Defining the Future: The AI Act's Potential in equitably Safeguarding Fundamental Rights and Promoting AI Innovation

Navigating the Act's legislative journey in view of the Forthcoming Regulation

Marinos Emmanouil Kalpakos[*]

Abstract

The recent advancements in the realm of Artificial Intelligence (AI) feature a landscape filled with exciting prospects and lurking dangers in equal measure, necessitating a nuanced, human-centred regulatory approach. The EU is the first organisation to have undertaken this challenge with binding law through its AI Act. This paper delves into the complex task of this founding stone of EU AI law to balance promoting innovation and safeguarding fundamental rights. It addresses both the Act's pre-history – encompassing the

* *Marinos Emmanouil Kalpakos* is an EU Qualified Lawyer and Doctoral Researcher at the Department of Law of the Faculty of Law, Economics and Finance of the University of Luxembourg under the supervision of *Professor Mark David Cole*. This project has been supported by the Luxembourg National Research Fund PRIDE19/14268506.
Contact at: marinos.kalpakos@uni.lu

UFITA 1/2023, DOI: 10.5771/2568-9185-2023-1-128

relevant policy documents before the European Commission's proposal – and its history, which includes the negotiations and amendments leading up to the final trilogue session on the 6th of December 2023 and the consequential provisional agreement, followed by its approval by COREPER and the IMCO and LIBE committees and its final upvoting in the European Parliament on the 13th of March 2024. Keeping track of the Act's legislative journey, including the context in which said journey evolved, shall be crucial in order to comprehensively understand, interpret, and apply what is expected to be one of Europe's most impactful pieces of legislation to date.

The paper aims to bridge a gap in scholarly discussion around the AI Act, which has so far focused solely on evaluating fundamental rights protection or assessing cost-efficiency, providing a practical approach that consolidates the legislative process and appraises the AI Act's overall potential in protecting fundamental rights, while equitably maintaining AI innovation. The research question it seeks to answer is: How can the AI Act promise sufficient fundamental rights protection without compromising AI innovation and overburdening enterprises?

The paper is structured to provide a comprehensive analysis of the Act in view of answering the research question. The Overview section (II.) will give the historical background of AI and track the main policy documents that led to the AI Act's proposal offering a source also for retrospective research in future. It will also identify specific key points of the Commission's original proposal that call for examination and establish its connection to fundamental rights protection, in view of being presented as a product safety regulation. The section on the Legislative State of the Art and Remaining Pitfalls (III.) will present the fruits the legislative process has yielded up until the confirmation of the agreed final draft by the IMCO and LIBE committees on the 13th of February 2024 and its subsequent approval in the Parliament a month later. It will enumerate the main points introduced with the negotiated amendments to then identify still existing potential pitfalls with regards to the Act's scope, enforcement mechanism and future proofness, and how this creates weaknesses for innovation and fundamental rights. The Balancing section (IV.) offers suggestions to counter the weaknesses that were identified in the previous part, uncovering the balancing point between fundamental rights protection and innovation. Finally, the paper will conclude (V.) by summarising the findings and the answer to the research question while leaving room for future discussion.

Die jüngsten Fortschritte im Bereich der Künstlichen Intelligenz (KI) bieten ein breites Spektrum an zugleich aufregenden Möglichkeiten und lauernden Gefahren, weshalb es eines differenzierten, auf den Menschen ausgerichteten regulatorischen Ansatzes bedarf. Mit ihrem verbindlichen KI-Rechtsrahmen (AI Act) ist die EU die erste Organisation, die sich dieser Herausforderung stellt. Im vorliegenden Aufsatz soll die komplexe Zielsetzung dieses grundlegenden KI-Rechtsrahmens auf EU-Ebene untersucht werden, Innovation zu fördern und gleichzeitig die Grundrechte zu schützen. Dabei werden sowohl die Vorgeschichte des Rechtsakts, einschließlich der dem Vorschlag der Europäischen Kommis-

sion vorausgegangenen einschlägigen Strategiepapiere, als auch sein Werdegang – die Verhandlungen und Änderungsvorschläge, die abschließende Trilog-Sitzung am 6. Dezember 2024 und die dabei erzielte vorläufige Einigung sowie seine Billigung durch den COREPER und den IMCO- sowie den LIBE-Ausschuss und die endgültige Abstimmung im Europäischen Parlament am 13. März 2024 – beleuchtet. Einzig durch die Nachvollziehung seines legislativen Werdegangs sowie der entsprechenden Rahmenbedingungen kann dieser Rechtsakt, der eines der einschneidendsten Gesetzespakete auf europäischer Ebene darstellen dürfte, umfassend verstanden, ausgelegt und umgesetzt werden.

Der Aufsatz zielt darauf ab, eine Lücke in der wissenschaftlichen Diskussion um den AI Act zu schließen, deren Gegenstand bislang lediglich die Bewertung des Schutzes der Grundrechte sowie die Beurteilung von Kosteneffizienzbelangen waren. Konkret soll ein praktischer Ansatz verfolgt werden, der eine Konsolidierung des Gesetzgebungsprozesses sowie eine Bewertung des Gesamtpotenzials des KI-Rechtsrahmens zur Sicherung der Grundrechte bei gleichzeitiger Aufrechterhaltung der Innovation im KI-Bereich umfasst. Dabei soll der Fragestellung nachgegangen werden, wie der KI-Rechtsrahmen einen ausreichenden Schutz der Grundrechte bieten kann, ohne dass es zu Beeinträchtigungen der Innovation im KI-Bereich oder zu einer übermäßigen Belastung von Unternehmen kommt.

Die Struktur des Aufsatzes soll eine umfassende Analyse des Rechtsakts im Hinblick auf eine Bearbeitung der genannten Fragestellung bieten. Im Abschnitt zum Überblick (II.) werden der historische Hintergrund der KI sowie die wichtigsten Strategiepapiere beleuchtet, die zum Vorschlag des KI-Rechtsrahmens geführt haben, wodurch eine Quelle für künftige retrospektive Forschung geschaffen wird. Zudem werden zentrale Aspekte des ursprünglichen Kommissionsvorschlags, der als Entwurf für eine Produktsicherheitsverordnung vorgestellt wurde, aufgezeigt, die einer Überprüfung bedürfen, und ein Zusammenhang zum Schutz der Grundrechte hergestellt. Im Abschnitt zum Stand der Gesetzgebung und zu verbleibenden Problemen (III.) werden die Ergebnisse des Gesetzgebungsprozesses bis hin zur Bestätigung des vereinbarten endgültigen Entwurfs durch den IMCO- und den LIBE-Ausschuss am 13. Februar 2024 und der darauffolgenden Zustimmung des Europäischen Parlaments einen Monat später präsentiert. Dabei werden die wichtigsten neuen Elemente aufgeführt, die mit den ausgehandelten Abänderungen implementiert wurden, wobei weiterhin bestehende mögliche Probleme in Bezug auf den Geltungsbereich des Rechtsakts, den Mechanismus für seine Durchsetzung und seine Zukunftssicherheit identifiziert und mit möglichen Risiken für Innovation und Grundrechte in Verbindung gebracht werden. Im Abschnitt zur Ausgewogenheit (IV.) werden Vorschläge unterbreitet, mit denen die im vorherigen Abschnitt identifizierten Schwachstellen angegangen und ein ausgewogenes Verhältnis zwischen dem Schutz der Grundrechte und der Innovation geschaffen werden können. Im abschließenden Abschnitt (V.) werden die Ergebnisse zusammengefasst und die Antwort auf die zugrunde liegende Fragestellung formuliert, wobei Raum für zukünftige Diskussionen gelassen werden soll.

I. Introduction

> "... The underlying advancements of this era are more than just research experiments. They are no longer fantasies of science fiction. They are real and present. The promises of curing cancer, or developing new understandings of physics and biology, or modelling climate and weather. All very encouraging and hopeful. But we also know the potential harms. And we've seen them already. Weaponized disinformation, housing discrimination, harassment of women, and impersonation fraud, voice cloning, deep fakes. These are the potential risks, despite the other rewards. And for me, perhaps the biggest nightmare is the looming new industrial revolution. The displacement of millions of workers, the loss of huge numbers of jobs, the need to prepare for this new industrial revolution in skill training and relocation that may be required. Already, industry leaders are calling attention to those challenges. To quote Chat GPT, this is not necessarily the future that we want. We need to maximise the good over the bad. Congress has a choice now…"[1]

This quote from US senator *Richard Blumenthal*'s opening statement in a Senate hearing on AI paints an accurate portrait of the AI fever that was caused since *ChatGPT* first launched. Drawing what seems to be a bitter-sweet image around AI, Senator *Blumenthal* acknowledges the merits and the risks, while urging Congress to maximise the former over the latter.

Indeed, AI is *en route* to affecting nearly all areas of life. The promise of AI-induced automaticity and enhanced efficiency carries benefits and risks that need to be balanced against each other. In Europe, the gradual development of 'the EU Digital Acquis' has culminated in a first attempt at regulating AI at EU level that aims to strike such a balance, namely the AI Act, as a fundamental rights oriented, product safety regulation.[2]

Ever since the disclosure of the Commission's AI Act proposal scholars, policy makers and non-profit organisations have reinvigorated their contemplation on the intersection of law and technology, from different perspectives, devising in equal measure arguments in favour and against its adequacy with regards to fundamental rights. These arguments scale from a lack of sufficient safeguards, to a misguided or ill-designed focus of said safeguards, to a promising first step towards responsible applications and finally to an over extended regulatory framework with the potential to do more harm than good.[3] Arguments brought forward for the last point are the consideration of the consequences of restricting national legislative initiative, imposing unrealistic expectations on providers and making the EU market in digital technologies considerably less competitive on the world stage.[4]

1 *CBS News*, 'OpenAI CEO Sam Altman testifies at Senate artificial intelligence hearing | full video' (16 May 2023), https://www.youtube.com/watch?v=TO0J2Yw7usM, accessed 19 February 2024.

2 *Bogucki* et al., 'The AI Act and the Emerging EU Digital Acquis' (14 September 2022), https://www.ceps.eu/ceps-publications/the-ai-act-and-emerging-eu-digital-acquis/, accessed 19 February 2024.

3 *Meeri* & *Bruson*, 'Reflections on the EU's AI Act and How We Could Make It Even Better' (14 March 2022), https://www.competitionpolicyinternational.com/reflections-on-the-eus-ai-act-and-how-we-could-make-it-even-better/, accessed 19 February 2024.

4 *Veale* & *Borgesius*, 'Demystifying the Draft EU Artificial Intelligence Act - Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22(4) Computer Law Review International, 97-112.

Scholarly discussion has so far concentrated on an evaluation of either adequate fundamental rights protection or cost-efficiency. What is therefore needed is a more practical approach that consolidates the legislative process the Act has undergone and appraises its overall fundamental rights potential in light of equitably maintaining the merits of AI innovation. This paper aims to fill this gap, addressing the dilemma that was identified by Senator *Blumenthal* to discover a balancing point between the two seemingly counterbalancing notions of fundamental rights protection and innovation.

