding traditional forms of centralisation to online interaction, new forms of payment, such as PayPal, made the legal regulation obsolete: the EU legislators in Brussels had to amend themselves with a new directive, n. 110 from 2009. Likewise, in the field of civil aviation with the regulation of drones, Reg. (EU) 2008/216 established a highly decentralised network revolving around the powers of member states and national civil aviation authorities that soon led to the fragmentation of the system. The EU law makers had to intervene with Reg. (EU) 2018/1139 on common rules for drones in the field of civil aviation (Bassi, 2019).

There are several legal techniques that help the law prevent failings in technological regulation. The overall intent is to make the general and abstract commands of the law responsive to the challenges of technology. Such techniques include

(i) Technological indifference, or the principle of technological neutrality, according to which legal regulations apply in identical ways, whatever the technology. Both Recital 15 of the GDPR and the right to authorise communication of a work to the public in the field of copyright law illustrate this legal technique against risks of inefficacy and normative obsolescence;
(ii) Implementation neutrality, so that regulations are by definition specific to a technology and yet, they do not favour one or more of its possible implementations, for example, the signature of e-documents;
(iii) Potential neutrality of the law that sets up a particular attribute of a technology, e.g., consumer friendly interfaces in e-commerce law, although law makers can draft the legal requirement in such a way that even non-compliant implementations can be modified to become compliant (Reed, 2012).

Since the proposal of the European Commission in April 2021, the AI Act has revolved around the principle of technological neutrality and its corollaries. In the phrasing of Section 5.2.1 of the first draft of the Act, "the definition of AI system in the legal framework aims to be as technology neutral and future proof as possible, taking into account the fast technological and market developments related to AI." The aim of the AI Act is not to govern the technology as such, but rather, the uses of AI systems. To attain this end, according to the Explanatory Memorandum of the Act, the classification system of the regulation tailors requirements and obligations based on a "risk-based approach." Therefore, the AI Act endorses the Ladder, rather than the Switch approach. The overall idea is, on

the one hand, to classify uses of AI in accordance with different levels of risk so that, on the other hand, different kinds of governance apply to such risk levels. The first draft of the AI Act clustered such risks into four categories, i.e., (i) unacceptable risks with corresponding bans, or prohibitions; (ii) high risk level with certification duties and further administrative burdens for developers, manufacturers, or end-users of AI systems; (iii) low risk uses of AI; and between high and low risk AI systems, (iv) transparency duties for some uses of the technology (Ebers et al., 2021). The 2024 consolidated text that followed the institutional trialogue between the Council, the Parliament, and the Commission specified a new risk, i.e., "systemic risk" (AI Act, 2024). According to the final wording of Recital 97 of the Act, "this Regulation provides specific rules for general-purpose AI models and for general-purpose AI models that pose systemic risks, which should apply also when these models are integrated or form part of an AI system." The new level of risk has to do with either the "high-impact capabilities" of the model measured in floating point operations (FLOPs), or "significant impact on the internal market due to its reach" (Recital 111).

Drawing on this basis, the purpose – and hope – of the EU institutions with their "Ladder approach" is fourfold. First, the aim of the AI Act is to provide certainty through a unified legal framework. Contrary to sectorial or vertical approaches of regulation, as occurs in U.S. law, the AI Act intends to set up clear rules for all AI systems and fields of regulation: "certainty" is a mantra of the first draft of the AI Act. The same holds for the final version of the legislation, according to Recitals 3, 12, 83, 84, 97, 139, 177 of Reg. (EU) 2024/1689.

Second, the risk-based approach of the Regulation will reinforce the future-proofing of the law to tackle advancements of technology. Rather than focusing on the technicalities and different approaches of AI, from logic-based to statistical methods, focus is on the different uses of technology with probabilities of events, consequences, and costs. In the Explanatory Memorandum of the 2021 draft, it is argued that the AI Act "proposes a single future-proof definition of AI. Certain particularly harmful AI practices are prohibited as contravening Union values, while specific restrictions and safeguards are proposed in relation to certain uses of remote biometric identification systems for the purpose of law enforcement" (Section 1.1 of the Memorandum). Whatever the means, it is the impact of AI systems on citizens, consumers, institutions, and society at large that which ultimately matters. This overall approach represents the backbone of Reg. (EU) 2024/1689. The establishment of regulatory sandboxes, pursuant to Art. 57 of the Act, should reinforce this proactive approach.