In search of this balance, this contribution aims to answer the following Research Question: How can the AI Act promise sufficient fundamental rights protection without compromising AI innovation and overburdening enterprises?

For the purpose of clarifying the scales that are weighted against each other, 'AI Innovation' shall be defined as facilitating AI developers in introducing new competitive AI products and promoting technological advancements while preserving the EU internal market as a welcoming field for investments in AI.

Taking into account the legislative journey of the draft act and the feedback of non-profit organisations and academics, while examining it in tandem with the most relevant legislation and case-law under EU digital law, the article will build up to answering the Research Question in the following structure: The Overview section (II.) will give the historical background of AI and track the main policy documents that led to the AI Act's proposal. It will also identify specific key points of the Commission's original proposal that call for examination and establish its connection to fundamental rights protection, in view of being presented as a product safety regulation. The section on the Legislative State of the Art and Remaining Pitfalls (III.) will present the fruits the legislative process has yielded up until the approval of the final text in February 2024 by the Committee of Permanent Representatives (hereinafter 'COREPER'), the Committee on Internal Market and Consumer Protection (hereinafter 'IMCO') and the Committee on Civil Liberties, Justice and Home Affairs (hereinafter 'LIBE') followed up by its upvoting in the Parliament in March 2024. It will enumerate the main points introduced with the negotiated amendments to then identify lurking pitfalls with regards to the Act's scope, enforcement mechanism and future proofness, funneling them to weaknesses for innovation and fundamental rights. The Balancing section (IV.) shall offer suggestions to counter the weaknesses that were identified in the previous part, uncovering the balancing point between fundamental rights protection and innovation. Finally, the paper will conclude (V.) by summarising the findings and the answer to the research question while leaving room for future discussion.

II. Overview: The Journey Towards the AI Act

1. A Journey Through Time: Evolution of AI from Talos to ChatGPT

In order to understand the need for regulatory governance of AI, a historical analysis of the subject matter that is meant to be regulated, spanning from its roots to its modern-day applications, is presented first.

The concept of Artificial Intelligence is often mistakenly perceived as a recent creation. This however could not be further than the truth, as its origins can be traced back centuries to the ancient Greek myth of *Talos*, the automaton guarding princess *Europa* and patrolling the shores of Crete, fending off invaders and would-be kidnappers.[5] Ironically, it is nowadays Europe that strives to protect itself from artificially intelligent automatons.

The modern story of AI truly began in the early 1940s, when American science fiction writer *Isaac Asimov* introduced the contemporary concept of robotics and AI in his short story 'Runaround' and most importantly in his 'Three Laws of Robotics' exhibited therein, which could arguably be considered a precursor of AI ethics.[6] *Asimov*'s science fiction would inspire scientists and researchers paving the road for the establishment of AI as an official academic discipline in the 1950s.[7] At the same time, *Alan Turing* developed 'The Bombe' in order to decrypt the German Enigma code during the Second World War.[8]

AI was established as an academic field in the 1956 Dartmouth Summer Research Project on Artificial Intelligence (hereinafter 'DSRPAI'), a conference held by *Marvin Minsky* and *John McCarthy* at Dartmouth College in New Hampshire with the goal of forming a new research area to develop machines capable of emulating human intelligence.[9] *John McCarthy* used the term 'Artificial Intelligence' during the conference, thus coining the field of AI as an academic discipline.[10] However, in the early 1970s, the new field was treated with scepticism that led to a decline in funding and public interest, marking the beginning of a period of stagnation in AI development.[11]

The explanation to early AI systems falling out of expectations and forfeiting the interest originally invested in them is that they were all Expert Systems, which is a category of artificial systems designed to emulate human intelligence formalising it through a series of top-down "if-then" statements.[12] Such systems, however, severely underperform in areas not suited to formalisation, where tasks require interpreting and learning from external data, characteristics that technically define modern day AI.[13]

The development of artificial neural networks and deep learning laid the groundwork for achieving true AI incorporating the aforementioned traits. Canadian psychologist

5 *Shashkevich*, 'Ancient myths reveal early fantasies about artificial life' (28 February 2019), https://news.stanford.edu/2019/02/28/ancient-myths-reveal-early-fantasies-artificial-life/#:~:text=The%20myth%20describes%20Talos%20as,boulders%20at%20approaching%20enemy%20ships, accessed 19 February 2024.

6 *Haenlein* & *Kaplan*, 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) 61 California Management Review, 5, 6.

7 ibid, 5.

8 ibid, 6-7.

9 ibid.

10 *Walters* & *Novak*, 'Artificial Intelligence and Law' in Cyber Security, Artificial Intelligence, Data Protection & the Law (Springer Singapore, 2021), 39-69.

11 *Haenlein* (n 6) 7.

12 ibid, 8.

13 *Kaplan* & *Haenlein*, 'Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence' (2019) 62(1) Business Horizons, 15-25.

Donald Hebb developed the 'Hebbian learning' theory that mimics the learning process used by human brain neurons, inciting further research on the field.[14]

AI nowadays has achieved breakthroughs only considered to be possible in the realm of science fiction up until merely a year ago. Colossal investments have been made by major companies like *Google* and *Microsoft* competing for the leading role in the AI revolution, making the matter of AI regulation ever more relevant and timely. Human-like AI generated content is currently perhaps the most talked about AI application, with *OpenAI*'s *ChatGPT* and *Dall-E* systems causing global enthusiasm and catching the eye of regulators and tech giants alike along the way.

2. A Journey through the Precursors: Building up to the proposal of the AI Act

Before the European Commission proposed its draft of the AI Act, the EU's approach to AI governance had already started to take form through a series of preceding policy documents. It is therefore important to examine these precursors to the Act's legislative journey before reflecting on the latter's capacity to protect fundamental rights and promote innovation.

The first noteworthy such document was the European Parliament's resolution of 2017 on 'Civil Law Rules on Robotics'.[15] In spite of not mentioning AI by name, it advocated for the incorporation of ethical principles in the EU's legal framework regarding AI and robotics.[16] In the same year, the European Economic and Social Committee (EESC) presented an 'Opinion on AI',[17] underlying the necessity of oversight for AI.[18]

In 2018, the European Group on Ethics in Science and New Technologies (hereinafter 'EGE') placed the next stepping stone through its statement on 'AI, robotics, and autonomous systems',[19] calling for an AI EU framework with emphasis on ethics.[20] Also in 2018, the Commission presented the 'Digital Day Declaration on Cooperation on AI' (hereinafter 'the Declaration') and its AI strategy which was included in the communication

14 *Haenlein* (n 6) 8.

15 European Parliament, 'Report - A8-0005/2017' (27 January 2017), https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html, accessed 19 February 2024.

16 *Stix*, 'The Ghost of AI Governance Past, Present and Future: AI Governance in the European Union' (2021) Pre-print draft, v.6.1., last updated June 18th, 2021, https://arxiv.org/abs/2107.14099, accessed 19 February 2024, 2-12.

17 European Economic and Social Committee, Opinion on 'Artificial Intelligence – The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society' (Own-initiative Opinion) [2017] OJ C 288, 31.8.2017, 1-9.

18 *Stix* (n 16).

19 European Commission, Directorate-General for Research and Innovation, European Group on Ethics in Science and New Technologies, 'Statement on artificial intelligence, robotics and 'autonomous' systems' (2018), Publications Office of the European Union, https://data.europa.eu/doi/10.2777/531856, accessed 19 February 2024.

20 *Stix* (n 16).

'Artificial Intelligence for Europe' (hereinafter 'AI Strategy').[21] The former document arguably marked the first international cooperation on AI governance, while the latter made mention of regulatory sandboxes and aligning AI governance with EU values and fundamental rights.[22] A follow-up document was the 'Coordinated Plan on Artificial Intelligence',[23] which aimed at upgrading the importance of rights emanating from the Charter of Fundamental Rights of the EU while tackling fragmentation and making the EU more competitive globally.[24] Moreover, in 2019, the European Parliament's Committee on Industry, Research, and Energy (hereinafter 'ITRE') issued a report on AI and robotics, stating the need for a 'robust legal and ethical framework for AI'.[25]

Perhaps the most important preceding document was the 'Ethics Guidelines for Trustworthy AI' that was published by the High-Level Expert Group on AI (hereinafter 'AI HLEG Guidelines').[26] The Guidelines state that 'trustworthy AI' is meant to be lawful, ethical and robust.[27] The 'ethical' condition encapsulates four principles that the AI system needs to abide by in order to be trustworthy, namely 'respect for human autonomy, prevention of harm, fairness and explicability'.[28]

These four core principles in turn lay down seven key requirements for achieving trustworthy AI development: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) societal and environmental well-being, and (7) accountability.[29]

The AI HLEG's Guidelines were accompanied by the AI HLEG 2020 'Assessment List for Trustworthy AI for Self-Assessment'.[30] This was the first time an EU tool took into ac-

21 European Commission, 'Digital Day 2018: EU countries to commit to doing more together on the digital front' (9 April 2018), https://ec.europa.eu/commission/presscorner/detail/en/IP_18_2902, accessed 19 February 2024; European Commission, Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions, 'Artificial Intelligence for Europe', COM/2018/237 final.

22 *Stix* (n 16). For other approaches on international level see also *Cole*, AI Regulation and Governance on a Global Scale: An Overview of International, Regional and National Instruments, 1 AIRe (Journal of AI Law and Regulation) (1), 2024, 126-142.

23 European Commission, Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions, 'Coordinated Plan on Artificial Intelligence', COM/2018/795 final.

24 *Stix* (n 16).

25 ibid; European Parliament, 'Report - A8-0019/2019' (30 January 2019), https://www.europarl.europa.eu/doceo/document/A-8-2019-0019_EN.html, accessed 19 February 2024.

26 European Commission, 'Ethics guidelines for trustworthy AI - High-Level Expert Group on AI' (8 April 2019), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, accessed 19 February 2024.

27 *Stix* (n 16).

28 ibid.

29 ibid.

30 European Commission, 'Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment' (17 July 2020), https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment, accessed 19 February 2024.

count the concept of 'trustworthy AI' throughout an AI system's life cycle and attempted to transpose ethical principles into actionable measures for all involved stakeholders in AI.[31] The Commission itself endorsed the approach of 'trustworthy AI' in its communication on 'Building Trust in Human-Centric AI'.[32]

Finally, on February 19, 2020, the Commission published its 'White Paper on Artificial Intelligence', laying down most of the features that are also found in the AI Act, such as the risk categorization and the horizontal requirements for high-risk AI systems.[33] Largely influenced by the AI HLEG Guidelines, the White Paper paved the road for the Commission's proposal of the AI Act on April 21, 2021.[34]

3. Finally there? Key Points of the AI Act and Fundamental Rights Connection

Having gone through the historical evolution of AI and the background of the AI Act's formulation, before appraising it, this subsection shall give a bird's eye view to the main points of the Commission's original proposal while laying down its connection with fundamental rights.

After the release of the White Paper, the Commission conducted impact assessments and initiated a stakeholder consultation based on the feedback of which it proposed the long awaited AI Act on April 21, 2021.[35] Being a product safety regulation based on Article 114 TFEU, its general objective is, according to the Explanatory Memorandum of the Commission's proposal, 'to ensure the proper functioning of the single market by creating conditions for the development and use of trustworthy AI in the Union'.[36]

The original draft pursues its objective following a risk-based approach to establish a regulatory framework consisting of four risk categories, complemented by a layered enforcement mechanism.[37] In this context, stricter requirements are applied as the level of risk increases.[38] Systems with an unacceptable level of risk are entirely prohibited and mandatory requirements are imposed only on what qualifies as 'high-risk' AI systems,

31 *Stix* (n 16).

32 ibid; European Commission, Communication to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, 'Building Trust in Human-Centric Artificial Intelligence', COM/2019/168 final.

33 European Commission, White Paper on Artificial Intelligence - A European Approach to Excellence and Trust, COM/2020/65 final.

34 *Stix* (n 16) 12-19.

35 ibid.

36 European Commission – Explanatory Memorandum of the AI Act, 'Proposal For A Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts', COM/ 2021/ 206 final, 9.

37 *Kop*, 'EU Artificial Intelligence Act: The European Approach to AI' (2021) Issue Stanford-Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021, 3.

38 ibid.

whereas providers of non-high-risk AI systems can voluntarily opt for compliance with a code of conduct.[39] Systems posing limited risk are only subject to certain transparency requirements, with only minimal risk AI systems escaping transparency obligations altogether.[40]

The requirements addressed to high-risk AI systems are articulated in Title III Chapter II of the Commission's original proposal and closely resemble the ones of the Ethics Guidelines, namely: 'data and data governance, technical documentation, record keeping, transparency and provision of information to users, human oversight and accuracy, robustness and cybersecurity'.[41] On top of these requirements, actors responsible for compliance need to build their systems in accordance with Title III Chapter II ('by design' principles can be spotted here), carry out internal conformity assessments, establish a post market monitoring system and register their AI system with the EU Database filling the EU Declaration of Conformity prior to the system receiving the CE marking.[42] Constant monitoring of the AI system, both upstream and downstream, seemed to be the end goal.[43] The original draft also includes regulatory sandboxes attempting to afford breathing room to research institutions and SMEs.[44]

In spite of the Act being designed as a product safety regulation based on the internal market, concepts such as 'trustworthy AI' and 'responsible innovation' as well as direct references to AI ethics and the policy documents endorsing them are abundant in its Explanatory Memorandum. Although evangelising the promotion of market integration, the Act is, to its greater extent, a fundamental rights oriented regulation aiming to foster responsible AI innovation in the single market. As explicitly stated in the Explanatory Memorandum, the Act 'seeks to ensure a high level of protection for those fundamental rights and aims to address various sources of risks through a clearly defined risk-based approach'.[45] The AI Act thus acknowledges the potential impact of AI systems on the Charter of Fundamental Rights.

III. AI Act Unveiled: Traversing the Legislative Process and Identifying Remaining Pitfalls

The author as any researcher and observer active in the years following the proposal had the privilege to experience the dynamic legislative journey of the AI Act. At the time of writing this contribution, the legislative process had culminated into the trilogue session of the 6th of December 2023 with the subsequent approval of the compromise text by COREPER and

39 Explanatory Memorandum (n 36) 9.
40 *Kop* (n 37) 5.
41 *Stix* (n 16) 16-17.
42 ibid.
43 *Kop* (n 37) 4.
44 ibid, 9.
45 Explanatory Memorandum (n 36) 14-16.

its endorsement by the Parliament's IMCO and LIBE committees in February 2024. The produced final draft was upvoted in the Parliament on the 14th of March 2024.

Going back one year before the final trilogue, the Council adopted its General Approach on the Commission's original proposal on the 6th of December 2022, while on the 11th of May 2023 IMCO and LIBE adopted a draft negotiating mandate including amendments to the Commission's proposal.[46] These amendments where upheld in the plenary vote of the 14th of June by the European Parliament forming the backbone of the Parliament's mandate in the following trilogue sessions. As per the relevant press release, the MEPs' aim with the proposed amendments was 'to ensure human oversight of AI, systems that are safe, transparent, traceable, non-discriminatory, and environmentally friendly and a uniform, future proof, technology-neutral definition for AI'.[47] Specifically, according to *Dragoș Tudorache*, the rapporteur in the LIBE committee, the AI Act is a crucial legislation with the potential of leading the way at making AI human-centric, trustworthy and safe.[48] The rapporteur further claimed that under the Parliament's amendments, the AI Act fosters AI innovation in Europe allowing start-ups and SMEs space to grow, while ensuring fundamental rights, democratic oversight, robust AI governance and enforcement across the EU.[49] *Svenja Hahn*, the shadow rapporteur in the IMCO committee, praised the balanced approach of the Parliament's version in ensuring civil rights protection and promoting innovation and economic growth and highlighted the capacity of regulatory sandboxes to turn the EU into 'a hotspot for research and innovation on artificial intelligence'.[50]

Undoubtedly a legal milestone as the first legislative endeavour to regulate AI and with the Parliament having suggested significant amendments towards what seems to be the right direction, the AI Act may nonetheless appear to fall short in comparison to the confident and optimistic statements of the rapporteur and shadow rapporteur. This can be especially argued in light of many relevant parts of the Parliament's mandate not having survived the negotiations that followed since. Keeping track of the key advancements the legislative process has yielded is therefore necessary in order to dive into the 'could haves' and 'should haves' of the Act's final version, putting those statements to the test.

After introducing (1.) the legislative state of the art along with the main developments the legislative process has yielded, this section shall identify the remaining main weak points of the Act, examining them into three overarching bundles. These will be (2.) the Act's under-comprehensive scope and apparent inconsistencies with EU digital legislation, (3.) its inadequate ecosystem of oversight and (4.) frailties jeopardising the regulation's future proof capacity. The section shall then conclude with (5.) a classification of the identi-

46 European Parliament News, 'AI Act: a step closer to the first rules on Artificial Intelligence' (11 May 2023), https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence, accessed 19 February 2024.

47 ibid.

48 *Rhawi*, 'AI Act: EU paves the way for world's first comprehensive rulebook on AI' (11 May 2023), https://www.reneweuropegroup.eu/news/2023-05-11/ai-act-eu-paves-the-way-for-worlds-first-comprehensive-rulebook-on-ai, accessed 19 February 2024.

49 ibid.

50 ibid.

fied shortcomings into those concerning fundamental rights protection and those touching upon innovation, paving the road for exploring avenues to equitably reconcile both in the final part.

1. Legislative State of the Art

With the final trilogue spanning over 38 hours of negotiations, a provisional agreement was reached between the Council presidency and the European Parliament on the 9th of December 2023, officially making the EU the first regulator on AI. The relevant press releases of the Council and the Parliament along with the opening consolidation of the agreement as submitted to COREPER give an overview of the final text.

Specifically, AI systems are classified on the basis of the risk they pose, with an extended list of prohibited AI, yet maintaining the possibility of law enforcement to use remote biometric identification in public spaces, provided strict conditions are met.[51] The Act's scope has been confirmed with the definition of an AI system aligning with the one proposed by the Organisation for Economic Co-operation and Development (hereinafter 'OECD') and concrete exceptions from the Act's application regarding national security, military and defence, and research and innovation.[52] An updated governance regime with enforcement authorities at the EU level has been agreed. These include the AI Office, a function of the Commission with specific tasks about general purpose AI models, the AI Board, made up of Member States' representatives to ensure they shall have a say in the implementation of the regulation, and the Advisory Forum for stakeholders, which will include civil society, academia and start-ups to advise the AI Board.[53] The agreement also features fundamental rights impact assessments for high-risk AI systems, an obligation for registration of said systems with the EU database touching upon public entities and an additional obligation to inform natural persons that are being exposed to emotion recognition systems.[54] Furthermore, the provisions on AI regulatory sandboxes have been updated to allow for the testing of innovative AI systems in real world conditions, outside sandboxes, under specific safeguards.[55]

In the aftermath of the AI fever induced by *OpenAI*'s release of *ChatGPT* in November 2022, it is not surprising that public discussion has been mostly revolving around generative AI systems.[56] It is thus equally unsurprising to see new rules on general purpose AI and foundation models included in the compromise agreement. Regulating foundation

51 Council of the EU, 'Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world' (9 December 2023), https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/, accessed 19 February 2024.

52 ibid.

53 ibid.

54 ibid.

55 ibid.

56 *Frank*, 'Draft of the AI Act gets on the home stretch' (12 May 2023), https://www.taylorwessing.com/en/insights-and-events/insights/2023/05/draft-of-the-ai-act, accessed 19 February 2024.

models was the stumbling stone that nearly caused the entire regulation proposal to collapse towards the final phase of the negotiations.[57] This happened when representatives, most notably, from France and Germany, pressured by their leading national AI companies, *Mistral* and *Aleph Alpha* respectively, pushed against regulating foundation models altogether at a meeting of the Telecom Working Party on the 9th of November 2023.[58] The compromise that was reached is a blend of the Council's and the Parliament's mandates, adopting the former's terminology of 'general purpose AI models' instead of 'foundation models' and the latter's approach of imposing mandatory obligations.[59] General purpose AI and the general purpose AI models they may be based on, will have to comply with EU copyright law, draw up technical documentation and deliver summaries of the content on which they were trained.[60] High-impact general purpose AI models with systemic risk will be subjected to stricter obligations, including conducting model evaluations and adversarial testing, assessing and mitigating systemic risks and ensuring cybersecurity.[61]

It is important to keep track of the legislative context in which the Act has been adopted during the negotiations to better understand and interpret it. The most recent, currently available version of the AI Act is the final draft submitted to the Parliament for the plenary of the 13th of March 2024 which had already been approved by COREPER and, a few days later, endorsed by IMCO and LIBE. This contribution shall thus use this final draft along with the one submitted to COREPER and the final four-columns document consolidating the mandates of all the legislative bodies, to trail and appraise the legislative progress of the Act.[62] Recitals, Articles and parts of the Act mentioned henceforth shall refer to said

57 *Bertuzzi*, 'EU's AI Act negotiations hit the brakes over foundation models' (15 November 2023), https://www.euractiv.com/section/artificial-intelligence/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/, accessed 19 February 2024.

58 ibid.

59 *Keller*, 'A Frankenstein-Like Approach: Open Source in the AI Act' (14 December 2023), https://openfuture.eu/blog/a-frankenstein-like-approach-open-source-in-the-ai-act/, accessed 19 February 2024.

60 European Parliament News, 'Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI' (9 December 2023), https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai, accessed 19 February 2024.

61 ibid.