Third, the different forms of governance set up with the AI Act – namely, the establishment and implementation of policies, procedures, and standards for the proper development, use, and management of AI systems, according to their level of risk – shall properly address the complexity of the field. One of the main reasons for the failure of the law in governing advancements of technology has in fact to do with lack of "flexibility." The formula means that the hard tools of the law fall often short in coping with the challenges of technological innovation, either because unfit to meet such challenges, e.g., development of standards, or because legal provisions simply arrive "too late," as occurs in the paradox of Zeno about the speedy Achilles (the technology), and the Turtle (the law). The AI Act intends to prevent these risks endorsing all three methods of legal governance, i.e., (i) the strict top-down approach of hard law that hinges on the threat of physical or pecuniary sanctions, for example, prohibition of unacceptable uses of AI, or certification duties for high-risk AI systems; (ii) a co-regulatory approach, as much as occurs with Art. 5(2) of the GDPR, in the case e.g., of development of certain standards under Art. 40 of the AI Act; and, (iii) a self-regulatory approach at work, for example, with Art. 95 on codes of conduct for other than high-risk AI systems.

Last but not least, the all-encompassing legal framework of the AI Act should provide the basis for the respect and protection of the fundamental rights in EU law. The principle of legal certainty and a risk-based approach – as well as different kinds of legal governance for the risks posed by AI technologies – intend to attain this end through a new set of duties and obligations for designers, manufactures, and in certain cases, end-users of the technology. This is not to say that AI systems do not create "liability gaps," nor that new rights should not be established to protect individuals. Further

proposals of EU law, from the AI Liability Act to the revised Product Liability Directive, complement the provisions of the AI Act by establishing new procedural rights and safeguards, by equalising the protection against certain malfunctioning of AI software to defects of products, etc. (Pagallo, 2022).

However, both the adoption of the principle of technological neutrality and the Ladder approach beg the question about the material scope of the legislation: the aim is to govern AI technologies, after all. In the first draft of the AI Act, Art. 3(1) on definitions referred to AI as a software developed with the "techniques and approaches" listed in Annex I. Three years later, the final version of Art. 3(1) intends AI as a "machine-based system" that can operate with "varying levels of autonomy" and eventually adapt itself after deployment. The new definition of AI in Art. 3(1) must be complemented with the further definitions of "GPAI model" (Art. 3(63)); "high-impact capabilities" (no. 64); "systemic risk" (no. 65); and "general-purpose AI system" (no. 66). They provide the final response of the EU legislators to the buzz and disruption of foundation models, Gen AI, and the subcategory of LLMs over the past two years.

The aim of the next section is to illustrate how the problems of the AI Act with the definition of technology are related to a more complex issue than that of multiplying the Switch upon which the Ladder approach rests. This more complex issue revolves around the normative design of the Act. The next section aims to illustrate the many facets of this problem from the "single future-proof definition of AI" in the Explanatory Memorandum of April 2021 to the five AI-related definitions inserted in the final text of June 2024.

5. The troubles of the AI Act with AI

The problems of EU legislators with the regulation of AI can be divided in this context – i.e., dealing with the challenges of LLMs as a subcategory of Gen AI and foundation models – into two groups. The first set of issues has to do with the "horizontal approach" of the AI Act; the second set specifically concerns the regulation of LLMs. The next subsections examine each of these problems separately.

5.1 The incongruities of the horizontal approach

Almost all articles of the first draft of the AI Act, i.e., the proposal of the Commission in April 2021, were the target of criticism and subject of vibrant discussions. The number of amendments passed by the Parliament in March and June 2023 down to the consolidated text of the Act from January 2024 illustrate this point. The intent of this subsection is not to cover this institutional debate, but rather, to sum up some crucial inconsistencies of the new legislation in accordance with a meta-regulatory approach. This perspective draws attention to four problems of the Act that concern the protection of fundamental rights, governance mechanisms, legal certainty, and the material scope of the legislation.