62 Final Draft: Amendments by the European Parliament to the Commission Proposal, Amendment 808, Report - A9-0188/2023, 'Regulation (EU) 2024/... of the European Parliament and of the Council of ... 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)' (6 March 2024), as submitted to the European Parliament in view of the final vote. Available at https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_EN.html?redirect, accessed 20 March 2024;

documents unless specified otherwise while their numbering is still subject to change in view of the final publication in the EU Official Journal. The next subparts shall depict the outline of the most relevant outcomes of the negotiations along with the respective parts of the final text. Doing so in an allegorical reference to a well-known 1960s movie, with inspiration having been drawn for this from a contribution by *Veale*, said outcomes shall be tiered as good, bad, and ugly.

a) The Good

The long-awaited definition of AI was a focal point of the negotiations. Article 3 of the Parliament's mandate included a detailed catalogue of definitions endorsing the OECD definition of AI, which was upheld in the final trilogue. The latter defined an AI system as 'a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments'.[63] The definition that had been suggested in the Council's mandate did away with the terms 'machine-based' and 'varying levels of autonomy'. Instead, it sufficed itself to 'elements of autonomy' explicitly mentioning the use of 'machine learning and logic-and knowledge-based approaches' to produce outputs based on human or machine provided inputs.[64] This version of the definition could possibly have been meant to specify that an AI system is not just any software application but one that employs such sophisticated computational techniques to perform its tasks. In the final draft, an AI system is defined as 'a machine-based system 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'. It can be observed that the final definition follows up on the one suggested by the Parliament with clarifications regarding adaptiveness and outputs.

Draft: Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Analysis of the final compromise text with a view to agreement' (Interinstitutional File: 2021/0106(COD)) (26 January 2024) (OR. en) 5662/24, as submitted to the Permanent Representatives Committee. Available at https://www.linkedin.com/posts/luca-bertuzzi-186729130_st05662en241pdf-activity-7157679718433980418-qrT2?utm_source=share&utm_medium=member_desktop, accessed 16 February 2024;

Final all mandates document: European Parliament and Council, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 2021/0106(COD), DRAFT [Final 4-column draft as updated on 21/01] (21 January 2024). Available at https://www.linkedin.com/posts/luca-bertuzzi-186729130_aiactfinalfour-column21012024pdf-activity-7155091883872964608-L4Dn?utm_source=share&utm_medium=member_desktop, accessed 16 February 2024.

63 ibid.

64 ibid.

When it comes to generative AI, although the eventual compromise opted for the Council's terminology of 'general purpose AI models', it is important to keep in mind that Article 3 (1c) and (1d) of the Parliament's mandate introduced definitions for both the terms of 'foundation model' and 'general purpose AI system'. These were respectively defined as 'an AI 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' and 'an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed'.[65] Even though it did not make it into the final text, it is noteworthy to mention that the former definition traces its origins to a report from the *Center for Research on Foundation Models* (*CRFM*) of Stanford University explicitly stating that the term 'foundation model' was chosen to express the important yet unfinished status of such models.[66] A more detailed definition of general purpose AI was put forward by the Council, indicatively enumerating certain functions that such AI can perform and the fact that it can be utilised and integrated across multiple contexts and other AI systems respectively, regardless of how it was distributed, including open source software. The final draft went with a definition describing a general purpose AI model as an AI model displaying significant generality, capable of various tasks regardless of market deployment and that can be integrated into a variety of downstream systems or applications, excluding pre-market AI releases for research, development and prototyping activities. A general purpose AI system is an AI system based on a such a model, capable of serving various purposes either for direct use or integrated in other AI systems.

However, the definition of generative AI *per se* still eludes the catalogue of Article 3 of the final draft, to be found instead only in Article 28b (4) of the Parliament's mandate,[67] as 'foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI")'. It could therefore have been deduced that the term 'generative AI' was planned to be a subsection of the term 'foundation model', subject to additional transparency requirements. Specifically, Article 28b (2) of the Parliament's mandate included seven requirements to be met by foundation models, namely: setting up a risk handling system, using proper databases, ensuring quality, applying energy efficiency standards, setting up technical documentation, establishing quality management and registering the foundation model. Additionally to those, generative AI systems, under Article 28b (4) would need to comply with transparency obligations pursuant to Articles 52 (1) and (2), to ensure adequate safeguards against the generation of content violating EU law and to publicly document the use of copyrighted training data. This last point could have become a starting

65 ibid.

66 *Bommasani* et al., 'On the opportunities and risks of foundation models' (2021) Center for Research on Foundation Models (CRFM) Stanford Institute for Human-Centered Artificial Intelligence (HAI) Stanford University, https://arxiv.org/abs/2108.07258, accessed 19 February 2024, 3-7.

67 *Frank* (n 56).

point for future discussions on the reach and limits of copyright protection.[68] Although the final text did away with these provisions entirely, it would not be impossible to see similar notions coming up in future guidelines.

Besides the newly added definitions, the Parliament's mandate introduced, in Article 4a, overarching general principles for AI development and use covering all AI systems, leaning towards the more principles-based regulatory approaches seen in the UK, US, Singapore, and Japan, as observed by technology law expert *Sarah Cameron*.[69] These principles included human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, and social and environmental well-being.[70] In this point, if the provision had made it to the final text, we could have observed the culmination of AI ethics from being initially expressed in *Asimov*'s fiction, to being included in the AI HLEG's 'Ethics Guidelines for Trustworthy AI', to finally formulating legally binding general principles that will soon be part of Europe's secondary legislation.

While these principles would have been unlikely to cause global disagreement, differing territorial understandings of them, such as what qualifies as 'technical robustness' being different in the US than in Europe, may have unnecessarily hindered establishing a 'coherent human-centric European approach'.[71] This is something that should be kept in mind during the enforcement of the Regulation, in view of better clarifying regulatory requirements and avoiding reference to territorial understandings,[72] in case general principles re-emerge through soft law.

The final text features a list of prohibited AI practises in Article 5.[73] This includes AI utilising subliminal, purposefully manipulative or deceptive techniques with the result of impairing a person's or group's ability to make informed decisions in a manner that is at least likely to cause significant harm. The list goes on with AI that exploits vulnerabilities due to age, disability or social or economic situations in a manner that is at least reasonably likely to cause significant harm. The addition of the word 'reasonably' results in an interesting phrasing hinting at a would-be tiered-like approach with manipulative AI being treated more strictly than AI exploiting vulnerabilities. Biometric categorisation for sensitive attributes is banned with an exception for lawfully acquired biometric datasets in the area of law enforcement. Social scoring is prohibited and real time biometric identification in public spaces for law enforcement is also banned unless necessary to find victims

68 ibid.

69 *Pinsent Masons*, 'MEPs' EU AI Act proposals focus on 'foundation models'' (16 May 2023), https://www.pinsentmasons.com/out-law/news/meps-eu-ai-act-foundation-models, accessed 19 February 2024.

70 Mandates (n 62).

71 *Frank* (n 56).

72 ibid.

73 Final draft (n 62).

of abduction, trafficking and missing persons, to prevent threat to life, safety and terrorist attacks and for identification of persons suspected of committing criminal offenses.

The final draft expanded the classification of high-risk AI in Annex III and Article 6 to include AI touching upon biometrics, critical infrastructure, education, employment, access to essential services, law enforcement, migration and administration of justice.[74] AI systems used to influence the outcome of an election, referendum or voting behaviour have been added as well to the high-risk list of Annex III, however the Parliament's attempt to also include AI used in social media recommender systems did not prevail.[75]

Amongst the key advancements introduced in the final draft is the obligation for deployers of AI systems, who are distinguished from AI providers, to conduct a fundamental rights impact assessment (hereinafter: 'FRIA') before deploying a high-risk AI system. This is mentioned in Recital 93, which highlights the better position of deployers to understand the real-world usage of the AI system and identify significant risks.[76] Article 27 lays down and details the obligation, and Annex VIII Section C (3) includes a summary of the findings of the relevant assessment into the information to be submitted for the registration of the AI system in the EU Database for High-Risk AI Systems.

The initially proposed European Artificial Intelligence Board was meant to be a supranational supervisory authority fashioned at the image of the European Data Protection Board of the GDPR and the European Board for Digital Services of the Digital Services Act.[77] Although the Council maintained the Commission's phrasing, in what appeared to be an effort to secure larger civil engagement, Article 56 of the Parliament's mandate attempted to replace the originally proposed Artificial Intelligence Board with an Artificial Intelligence Office, with a legal personality of its own, acting as an independent EU body within the Commission established in Brussels.[78] Given the considerable increase in the AI Office's duties this would entail, the Parliament seemed to want to draw further competencies to new EU bodies, in an especially dynamic field.[79]

However, what eventually came to pass in the final draft, is that Article 65 establishes the AI Board, comprised of one representative from each Member State, with the European Data Protection Supervisor as an observer. The subsequent Articles lay down the Board's tasks. Article 64 entrusts governance at the EU level to the AI Office, which, as per Article 3 (47), is now a function of the Commission contributing to the implementation, monitoring and supervision of AI systems, general purpose AI models and AI governance. According to said Article, the EU AI Office and the Commission are interchangeable in the text of the Act. Additional governance bodies established with an advisory role are,

74 ibid.

75 ibid.

76 ibid.

77 *Raposo*, 'Ex Machina: Preliminary Critical Assessment of the European Draft Act on Artificial Intelligence' (2022) 30(1) International Journal of Law and Information Technology, 91.

78 Mandates (n 62).

79 *Frank* (n 56).

in Article 68, the scientific panel of independent experts and, in Article 67, the advisory forum, comprised of diverse stakeholders including industry, start-ups, SMEs, civil society and academia.

When it comes to making sure that the AI Act shall not be a paper tiger, Article 99 of the final draft predicts considerable fines for non-compliance. Specifically, non-compliance with Article 5 is subject to fines of up to €35 million or, if the offender is a company, up to 7% of its total worldwide annual turnover, whichever is higher.[80] Samewise, the fine shall be €15 million or 3% of total worldwide annual turnover for violations contained in Article 99 (4).[81] Finally, providing incorrect, incomplete, or misleading information in answering a request by a notified body or national competent authority can result in fines of up to €7,5 million or 1,5% of total worldwide annual turnover.[82] More proportionate caps for SMEs and start-ups are also provided for in Article 99 (6) with the fine capped at whichever, amount or percentage, is lower instead of higher.[83]

b) The Bad

The negotiations have significantly changed the landscape of the AI Act. However, certain issues remain alarming or altogether unaddressed, in particular with regards to loopholes, standardisation bodies, remedies and civil society participation.

In spite of expanding the list of high-risk AI systems, Article 6 potentially creates a loophole for AI systems to not qualify as high-risk to begin with, since besides being intended to be used in one of the high-risk areas, AI systems now also need to pose a significant risk to health, safety or fundamental rights of natural persons to be considered high-risk.[84] The vagueness of this requirement could be abused by AI developing companies in order to circumvent their obligations under the AI Act, claiming that the risk their systems pose is not 'significant' enough.[85]

Furthermore, fundamental rights related issues have after all not been excluded under Article 40 from the scope of harmonised standards developed by private standardisation bodies.[86] Said bodies in the view of some seem to have insufficient fundamental rights expertise and an appetite for disregarding the participation of civil society in their procedures.[87] The status of such standardisation bodies, namely the European Committee for Standardization (hereinafter: 'CEN'), the European Committee for Electrotechnical Stan-

80 Council (n 51).

81 ibid.

82 ibid.

83 ibid.

84 *European Center for Not-for-Profit Law*, 'Hope on the horizon for digital civic space, as EU parliament advances protection of rights' (11 May 2023), https://ecnl.org/news/hope-horizon-digital-civic-space-eu-parliament-advances-protection-rights, accessed 19 February 2024.

85 ibid.

86 ibid.

87 ibid.

dardization (hereinafter: 'CENELEC'), and the European Telecommunications Standards Institute (hereinafter: 'ETSI'), has in fact been upgraded to having a permanent position in the advisory forum that participates in the creation of said standards as per Article 67 (5) and Recital 121 in conjunction with Article 40 of the final draft.

Although the inclusion of remedies is a significant step towards empowering people affected by AI, the expected protection is not yet complete. Specifically, Article 85 of the final draft grants affected natural persons and groups the right to lodge a complaint with a Market Surveillance Authority (hereinafter 'MSA') in case of an infringement. Additionally, Article 86 grants affected individuals the right to request from deployers a clear and meaningful explanation of a decision they are subject to, and which was based on output from a high-risk AI. However, the proposed right to individual redress does not receive a judicial dimension, since Article 68b of the Parliament's mandate stipulating a right to an effective judicial remedy against legally binding decisions of a national supervisory authority was not upheld in the final text. With the primary complaint only allowed to be lodged with the MSA, the deployer or AI provider is not the direct addressee of the scrutiny in any case, nor is there a right to directly claim damages from them enshrined in the yields of the AI Act negotiations, outside the AI Liability Directive.[88]

Finally, in spite of Article 110 of the final draft showing promise by including the AI Act to Annex I of Directive 2020/1828, the AI Act still does not explicitly enable affected parties to mandate public interest organisations to represent them. Civil society organisations still do not have the right to report violations directly, without representing a specific affected individual. As observed by the *European Center for Not-for-Profit-Law*, this is particularly important if the affected individual is in a vulnerable position where they may be afraid of the consequences of filing a complaint, such as if they are an activist, a refugee, or an employee.[89]

c) And the Ugly?

Having gone through the main points of the negotiations' outcome, it is possible to draft an outline of the weaknesses that are still clinging to the Act. We have observed the horizontal Act attempting to embrace general principles that are also applicable sectorally, which was to an extent peer pressured by regulatory initiatives elsewhere. We have also seen mere infant steps at securing redress for affected persons and establishing enforcement. All this was spectated from the point of view of a society that has been overwhelmed by the prospects and threats of AI, experiencing huge breakthroughs over the course of mere months. The capacity of the draft regulation to remain relevant, in spite of the staggering speed in which the realm it aims to regulate changes, is perhaps the most crucial challenge

88 Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM(2022) 496 final, 2022/0303 (COD).

89 *European Center for Not-for-Profit Law* (n 84).

it must meet. This has also been recognised, as we have noted, in the abandoned definition of 'foundation models'.

Succumbing to peer pressure, taking timid steps and pointing out uncertainty are all symptoms showing the draft Act still lacks confidence in accomplishing its ultimate task of effectively balancing fundamental rights protection and AI innovation in view of the AI revolution. With this realisation as a starting point, weak points begin to emerge that can be bundled into three distinct groups that shall be subsequently examined. Specifically:

(1) The remaining under-comprehensive scope of the Act. Understanding the real reach of the legislation is necessary in order to evaluate its impact and effectiveness.
(2) The lack of an effective ecosystem of enforcement and oversight. Adequate fundamental rights protection is heavily dependent on successful enforcement.
(3) The questionable future proof capacity of the draft regulation. A more than relevant point to consider given that legislation that aims to foster innovation in a dynamic field like AI needs to be equally dynamic.

2. The remaining under-comprehensive scope of the Act

In order to identify key remaining weak points in the Act, it is necessary to examine the scope it has achieved under the negotiations. This is so because the forthcoming legislation's range will need to be adequate to meet the challenge of effectively protecting individuals' fundamental rights while fostering innovation.

In the Explanatory Memorandum of the Commission's proposal, the AI Act was characterised as a 'balanced and proportionate horizontal regulatory approach to AI,' that is confined to the 'minimum necessary requirements to address the risks and problems associated with AI'.[90] AI was originally defined in Article 3 of the Commission's proposal as 'software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'. Albeit allegedly broad, the original definition was meant to cover only a marginal 5 % to 10 % of AI products and services likely to enter the market in the next few years.[91]

The original definition was more specific, focusing on software-based AI systems. The one of the OECD that was upheld in the final trilogue on the other hand, seems to have a greater reach. It defines AI as a machine-based system that can function with different degrees of autonomy impacting both real and virtual settings, regardless of the precise techniques or methodologies utilised in its development. Therefore, the amended version of the Act covers both software-based and machine-based systems.

Although the new definition of AI includes a wider range of AI systems, this expansion was originally triggered by the need to also cover large language models and generative AI in light of recent technological advances. Besides the issue of future proofness, which will

90 *Bogucki* (n 2) 2-6.
91 ibid.

be examined later on, the observation of the incidental nature of this inclusion also raises questions with regards to the adequacy of the Act's scope.

Circumstantially increasing the range of the draft regulation causes legal certainty deficiencies with regards to its scope, unravelling the underlying problematic reasoning of relying on a black-letter approach to regulate an unprecedented ever-growing industry. While at first glance the new definition of AI is more inclusive than its predecessor in the proposal stage, it has in fact only addressed an additional aspect of the Act's scope, it has not adequately expanded it *per se* in light of advances that are yet to be seen.

It can be argued that a similar realisation could have led to the adoption of a more flexible, principle-based approach, such as those of the UK, US, Singapore, and Japan, which had also partially been endorsed by the Parliamentary amendments, even though not reflected though the definition of AI. The comparison of the top down European approach of the AI Act to the more flexible industry or sector-led approach preferred elsewhere, also breeds questions as to the suitability of the regulation's horizontal character.[92] However, it has been argued that neither a completely horizontal, as endorsed by the EU, nor an entirely vertical approach, as somewhat preferred by China, would be enough on its own when it comes to fundamental rights protection.[93]

Moreover, the list of prohibited AI remains an exhaustive enumeration of a pre-identified, limited number of high-risk AI systems, which, on top of possibly allowing for dangerous loopholes as already explained, still does not fully address societal harm.[94] Specifically, societal harm takes place when an interest of society at large is wrongfully jeopardised.[95] Three examples have been identified by *Nathalie A. Smuha* with regards to societal harms potentially caused by AI, namely AI powered facial recognition, voter manipulation, and public decision making, all concerning different societal interests: equality, democracy and the rule of law respectively.[96] Safeguarding such societal interests subsequently contributes to the effective protection of fundamental rights.

Even though facial recognition has been prohibited and AI tools used to influence voters have been added to the high-risk AI list of the final draft, the amended version still does not offer a comprehensive inclusion of societal harm as one of the risks it offers protection against, besides a mere reference in Recital 5 of the final draft and Recital 9a of the Parliament's mandate. The latter did not make it into the final text while both appear to be inadequate to sufficiently guarantee protection against AI societal risks.

92 *Scott*, 'Digital Bridge: Global AI rulebook — US digital policymaking — Data rules' (Brussels, 20 April 2023), https://www.politico.eu/newsletter/digital-bridge/global-ai-rulebook-us-digital-policy making-data-rules/, accessed 19 February 2024.

93 *O'Shaughnessy & Sheehan*, 'Lessons From the World's Two Experiments in AI Governance' (14 February 2023), https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035?utm_source=substack&utm_medium=email, accessed 19 February 2024.

94 *Bogucki* (n 2) 2-6.

95 *Smuha*, 'Beyond the Individual: Governing AI's Societal Harm' (2021) 10(3) Internet Policy Review, https://doi.org/10.14763/2021.3.1574, accessed 19 February 2024, 6-12.

96 ibid.

Finally, as observed by *Artur Bogucki*, a noteworthy weakness in the scope of the AI Act is its focus on a specific, linear risk-based approach emphasising a one-to-one link between an AI system, its risk potential, and its user to address risks associated with a particular AI system.[97] This approach however neglects the potential risks for individuals and society at large that may arise when more than one AI systems interact with each other, known as interactive risks of AI, causing a potential gap in the legislation.[98]

3. The lack of an effective ecosystem of enforcement and oversight

The AI Act should not be limited to theory instead of practice. Criticism with regards to its limitations in terms of enforcement due to the nature of enforcement instruments and the lack of an effective ecosystem of oversight has been prominent since the Commission first announced its proposal.[99] The final draft introduced improvements through specifying the role of national competent authorities in Article 70, as did the explicit inclusion of rights of redress under Articles 85 to 87. However, the changes seem insufficient to establish an effective ecosystem of enforcement and oversight. Keeping up with the trilogue, inadequacies can be spotted with regards to post-market enforcement, accountability, bottom-up enforcement through Article 71's central database, as well as delegation of rule-making to private bodies through the standardisation process, public feedback and societal rights.

The EU AI Office, although not as powerful as envisioned by the Parliament, remains a notable feature when it comes to the enforcement of the regulation on a European level. Yet, little has changed in the current version of the Act when it comes to the bodies that are meant to operate it on a national level. Articles 28 to 39 set up certain national organs, specifically notified bodies and notifying authorities tasked with assessing the conformity of high-risk AI systems and designating, assessing and monitoring the notified bodies respectively.[100] The most noticeable changes with regards to notifying authorities are the explicit provisions in Article 28 (7) requiring qualified personnel with a sufficient background in fundamental rights protection to form their ranks and Article 34 (2) minimising burdens for providers in view of the undertaking's size, amongst other factors. The role of the Commission in the notification procedure has also been upgraded as Article 30 (5) stipulates that when objections are raised concerning the assessment, consultation including the Commission, the relevant Member State, and the conformity assessment body is needed in view of the Commission deciding whether the authorization is warranted.

Components of post-marketing controls, inspired by pharmacovigilance, that are all the more included in New Legislative Framework regimes, can also be traced in the AI Act.[101] The most prominent role in national enforcement lies with the national competent authorities which are meant to act as MSAs. MSAs are public bodies, usually government

97 *Bogucki* (n 2) 2-6.
98 ibid.
99 *Bogucki* (n 2) 25.
100 *Raposo* (n 77) 91.
101 *Veale* (n 4) 110 - 112.

departments and regulatory agencies, that have broad powers, spanning from information gathering, imposing administrative fines and withdrawing products to compelling intermediaries to cease offering non-compliant AI products or co-operate with the authority to mitigate risks.[102] Same as for notifying authorities, MSAs are required to employ personnel with adequate knowledge of fundamental rights besides AI technologies according to Article 70 (3), while Member States must biannually report to the Commission with regards to the financial, human resources and adequacy of the national competent authority as per Article 70 (6).