First, regarding the protection of the rights enshrined in the EU Charter of Fundamental Rights (CFRs), the 2021 proposal of the Commission simply overlooked some of them, for example, the right to a high level of environmental protection and the improvement of the quality of the environment under Art. 37 of the CFRs. Environmental risks fell outside of the 2021 proposal lest human rights were directly affected (Gailhofer et al., 2021). Correspondingly, in the series of amendments presented in March and June 2023, the European Parliament proposed to insert a fundamental rights impact assessment (FRIA) with a new Recital 58(a) and a new Article 29(a) of the Regulation (EP, 2023). An intense institutional trialogue between the Commission, the European Parliament, and the Council followed as a result. The hard-won compromise reached with the final version of the Act can hardly be presented as satisfactory. The FRIA of Art. 27 covers only some high-risk uses of AI, whereas Art. 95 on Codes of conduct hinges on the good will and voluntary initiatives of companies and corporations. Although the final goal of the Codes of conduct of Art. 95 is to progressively expand the set of standards and protection established for high-risk uses of AI to other than high-risk uses

of the technology, even a teleological interpretation of Art. 95 and 112(7) of the Act is not enough to make that expansion legally binding.

The troubles of the AI Act with the protection of rights entail further governance issues. As noted in the previous sections, the regulation concerns different kinds of governance in accordance with the risks posed by technology, namely, strict top-down regulations for high-risk uses of AI vis-à-vis self-regulatory solutions for other than high-risk uses of AI systems. However, it is apparent that also "low-risk" AI applications – in the original meaning of the 2021 proposal – may have a relevant impact on the environment of human societies. AI technologies often require massive computational resources and corresponding large computing centres that have a very high energy requirement and carbon footprint (Sokolowski, 2021). Even the Regulation seems to admit this fact with the provisions of Art. 73 on "serious incident." According to the definitions of the Act, the formula covers "incident or malfunctioning of an AI system that directly or indirectly leads to… the infringement of obligations under Union law intended to protect fundamental rights" (Art. 3(49)(c)); and, "serious harm to … the environment" (lett. d). It is likely that several environmental risks of AI systems and the protection of some fundamental rights may thus fall within the loopholes of the legal governance set up in Chapter VII of the Act. The assumption rests on the limited cases in which the protection of fundamental rights, such as environmental protection, is covered by the top-down provisions of the Regulation and, moreover, on how the governance of other than high-risk AI systems may fall short in coping with consequences of AI uses that, although serious, are not covered by the strict definitions of Art. 3(49)(c) and (d).

A further problem of the AI Act concerns one of its alleged strengths, i.e., legal certainty. The "horizontal approach" of the legislation necessitates a strict coordination with other pieces of EU law, either horizontal, e.g., the GDPR, or vertical, such as regulations in the financial sector, for medical devices, or drones. The first draft of the AI Act created some legal puzzles (Veale & Zuiderveen Borgesius, 2021). How to put together the panoply of technological regulations in EU law is indeed no easy task; just like squaring circles through stochastic methods in the legal domain. In fact, the premise is often given by hardly readable texts, for example, the provisions that amend previous acts of the Union in civil aviation (Art. 102 and 108 of the AI Act); new agriculture and forestry vehicles (Art. 103); marine equipment (Art. 105); or rail systems interoperability (Art. 106). To be sure, the extremely technical wording of such legal texts does not entail that every coordination among different pieces of legislation is impossible, for example, between the assessment of multiple levels of risk triggered by AI systems in accordance with the different parameters of the AI Act, of the GDPR, or of the regulation of AI as a medical device. Further problems of legal certainty and coordination may arise with Recital 11 of the AI Act, according to which the regulation should be understood without "prejudice" to Reg. 2022/2065 on a single market for digital services. Likewise, Recital 141 establishes that the safeguards of the AI Act have to be intended "in accordance with" Reg. 2022/868 on data governance and Reg. 2023/2854 on fair access to and use of data.

The principle of legal certainty of the AI Act is also under stress because of the material scope of the legislation. Since the introduction of this paper, we noted that the disruption of Gen AI and LLMs between 2022 and 2023 ignited not only the imagination of everyday people but also that of law makers and institutions. In the amendments adopted by the European Parliament on 14 June 2023 (EP, 2023), the counterproposals to the 2021 text did not only stress the incongruities of the first draft, e.g., the lack of a FRIA, but also the reasons why a new array of provisions had to cope with that which fell outside the proposal of the Commission, just as LLMs. The consolidated text from January 2024 and then, the official version of the AI Act intended to square the legal circle with the overall notion of "GPAI" model (and GPAI system). The previous section of this paper mentioned the definitions no. 63 (on GPAI model); no. 64 (on its high-impact capabilities); no. 65 (systemic risk); and no. 66 (GPAI system) of Art. 3 of the AI Act.