Literature has considered the post-market enforcement processes of the original version of the AI Act weak, with MSAs being deemed less active than data protection authorities (hereinafter 'DPAs') due to the latter's increased experience with AI and data related matters.[103] The negotiations did not come a long way in considering this critique, with only a minor addition in the text of Article 59 (8) of the Parliament's mandate making the European Data Protection Supervisor competent also for the coordination besides just the supervision of EU bodies that fall under the scope of the Regulation. However, this did not make it into the final text.

Highlighted as a powerful instrument for the protection of fundamental rights by *Vera Lúcia Raposo*, stakeholder's accountability in the AI Act assigns responsibilities and duties to both AI providers and deployers.[104] Most notably, deployers of AI systems are called to monitor and report any new risks they discover to providers, while providers must inform the MSA if post-marketing monitoring shows risks or non-compliance.[105]

However, in spite of affected individuals now having a right to complain to an MSA under Articles 85 to 87, there remains to be no civil liability mechanism to claim compensation directly from AI providers and deployers, while the concept of AI personhood is still left unaddressed.[106] There is perhaps hope for the AI Liability Directive that is currently being prepared to address this gap.[107]

Article 70 introduces a central database for the registration of all high-risk AI systems. This database, fashioned after the standards in the new Medical Devices Regulations, is intended to assist MSAs in keeping track of non-compliant systems.[108] It establishes further accountability measures for deployers who are also acting as providers, obliging them to register internally developed systems for their own use.[109] However, this expansion may lead firms to contest this requirement in court, citing violations of trade secrets.[110]

102 ibid.
103 *Bogucki* (n 2) 25-27.
104 *Raposo* (n 77) 101-103.
105 *Veale* (n 4) 110-112; Final Draft (n 62) Articles 72 (1), 26 (5).
106 *Raposo* (n 77) 101-103.
107 AI Liability Directive (n 88).
108 *Veale* (n 4) 110-112.
109 ibid.
110 ibid.

Moreover, as long as clear complaint rights remain haphazard, bottom-up enforcement on the basis of the database shall be severely hampered.[111]

According to Article 43, AI providers of high-risk AI applications have an obligation to carry out conformity assessments. Article 40 stipulates that there is a presumption of conformity with the provisions of Section 2 of Chapter III when a high-risk AI system has been developed in conformity with harmonised standards. As per Article 43, where the provider has applied said standards or common specifications referred to in Article 41, they can also opt for the internal control conformity assessment procedure of Annex VI. Standardisation organisations and notified bodies monitoring conformity are therefore key actors in the assessment of conformity and subsequently in the enforcement and oversight of the regulation.[112]

CEN and CENELEC are two of the three European Standardization Organizations (the third one being ETSI) that can be mandated by the Commission to develop harmonised standards which can cover the entirety of a legal instrument or specific areas as per Regulation No 1025/2012.[113] However, delegating rule-making to private bodies like CEN and CENELEC can be problematic, as it can limit stakeholder involvement and raise constitutional concerns regarding the legitimacy of the rule-making process.[114]

Further shortcomings of the Act consist of the current absence of a mechanism for public feedback on risk assessments.[115] There is no explicit establishment of a public monitoring mechanism to gather and disseminate independent information on the adverse societal effects of AI.[116] MSAs are meant to monitor compliance, yet they are not explicitly tasked with investigating or gathering data on the long-term social implications of AI implementation, a gap that does not seem to be sufficiently addressed by the EU central database.[117]

Furthermore, procedural rights with a societal dimension, such as access to justice, access to information, and participation in public decision-making, are still effectively lacking in the proposal.[118] The negotiations did introduce the possibility to lodge a complaint with the MSA, but the complainant still needs to prove an individual harm caused by AI, disregarding the need for protection against AI's societal impacts.[119]

Prominent amongst rights with a societal dimension as described by *Nathalie A. Smuha*,[120] the procedural right to a reasoned decision as expressed in Article 41 of the Charter and the right of access to justice through individual complaints have been identified

111 ibid.
112 ibid, 104-106.
113 ibid.
114 ibid.
115 *Smuha* (n 95) 20-23.
116 ibid.
117 ibid.
118 ibid.
119 ibid.
120 ibid, 15-19.

by *Melanie Fink* as not adequately guaranteed in the Act's original version.[121] Article 68b of the Parliament's mandate was likely to have satisfied a part of *Fink*'s concerns, however, it did not make it into the final text. As for the provisions that did, Article 85 does not provide for a societal dimension of redress nor does Article 86 grant a right to an explanation with regards to societal harm caused by AI.

4. The questionable future proof capacity of the AI Act

According to Recital 1 and Article 1 of the final draft, the purpose of the AI Act is to improve the functioning of the internal market and promote human centric and trustworthy AI ensuring protection of health, safety, fundamental rights, democracy and rule of law, and the environment, while supporting innovation. The tech industry is admittedly one of the fastest growing and changing fields of today's economy. With the AI revolution having already been launched at a staggering pace, it is of crucial importance for the AI Act to effectively remain relevant in such a dynamic field and foster AI innovation along with preserving fundamental rights.

The draft Act effectively employs only two instruments for the promotion of innovation, namely reliefs for SMEs and regulatory sandboxes.[122] Regulatory sandboxes have been accurately compared to clinical trials by *Vera Lúcia Raposo*, as frameworks established by regulators in view of allowing start-ups and innovators to experiment in a controlled environment with the potential to transparently test new technologies and improve law-making.[123]

These two instruments have been of key focus in the negotiations. Specifically, Recital 138 makes mention of 'responsible innovation', acknowledges the rapidly developing nature of AI and underlines the necessity of establishing AI regulatory sandboxes in order to ensure a future proof legal framework that promotes AI innovation. It also calls for adequate financial and human resources for sandboxes and even virtual or hybrid sandboxes. Recital 139 enumerates the objectives of these sandboxes, amongst which 'to facilitate regulatory learning for authorities and companies, including with a view to future adaptions of the legal framework'. It awards particular significance to the accessibility of sandboxes to SMEs, also stating that the Member States should strive to make them widely available throughout the Union.

These objectives are also transposed in Article 57 (5) while paragraph (16) lays down an obligation for national competent authorities to carry out an annual report so as to monitor the performance of the sandbox. As per Article 58, the Commission is tasked with ensuring that regulatory sandboxes are accessible to AI providers that meet the eligibility and selection criteria, whereas SMEs and start-ups gain free access to the regulatory

121 *Fink*, 'The EU Artificial Intelligence Act and Access to Justice', EU Law Live, 10 May 2021; *Fink* & *Finck*, 'Reasoned A(I)dministration: Explanation Requirements in EU Law and the Automation of Public Administration' (2022) 47(3) European Law Review, 376-392, retrieved from: https://hdl.handle.net/1887/3439725.

122 *Raposo* (n 77) 105-106.

123 ibid.

sandbox and aid, such as access to pre-deployment services and value-adding services. These points have been transposed to the final draft from Articles 53a (2) (a), 53a (2) (c) and 53a (3) of the Parliament's mandate. Further facilitations for SMEs were laid down in Article 28a of the Parliament's mandate stipulating unfair unilateral terms imposed on SMEs which are to be reviewed and kept up-to-date by the Commission as per paragraph 8.[124] Moreover, according to Recital 58a and Article 29a (4) of the same mandate, FRIAs would pose no requirement of prior consultation of relevant stakeholders and national supervisory authorities to SMEs.[125] However, these suggested provisions were not included in the final text.[126]

In spite of the amendments, these two means alone do not seem to be enough.[127] SMEs are already struggling to compete in the global tech market, while complying with the EU's necessary but nonetheless restricting regulations. Facilitating compliance with yet another necessarily restricting legislation trims the branches of the tree without addressing its roots, namely the limited resources of SMEs that already make it difficult to fully avail themselves of the opportunities provided by regulatory sandboxes. AI innovation cannot be fostered by supporting SMEs alone as the field of AI technology consists of a never-ending dialogue amongst the industry on the one hand, including large enterprises besides small ones and start-ups, and academics, government institutions, and NGOs on the other. This reality does not seem to be depicted under Chapter VI.

The initial criticism on the lack of an Article 89 GDPR equivalent for derogations in the public interest, research or statistical purposes,[128] was partially addressed through the abandoned Article 54a of the Parliament's mandate, providing for promotion of research in support of 'socially or environmentally beneficial outcomes'.[129] Research, except for testing in real world conditions, has been excluded from the scope of the Act as per Article 2 (6) of the final draft, yet inclusion of the Parliament's suggested provision to the final text could have presumably been welcomed by critics, even in spite of its vagueness, in view of encouraging innovation. Recital 142 of the final draft makes a similar mention, however without clarifying what qualifies as a 'socially or environmentally beneficial outcome', while a relevant definition eludes Article 3.

AI innovation in Europe can be hindered instead of fostered in view of this inadequacy.[130] This is further illustrated by reports on the possible compliance costs of the AI Act for enterprises, particularly SMEs. The reports that concerned compliance costs of the Act's original version can still be relevant in outlining the broader picture. Specifically, the Digital SME's reply to the AI Act consultation found the draft regulation particularly restrictive

124 Mandates (n 62).
125 ibid.
126 Final draft (n 62)
127 *Raposo* (n 77) 105-106.
128 ibid.
129 Mandates (n 62) Article 54a.
130 *Raposo* (n 77) 108-109.

towards SMEs in terms of compliance costs that cannot be passed on to customers, even identifying a potential to drive SMEs out of business entirely.[131]

Meeri Haataja and *Joanna J. Bryson* have carried out an analysis of the costs that are to be expected from the AI Act for affected organisations, in light of the Commission's Impact Assessment of the AI Act[132] and its supporting study[133] for an impact assessment of regulatory requirements for artificial intelligence in Europe.[134] As per said analysis, if an enterprise is required to adapt an existing Quality Management System (hereinafter 'QMS') in view of compliance with the AI Act, the cost for enterprises could be calculated between €117,750 - €174,800, and for SMEs with 50 employees, between €88,050 - €130,850.[135] Not having a QMS and opting for a third party assessment would raise these costs to €193,000 - €330,050 for enterprises and €144,350 - €247,150 for SMEs with 50 employees.[136]

Consultation of the reports themselves is advised for a more detailed analysis, yet these values are indicative of the fact that, in actuality, it is the smaller businesses that will sustain qualitatively the biggest hit, impairing innovative initiatives and consequently the Act's future proof capacity, when considered in tandem with the limited means to support innovation that are included in the Act.

The Act's risk to cause overregulation has been discussed in literature.[137] Its preemptive effect as a horizontal regulation forcing Member States to market compliant AI products that potentially conflict with national transparency rules, causes uncertainty with regards to the leeway allowed to apply higher transparency standards, as the full or partial harmonisation outreach of the forthcoming regulation is debated.[138] The avenues transversed by the Court of Justice of the European Union (hereinafter 'CJEU') in *Philip Morris*[139] to

131 *Digital SME*, 'Digital SME reply to the AI Act consultation' (6 August 2021), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665574_en, accessed 19 February 2024.

132 European Commission, 'Commission staff working document impact assessment. Accompanying the proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts' (22 April 2021), https://ec.europa.eu/transparency/documents-register/detail?ref=SWD(2021)84&lang=en, accessed 19 February 2024.

133 *Renda* et al., 'Study to support an impact assessment of regulatory requirements for Artificial Intelligence in Europe' (April 2020), https://op.europa.eu/en/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1, accessed 19 February 2024.

134 *Haataja* & *Bryson*, 'What Costs Should We Expect from the Eu's AI Act?' (27 August 2021), https://osf.io/preprints/socarxiv/8nzb4, accessed 19 February 2024.

135 ibid.

136 ibid.

137 *Ruschemeier*, 'AI as a Challenge for Legal Regulation – The Scope of Application of the Artificial Intelligence Act Proposal' (2023) 23(3) ERA Forum, 361-376, https://doi.org/10.1007/s12027-022-00725-6, accessed 19 February 2024.

138 *Veale* (n 4) 108-110.

139 Case C-547/14 *Phillip Morris* v *Secretary of State for Health* [2016] ECLI:EU:C:2016:325, para 70-72.

interpret Directive 2014/40/EU as partially harmonising certain areas, cannot seem to be used to interpret the AI Act in a similar, maximum harmonisation-escaping manner.[140] Regardless of its partial or full harmonisation effect and the subsequent preemption, EU internal market law would still allow affected parties to challenge more adverse limitations imposed by Member States as measures equivalent to a quantitative restriction on trade.[141] These would then have to be justified through the avenues of objective justifications or public interest requirements.[142] Nonetheless, the unclarity of the forthcoming regulation's preemptive effect is an impediment in securing its future relevance.

What has been eloquently described by *Jerome De Cooman* as 'the humpty dumpty fallacy' refers to the fact that the exhaustive enumeration of high-risk AI systems in the Act does not mean that the residual category is indeed non-high-risk.[143] It is a fundamental element of the AI Act's risk-based approach that becomes particularly conspicuous as a fault when taking into account that an exhaustive classification is not likely to cover dangers that are yet to emerge. *De Cooman*'s analysis, drawing inspiration from *Niels van Dijk*'s similar arguments on privacy by design,[144] is particularly critical of the voluntary endorsement mechanism that the Act provides for non-high-risk systems, as it might lead to bluewashing, a strategy similar to greenwashing, in which organisations utilise voluntary codes of conduct to falsely portray themselves as more compliant than what they really are.[145]

Finally, addressing the elephant in the room is in order. What the AI Act is still missing, being a fundamental rights oriented regulation as already established, is an effective answer to the question that has been on many people's mind since *OpenAI*'s applications entered the spotlight: 'is AI going to take my job?'. AI induced job displacements can be related to the Charter as they raise numerous fundamental rights concerns. As per Article 15 of the Charter, individuals have a right to engage in work and choose an occupation, which, in light of current developments, could be read to imply that the prognosed upcoming, unprecedented loss of jobs constitutes an infringement of said right if no fair compensation and opportunities for retraining are guaranteed. Article 31 of the Charter dictates that workers should enjoy fair and just working conditions, which could also be jeopardised if AI systems are designed oblivious to the EU social acquis.

140 *Veale* (n 4) 108-110.

141 ibid.

142 ibid.

143 *De Cooman*, 'Humpty Dumpty and High-Risk AI Systems: The Ratione Materiae Dimension of the Proposal for an EU Artificial Intelligence Act' (2022) 6 Mkt & Competition L Rev, 49, 50.

144 *van Dijk* et al., 'Right Engineering? The Redesign of Privacy and Personal Data Protection' (2018) 32(2-3) International Review of Law, Computers & Technology, 230, 251.

145 *De Cooman* (n 143) 50.

According to a *Goldman Sachs* report, an estimate of two thirds of jobs in Europe and the US and 300 million jobs in the world could potentially be automated by AI.[146] Of course, as history has proven with every other major technological breakthrough, the AI revolution is bound to create new jobs apart from rendering already existing ones obsolete. Specifically, AI is expected to create 97 million new jobs.[147] However, 85 million of them are likely to succumb to automation, eventually leaving for an actual gain of 13 million.[148] The ongoing trend when it comes to new AI jobs is that of prompt engineers, people that can skilfully communicate the right questions to generative AI in order to obtain desired outcomes.[149] This change of pace in the economy that has already been kickstarted towards a steady future course is left unaddressed by the AI Act at this point.

5. Ecce difficultates

Tailing the legislative journey of the AI Act has yielded several remaining pitfalls concerning scope, enforcement and future-proofness. At this point, it is possible to compile and categorise them into pitfalls for innovation and fundamental rights respectively.

As far as innovation is concerned, albeit a product safety regulation with the aim to promote innovation in AI, the Act only employs compliance reliefs for SMEs and regulatory sandboxes in pursuit of innovation, while the overall ecosystem that contributes to fostering AI innovation is not taken into account. SMEs are put in the spotlight of the Act's protective range, but receive no financial aid to make up for the uneven playing field that is created as a result of the uneven means at their disposal compared to larger enterprises and tech giants. Moreover, facilitating research has been taken into consideration in the amendments, but there remains to be no equivalent to GDPR's Article 89 for testing in real world conditions, while the abandoned Article 54a of the Parliament limited this consideration to undefined socially or environmentally beneficial outcomes, a vague phrasing that was maintained in Recital 142 of the final draft. All this takes place while the Act's unclarified preemptive effect keeps Member States from applying less stringent standards to foster innovation.

When it comes to fundamental rights, the definition of AI might be more inclusive, but the concept it defines is prone to change, leaving room for uncertainty and under-com-

146 *Hatzius* et al., 'The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)' (26 March 2023), https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html, accessed 19 February 2024.

147 *Zahidi*, 'The Jobs of Tomorrow' (December 2020), https://www.imf.org/en/Publications/fandd/issues/2020/12/WEF-future-of-jobs-report-2020-zahidi#:~:text=The%20robot%20revolution%20will%20create%2097%20million%20new%20jobs.&text=The%20emerging%20professions%20reflect%20the,cloud%20computing%2C%20and%20product%20development, accessed 19 February 2024.

148 *Confino* & *Burton,* 'A.I. might not replace you, but a person who uses A.I. could' (25 April 2023), https://fortune.com/2023/04/25/artificial-intelligence-ai-replace-humans-prompt-engineers-chatgpt/, accessed 19 February 2024.

149 ibid.

prehensiveness, thus undermining the Act's capacity to protect them. A possible loophole left in Article 6, requiring AI systems to pose a fundamental rights risk so as to qualify as high-risk, could be exploited to the detriment of fundamental rights protection. Private standardisation bodies maintain control over fundamental rights related issues albeit having limited expertise and democratically-obtained competence, while over-reliance on them raises legitimacy and democracy concerns. Furthermore, societal harm is not effectively included in the risks the Act protects against, contributing to a gap in effective fundamental rights protection. Procedural rights with a societal dimension are effectively lacking and interactive risks of AI are disregarded. There still is no right to claim damages from the immediate actors and no direct judicial scrutiny against the AI provider or deployer. Civil society organisations also do not have a right to directly call out infringements without representing a specific affected individual, while there is no mechanism to receive public feedback and the enforcement mechanism is still considered weak. MSAs are likely not as appropriate enforcement agents as DPAs, while bottom-up enforcement through the central database, which already raises concerns on trade secrets, is severely impaired as long as complainant rights are not sufficiently secured. Finally, the 'humpty dumpty fallacy' can lead to bluewashing, as explained, with businesses evading their responsibilities by striving to brand their systems as non-high-risk, whereas the imminent massive job displacements and the fundamental rights concerns they cause remain unaddressed.

It is these remaining difficulties that must be addressed for the endeavour of achieving a balance between the two concepts to succeed. Although the text of the Act is unlikely to change any time soon in the nearer future, the following section shall aspire to contribute useful points for contemplation regarding the ongoing discussion of regulating AI.

IV. AI Act Appraised: Balancing Innovation and Fundamental Rights

The research conducted for this contribution took place in parallel to tech competitors constantly launching new applications, on the one hand, and regulators all over the world striving to be the first to figure out how to keep them in check, on the other. This dance between competitive supremacy and regulatory oversight has already yielded interesting developments to consider when answering how the AI Act can protect fundamental rights without sacrificing innovation and overwhelming businesses.

The *Future of Life Institute* published an open letter calling for a pause of at least 6 months in developing AI systems more powerful than *GPT-4*.[150] With key actors in the AI stage such as *Elon Musk* and professor *Yoshua Bengio* amongst its signatories, the letter argued that further developments should not be left solely to tech giants but instead involve independent experts and policy-makers.[151] Poetically describing it as 'enjoying an AI summer instead of rushing into a fall', the letter did not call for a general pause on

150 *Future of Life*, 'Pause Giant AI Experiments: An Open Letter' (22 March 2023), https://futureoflife.org/open-letter/pause-giant-ai-experiments/, accessed 19 February 2024.

151 ibid.

AI development. Instead, it asked for shifting the focus on making the existing systems safer.[152]

Meanwhile, CEO of *OpenAI Sam Altman* initially found the EU's upcoming legislation restrictive to the extent that withdrawing his company from the EU altogether was considered an option.[153] This strategic statement was eventually withdrawn and replaced by a far more welcoming attitude towards the EU's legislative initiative, when regulators held their ground and deflected what was characterised by MEP *Kim van Sparrentak* as 'a blackmail by American Companies'.[154] *Mr. Altman* has since even expressed interest in establishing an *OpenAI* office in Europe.[155] Meanwhile, *OpenAI*'s participation in the EU's first sandbox in Spain has been discussed.[156]

These are developments in both the industry and the legislative field. On the one hand, tech innovators acknowledge the need for regulatory intervention and the fact that progress does not have to cease, just be refocused. On the other hand, the abrupt change of heart by one of the currently leading figures in AI at the face of the regulators' persistence could be interpreted as showing that innovators are willing to listen, if regulators are willing to stand firm.

The key take-away from looking at these recent relevant occurrences is that regulating AI can be steadfast without freezing innovation, just orienting it towards human centred objectives. This realisation means that regulators should aim for effective protection and that innovators are already willing to accept it. With that mind, and in view of the guidelines and recommendations to be expected on the application of the regulation, this section shall offer points to consider in tackling the previously identified weaknesses of the Act with regards to fostering innovation and protecting fundamental rights. By appropriately tilting the scales of both sides (1. and 2.), the key to reaching the balancing point shall emerge towards the end (3.).

1. Tilting the Scales: Innovation

We have identified that the Act employs rather few measures to foster innovation while SMEs lack the necessary funding to comply with it and the overall AI ecosystem is not sufficiently taken into account. In parallel, the extent of permitted research is not clarified and

152 ibid.

153 *Bastian*, 'OpenAI has 'no plans' to withdraw from EU, says Sam Altman' (27 May 2023), https://the-decoder.com/ai-overregulation-openai-ceo-sam-altman-sees-eu-exit-an-option/, accessed 19 February 2024.