Next, focus is on the legal hole patch followed as a result of the hard-won compromise between the Triad, i.e., the Council, the Parliament, and the Commission. Since the good will of the EU institutions is out of the question, attention should be drawn to some side-effects of the Regulation.

5.2 A new sorcerer's apprentice?

The summer 2023 amendments of the European Parliament do not refer to the use of LLMs, and only partially to the notion of Gen AI. Rather, the focus is on foundation models. The formula was coined by (Bommasani et al., 2021), to identify AI models that, on one hand, are fruitful across multiple fields of application and present emergent capabilities when scaled up; on the other hand, such models can be adapted to undefined downstream tasks by a multitude of actors (Bommasani & Liang, 2021). In the wording of the Parliament's amendment – i.e., Art. 3(1)(1c) of the 2023 amendments – foundation model "means an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks."

The Parliament proposed a full array of new regulatory measures for foundation models: the term is employed 62 times in the series of amendments. Many of them made sense. Since the very beginning of this paper, attention was drawn to the normative challenges of LLMs that should be properly tackled. The result was the consolidated text from January 2024 which is copied and pasted by Art. 3(63) of the Regulation. The definition refers to

> an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, 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 and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market.

The AI Act sets up specific provisions for LLMs falling under the all-encompassing formula of GPAI model. Such a formula should be complemented with the notion of "high-impact capability," i.e., models that match or exceed the capabilities of the most advanced GPAI models (Art. 3(64)), which can further trigger a "systemic risk" due to its "significant impact on the internal market" and because of "negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain" (Art. 3(65)). Therefore, Art. 51 of the Regulation is devoted to the classification of these AI models; Art. 52 sets up conditions and procedure for the use of GPAI models with systemic risks; Art. 53 establishes the obligations for providers of GPAI models; Art. 54 specifies the obligations for providers of third countries; Art. 55 determines the obligations for providers of GPAI models with system risk; Art. 56 defines the Codes of practice. The overall outcome of this normative effort looks like a legal monster.

The stratification of the horizontal or "Ladder approach" in the AI Act is strengthened by the multiplication of "legal switches" on what is AI and what is not; what is AI and what is a GPAI model; what is a GPAI model and a GPAI system; or a GPAI model with or without systemic risks; or, with or without high-impact capabilities. The threat is to end up with all drawbacks of both the vertical and horizontal approaches to the regulation of technology. In fact, as regards the original aim of the Act on "rules of harmonisation" in EU law, Chapter V of the legislation on what US law dubs Gen AI adds to the coordination problems of the horizontal regulation, for example, the new obligation of GPAI providers, i.e., of GPAI models, including LLMs, to "put in place a policy to comply with Union law on copyright and related rights" (Art. 53(1)(c) of the AI Act). On the other hand, multiple definitions of manifold AI systems intend to flesh out the material scope of the legislation, and yet, they exponentially increase the troubles with the regulation of AI.

The multiplication of regulatory efforts on general-purpose models, systemic risks, high-impact capabilities, or the cumulative amount of compute used for training these models, e.g., the greater

than 10^25 FLOPs of the AI Act (Art. 51(2)), does not prevent the main threat of technological regulation, namely, the over-frequent revision of the legal text. The EU law makers have precautionarily established the legal presumption that all LLMs, such as GPT-4, have "high impact capabilities" pursuant to Art. 51(1) of the legislation – in fact, the pre-training for GPT-4 just required about 10^25 FLOPS in 2024. And still, it is sort of paradox that Recital 4 of the Act shall remind us of how "AI is a fast evolving family of technologies that contributes to a wide array of economic, environmental and societal benefits across the entire spectrum of industries and social activities." It is noteworthy that the principle of technological neutrality – and for that matter, every legal technique that may help the law prevent the obsolescence of regulations aiming to govern advancements of technology – vanished from the AI Act. Considering the exponential growth of innovation, from computing power to data availability for further AI models, systems, and approaches that shall be expected in the next future, the threat is that law makers will arrive once again too late in this cat-and-mouse, Achilles-and-the turtle game. Can the AI Act catch up with this complexity and prevent the threat of its own obsolescence?

6. Conclusions

In his 1847 conference *Court Resolutions Are Not Science*, the German philosopher of law Julius Hermann von Kirschmann claimed, "three rectifying words from the legislators and whole libraries become wastepaper" (von Kirschmann, 1881). This has been the risk with all discussions about the AI Act, from the first draft to the series of amendments by the European Parliament in 2023, from the consolidated text of January 2024 to the version of the Parliament from March of that year, down to the final text of July 2024. At the time of writing, it is an open issue how critical provisions of the regulation will be interpreted and implemented, for example, as regards the new AI standards of the regulation. Also, further provisions shall be expected with delegated or implementing acts of the Commission under Art. 58, 60, 68, 72, 92, or 101 of the Regulation. This uncertainty, however, *pace* von Kirschmann, does not mean that we should simply wait for the "three rectifying words" of the legislator to assess a legal framework. The paper has insisted on some specific inconsistencies of the AI Act in accordance with a meta-regulatory approach. In the spirit of Goethe's *Sorcerer's Apprentice*, the aim was to illustrate the threat that legislators may worsen the challenges of technology with their own provisions. The risk materialises every time that regulations either stifle sound technological innovation or need over-frequent revision to tackle the speed of such progress.

To corroborate the assumption, the analysis illustrated pros and cons, strengths and drawbacks of the current regulatory efforts in EU law. The benefits regard a single piece of legislation – horizontal, in the jargon of EU lawyers – that shall provide legal certainty, proper governance, and strong protection of fundamental rights. The pitfalls regard limits that are endogenous or exogenous. The limits start with the definition of the material scope of the legislation, namely, the preliminary definition of that which should be governed. In addition to the growing list of definitions in the AI Act regarding the use of AI systems and of GPAI models, further models and subcategories of AI must be expected in the foreseeable future. The AI Act provides the technique of Annexes to the legal text to keep the list of definitions growing with the implementing powers of the Commission. The problem does not depend, however, on whether the list is more or less complete, but rather, on whether the original aim of the legislation – the risk-based, or Ladder approach of the first draft of the Act – will be attained, or still makes sense. Among the most relevant endogenous limits, or defects of the final AI Act, it seems fair to concede that a straight risk assessment of the uses of technology remains problematic. Whether or not the AI Act is considered a risk-based regulation, crucial aspects of technological governance remain open, such as (i) uncertainties associated with the notion of risk and corresponding protection of rights; (ii) the difficulty to assess all risks raised by specific uses of AI; (iii) the costs related to such assessment; and (iv) the future-proofing of the law.

All in all, what legal technique should prevent the threat of legislators to over-frequently revise their own texts remains uncertain in the case of the AI Act. We noted that the Commission adopted the principle of technological neutrality in the first draft of the Act and then the EU legislators discarded the principle along the law making process. There is no reference to any technological neutrality in the amendments approved by the European Parliament in June 2023 (EP, 2023), in the consolidated text of January 2024 (AI Act, 2024), in the version of the Act released by the Parliament from March of that year, down to the official text of July 2024. Although implemented through a differentiated transition period for certain uses of AI according to Art. 113 of the Act, the risk of legal obsolescence is real considering that legislation will be at full speed on 2 August 2026. New technological advancements can be reasonably expected in this span of time, triggering once again a cat-and-mouse game in which lawmakers may arrive "too late" or worsen the challenges of technology with their regulations.

Such risks were illustrated with the EU e-money directive 46 from 2000 and its revision in 2009; together with the revision of the EU 2008 regulation on the use and governance of drones with Reg. (EU) 2018/1139, i.e., today's General Regulation on civil aviation. The paper discussed some of the reasons why (certain provisions or chapters of) the AI Act may meet a similar fate, i.e., the self-defeating outcome of every sorcerer's apprentice. To be fair, the Regulation establishes several legal mechanisms to tackle both the known-unknowns and even the unknown-unknowns of the legislation. We noted that a considerable amount of normative power has been devolved back to the European Commission through the means of delegated or implementing acts. Further provisions concern the detailed evaluation and review of the Regulation under Art. 112. Every year the Commission will have to assess whether the list of Annex III and that related to Art. 5 should be amended (Art. 112(1)). Every four years, by the end of this decade, the Triad should discuss, among other things, whether "extending existing area headings or adding new area headings in Annex III" (no. 2). Reports should include the assessment of national competent authorities, state of penalties, harmonised standards, and "the number of undertakings that enter the market after the entry into application of this Regulation, and how many of them are SMEs" (no. 4). 2 August 2028 will also be particularly relevant because that's the first deadline for the Commission's reports on the functioning of the AI Office, progress on the development of standards, and the impact and effectiveness of the Codes of conduct (no. 5–7).

Therefore, the question does not revolve around whether the EU lawmakers will be forced to amend themselves in accordance with the provisions of the AI Act, but rather, how soon we should expect this process of revision to begin with the complex interaction of Art. 58, 60, 68, 72, 92, and 101 of the Regulation – on the Commission's delegated powers – and the seven fronts of Art. 112. The over-frequent revision of the law is not only highly likely but seems even necessary, or inevitable. Leaving aside the powers of the Commission, we know even the starting time of this complex process of amendment and revision because Art. 5 and Annex III of the Regulation shall be assessed by August 2027. By then, it may well happen that which occurs from time to time in business and Greek legends with "sons" devouring and replacing "fathers," for example, Tim and Telecom Italia in the field of telecommunications; much as Zeus and Cronos in the Greek myth. As regards the AI Act, this may occur with LLMs of Chapter V superseding the AI systems of Art. 3(1) of the Regulation. However, in addition to the pace of technological innovation, the chapter insisted on problems of legal governance and legal certainty, and the troubles with the protection of fundamental rights. Law makers in Brussels are spoiled for choice as to where they will start amending the Act.

Funding statement. No funding was received for publication of this article.

Competing interests. No situations —- including financial, professional, contractual, or personal relationships or situations —- have exerted an undue influence on the content or publication of this work.

References

AI Act. (2024). Consolidated text of 2021/0106 (COD), 24 January, Brussels.

Avin, S., Belfield, H., Brundage, M., Krueger, G., Wang, J., Weller, A., Anderljung, M., Krawczuk, I., Krueger, D., Lebensold, J., & Maharaj, T. (2021). Filling gaps in trustworthy development of AI. *Science*, *374*(6573), 1327–1329.

Bassi, E. (2019). European drones regulation: Today's legal challenges. In *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*, Atlanta, GA, USA, 443–450.

Bertomeu, J., Lin, Y., Liu, Y., & Ni, Z. (2023). Capital market consequences of generative AI: Early evidence from the ban of ChatGPT in Italy. Available at SSRN.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, A., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … **Liang, P.** (2021). On the opportunities and risks of foundation models. *ArXiv*, https://doi.org/10.48550/arXiv.2108.07258.

Bommasani, R., & Liang, P. (2021). *Reflections on Foundation Models*. Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/reflections-foundation-models.

Derczynski, L., Hannah Rose, K., Balachandran, V., Kumar, S., Tsvetkov, Y., Leiser, M. R., & Mohammad, S. (2023). Assessing language model deployment with risk cards. *ArXiv*, https://doi.org/10.48550/arXiv.2303.18190.

Dickson, B. (2023) *How Open-source LLMs Are Challenging OpenAI, Google, and Microsoft*. https://bdtechtalks.com/2023/05/08/open-source-llms-moats/.

Ebers, M., Hoch, V. R. S., Rosenkranz, F., Ruschemeier, H., & Steinrötter, B. (2021). The European commission's proposal for an Artificial Intelligence act—A critical assessment by members of the Robotics and AI Law Society (RAILS). *J*, *4*, 589–603.

EO. (2023). *Executive order on the safe, secure, and trustworthy development and use of Artificial Intelligence*. The White House Washington D.C., USA, October. 30.

European Parliament (EP). (2023). 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)), P9_TA(2023)0236. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

Evans, O., Cotton-Barratt, O., Finnveden, L., Bales, A., Balwit, A., Wills, P., … **Saunders, W.** (2021). Truthful AI: Developing and governing AI that does not lie. *ArXiv*. https://doi.org/10.48550/arXiv.2110.06674.

Gailhofer, P., Herold, A., Schemmel, JP., Scherf, CS., de Stebelski, CU., Köhler, AR., Braungardt, S. (2021) *The role of Artificial Intelligence in the European green deal, Study for the special committee on Artificial Intelligence in a Digital Age (AIDA)*, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92.

Hsieh, K. (2019). *Transformer poetry: Poetry classics reimagined by artificial intelligence*. Paper Gains Publishing.

Hu, Y., Jing, X., Ko, Y., & Rayz, J. T. (2021) Misspelling correction with pre-trained contextual language model. In *2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, 144–149.

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A watermark for large language models. *ArXiv*. https://doi.org/10.48550/arXiv.2301.10226.

Kirk, H. R., Jun, Y., Iqbal, H., Benussi, E., Volpin, F., Dreyer, F. A., … **Asano, Y. M.** (2021). Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models. *Advances in Neural Information Processing Systems*, *34*, 2611–2642.

Marsden, C. (2011). *Internet co-regulation and constitutionalism: Towards a more nuanced view*, 29 August, https://ssrn.com/abstract=1973328.

MetaAI. (2023). *System cards: A new resource for understanding how AI systems work*. https://ai.facebook.com/blog/system-cards-a-new-resource-for-understanding-how-ai-systems-work/.

Mökander, J., Sheth, M., Watson, D. S., & Floridi, L. (2023). The switch, the ladder, and the matrix: Models for classifying AI systems. *Minds and Machines*, *33*(1), 221–248.

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, *119*(8), e2120481119.

Organisation for Economic Co-operation and Development. (2022). *OECD's framework for the classification of AI systems*. https://doi.org/10.1787/cb6d9eca-en.

Pagallo, U. (2017) From automation to autonomous systems: A legal phenomenology with problems of accountability. In *International Joint Conferences on Artificial Intelligence Organization (IJCAI-17)*, Melbourne, 17–23.

Pagallo, U. (2022). The politics of data in EU law: Will it succeed? *Digital Society*, *1*(3), 20.

Pagallo, U., Casanovas, P., & Madelin, R. (2019). The middle-out approach: Assessing models of legal governance in data protection, Artificial Intelligence, and the web of data. *The Theory and Practice of Legislation*, *7*(1), 1–25.

Pagallo, U., & Ciani Sciolla, J. (2023). Anatomy of web data scraping: Ethics, standards, and the troubles of the law. *European Journal Data Privacy Law & Technology*, *2*, 1–19.

Pagallo, U., Ciani Sciolla, J., & Durante, M. (2022). The environmental challenges of AI in EU law: Lessons learned from the Artificial Intelligence Act (AIA) with its drawbacks. *Transforming Government: People, Process and Policy*, *16*(3), 359–376.

Pagallo, U., & Durante, M. (2016). The pros and cons of legal automation and its governance. *European Journal of Risk Regulation*, *7*(2), 323–334.

Perez, E., Huang, S., Song, F., Cai, T., Ring, R., Aslanides, J., … **Irving, G.** (2022) Red teaming language models with language models, *ArXiv*. https://doi.org/10.48550/arxiv.2202.03286.

Reed, C. (2012). *Making laws for cyberspace*. Oxford University Press.

Shevlane, T. (2022). Structured access. In J. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. Young & B. Zhang (Eds.), *The oxford handbook of AI governance* (pp. 604–618). Oxford University Press.

Sokolowski, M. M. (2021). Energy efficiency at energy production level: Promoting combined heat and power. In M. M. Roggenkamp, K. J. de Graaf & R. C. Fleming (Eds.), *Energy law, climate change and the environment* (pp. 753–763). Elgar.

Veale, M., & Zuiderveen Borgesius, F. (2021). 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.

von Kirschmann, J. H. (1881). *Zeitfragen und abenteuer*. FF Weber.

Wang, Y., Wang, W., Joty, S., & Hoi, S. C. H. (2021) CodeT5: Identifier-aware unified pre-trained encoder-decoder models for code understanding and generation. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, 8696–8708.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A. … **Gabriel, I.** (2021). Ethical and social risks of harm from language models. In *2021 ACM conference on fairness, accountability, and transparency*, 214–229.

Wilson, C., Marchetti, F., Di Carlo, M., Riccardi, A., & Minisci, E. (2020). Classifying intelligence in machines: A taxonomy of intelligent control. *Robotics*, *9*(3), 1–19.

Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). *Defending against neural fake news*. http://arxiv.org/abs/1905.12616.

Cite this article: Pagallo U. (2025). LLMs meet the AI Act: Who's the sorcerer's apprentice?. *Cambridge Forum on AI: Law and Governance* **1**, e1, 1–13. https://doi.org/10.1017/cfl.2024.6