154 ibid.

155 *Volpicalli*, 'ChatGPT boss wants HQ in Europe' (30 May 2023), https://www.politico.eu/article/open-ai-chatgpt-sam-altman-kicks-off-eu-charm-offensive-artifical-intelligence/?utm_source=POLITICO.EU&utm_campaign=62846722ea-EMAIL_CAMPAIGN_2023_05_25_07_46_COPY_02&utm_medium=email&utm_term=0_10959edeb5-2a25e84426-%5BLIST_EMAIL_ID%5D, accessed 19 February 2024.

156 ibid.

Member States are not allowed to impose more lenient standards. In order to contemplate suggestions to tackle these pitfalls, we need to look at them separately.

Quantity is not necessarily an indicator of quality. Before exploring additional avenues to foster innovation, the existing ones should be maximised to express their full potential. The *Future of Life Institute* had recommended further empowering the AI Office and forming a single European AI portal to a pan-European sandbox so as to facilitate registration and avoid fracturing the common market through different regional sandboxes.[157] It was also suggested that sandboxes be enhanced to offer additional services including legal support, insurance and fiscal incentives for research.[158] Another suggestion was increasing the public sector's capacity for AI development in view of ensuring efficient public oversight and improving public services.[159]

In line with further improving sandboxes, the *Future Society Institute* suggested that sandboxes be designed as attractive to foreign entrepreneurs and researchers apart from Europeans.[160] In light of this, it was recommended that compliance be facilitated through administrative assistance and supportive benefits.[161] Testing and experimentation facilities were put at the forefront, with adequate staffing and funding to stimulate innovation and trustworthiness and an interconnected network of national labs and EU centres to avoid fragmentation of the internal market, as prominent suggestions.[162]

It would still remain however that, in view of the already explained compliance costs, SMEs would be in need of financial support besides bureaucratic facilitation of compliance. Although SMEs have in principle the possibility to apply for the Commission's more general funding and tenders opportunities on AI,[163] a more tailored, narrowed down approach seems to be preferrable. The existing funding mechanisms of the Commission could have been complemented by an additional provision in the AI Act granting promising SMEs the possibility to apply for funding directly with the AI Office. Said provision could have been included under Article 64 as an additional task of the AI Office. Specific conditions according to which the application would be evaluated, and the funding granted would be laid down in this provision. These could have taken into account at least the purpose of the

157 *Future of Life*, 'FLI position on the Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)' (4 August 2021), https://futureoflife.org/wp-content/uploads/2021/08/FLI-Position-Paper-on-the-EU-AI-Act.pdf?x72900, accessed 19 February 2024.

158 ibid.

159 ibid.

160 *The Future Society*, 'Proposal for a regulation - 'Artificial intelligence – ethical and legal requirements' Trust in Excellence & Excellence in Trust Contribution by The Future Society' (August 2021), https://thefuturesociety.org/wp-content/uploads/2021/12/The-Future-Society-response-to-AI-Act-consultation.pdf, accessed 19 February 2024.

161 ibid.

162 ibid.

163 European Commission, 'Funding opportunities for small businesses', https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/key-enabling-technologies/artificial-intelligence-ai_en, accessed 19 February 2024.

systems developed, their competitiveness in the global stage, their potential impact to the economy of the internal market, and their alignment to the objectives of the AI Act.

Securing the seeds are planted is not enough on its own. The ground must also be fertile and the ecosystem friendly for innovation to grow. This realisation was not adequately encapsulated in Article 57 of the original version of the Act laying down the structure of the European AI Board. This led the *Digital SME* to suggest inclusion of multiple stakeholders such as research organisations, academia and large companies with a significant participation secured for SMEs.[164] Article 67 (2) of the final draft on the membership of the advisory forum clearly takes into consideration the aforementioned suggestion, minus the explicit minimum 40 % participation of SMEs that was recommended by the *Digital SME*. 'A balanced selection of stakeholders' is now required for the composition of the AI advisory forum of the EU AI Office. Although this is a welcomed change, diverse participation alone is not deemed enough to effectively cultivate an innovation-friendly ecosystem. Productive interaction allowing for fruitful outcomes is equally necessary to make participation in the advisory forum meaningful. It would therefore have been prudent that the Act stipulated that the advisory forum should, additionally to its existing tasks, hold more than a minimum of two regular meetings per year between the participating and other affected stakeholders. The purpose of these should be to keep all parties up-to-date with the state of the art in the field and to decide on non-binding, commonly accepted resolutions with regards to the preferred direction for AI innovation under equal voting rights.

Recital 142 promotes AI research and development in support of socially and environmentally beneficial outcomes. What constitutes such an outcome remains undefined in the Act. Suggesting the inclusion of a particular definition of a 'socially and environmentally beneficial outcome' in the Act itself would shift the focus of this contribution towards discussing delicate and broad philosophical concepts. This is perhaps the same reason why the legislature has avoided imprinting such a definition in the forthcoming legislation. Judicial interpretation of this provision considering the corresponding general principle of Article 4a (1) (f) of the Parliament's mandate could shed light to this unclarity. Yet the counterproductive uncertainty it provokes until then, could be tackled with an indicative enumeration of certain socially and environmentally beneficial outcomes in guidelines. A similar approach could be used to clarify the Act's preemptive effect.

2. Tilting the Scales: Fundamental Rights

We have also identified weaknesses of the Act with regards to ensuring full fundamental rights protection. Its definition and scope remain under-comprehensive, European Standardisation Organisations (hereinafter 'ESOs') are overpowered and civil society organisations are sidelined. Its enforcement mechanism still lacks an effective societal dimension and its future proof capacity seems inadequate to counter the 'humpty dumpty fallacy' and the upcoming job displacements. Once again, we need to take a step back and gradually assess the situation.

164 *Digital SME* (n 131).

The new definition of AI has shifted the focus mainly to covering general purpose AI models and generative AI in light of recent developments in the field. Inspiration for addressing the uncertainty caused by the circumstantial shifts the legislative process has exhibited can be drawn by looking overseas, specifically at the US Algorithmic Accountability Act (hereinafter the 'US AAA'). The US AAA centres around 'Automated Decision Systems' (hereinafter 'ADS') instead of going for the more futuristic term 'AI systems', managing nonetheless to achieve a more technology neutral and by extension future-proof outcome.[165] According to *Jakob Mökander*, the term 'ADS' encapsulates the relevant technical features of AI such as 'machine learning and hard-coded argumentation frameworks' without shifting attention to the pitfall of discussing the nature of intelligence *per se*.[166] The approach of the US AAA could be emulated, in guidelines to come, to make the scope of the AI Act more comprehensive, extending transparency obligations to include 'any decision that has significant legal or material effects on a consumer's life'.[167] This would additionally pave the road to better clarify the actors participating in the AI lifecycle and align the terminology of the AI Act with more internationally familiar concepts.[168] It would have also been welcomed if the negotiations did away with the requirement of posing a significant risk to fundamental rights for AI systems to be considered high-risk, so as to avoid loopholes causing limitations in their protection.

Turning our overseas inspection specifically to regulating AI, the US has yet not opted for binding legislation such as the one the EU has put forth. Instead, the National Institute of Standards and Technology (hereinafter 'NIST') has produced a voluntary set of recommendations called the AI risk management framework.[169] The interplay of this framework with the AI Act has been under discussion in the Trade and Technology Council (hereinafter 'TTC'), a regular meeting of top EU and US officials.[170] In view of the rapid advance of generative AI systems after the release of *ChatGPT*, European Commission executive vice president *Margrethe Vestager* called for 'an initiative to get as many other countries on board on an AI code of conduct for businesses voluntarily to sign up' during the latest TTC meeting.[171] *Vestager*'s initiative is a result of the realisation that even in the most optimistic scenario the AI Act will take considerable time to have full legal effect, which could be too long in the face of what has begun to be called 'a risk of extinction from

165 *Mökander* et al., 'The US Algorithmic Accountability Act of 2022 vs. the EU Artificial Intelligence Act: What Can They Learn from Each Other?' (2022) 32 Minds and Machines, 751, 752-753.

166 ibid.

167 ibid.

168 *Bogucki* (n 2) 19-21.

169 *Matthews*, 'US unveils light-touch strategy to deal with artificial intelligence risks' (31 January 2023), https://sciencebusiness.net/news/AI/us-unveils-light-touch-strategy-deal-artificial-intelligence-risks?utm_source=substack&utm_medium=email, accessed 19 February 2024.

170 ibid.

171 *Zubaşcu*, 'EU and US hatch transatlantic plan to rein in ChatGPT' (1 June 2023), https://scienceb usiness.net/news/AI/eu-and-us-hatch-transatlantic-plan-rein-chatgpt?utm_source=substack&utm_medium=email, accessed 19 February 2024.

AI'.[172] Both the US approach and the EU reaction underline a need to further emphasise voluntary codes of conduct.

Expanding the scope of the AI Act and strengthening voluntary codes of conduct does not have to come at the expense of innovation. A balance could be struck through a broader interpretation of the definition in tandem with the tailored guidance of the AI Office. This could be influenced by the US AAA's more comprehensive approach with an additional subsequent collaboration between sectoral EU regulators, experts, and stakeholders, spearheaded by the AI Office, to adequately define specific high-risk AI techniques in particular contexts.[173] This avenue combines the merits of a broad AI definition and tailored guidance for each high-risk category.[174] It additionally establishes a network of independent experts and sectoral regulators that would be ideal to offer insight concerning high-risk requirements, breaking the monopoly of private standardisation bodies such as CEN and CENELEC.[175]

Standardisation is of particular relevance when it comes to the Act's enforcement. Besides the merits of standardisation, such as the rapid transfer of technologies and the interoperability of systems, over-reliance on private standardisation bodies breeds prominent dangers as a result of the lack of democratic oversight.[176] In order for the European standardisation process to adequately reflect European values and fundamental rights, the *Robotics and AI Law Society* (hereinafter 'RAILS') called for legally binding provisions in the AI Act for essential requirements of high-risk AI systems which could in turn be further specified by ESOs.[177] The standardisation process could also become more inclusive through democratic input from society in general.[178] Existing standardisation bodies also lack the necessary expertise and democratic legitimacy to correctly interpret human rights law and relevant policy objectives.[179] *Christine Galvagna* has proposed increasing the participation of civil society organisations in the standardisation process to counterbalance this gap.[180] Possible options spanned from expanding Annex III and offering more funding

172 ibid; *Center for AI safety*, 'Statement on AI Risk', https://www.safe.ai/statement-on-ai-risk#open-letter, accessed 19 February 2024.

173 *Renda* & *Engler*, 'What's in a name? Getting the definition of Artificial Intelligence right in the EU's AI Act' (22 February 2023), https://www.ceps.eu/ceps-publications/whats-in-a-name/, accessed 19 February 2024, 3-5.

174 ibid.

175 ibid.

176 *Ebers* et al., 'The European Commission's Proposal for an Artificial Intelligence Act - A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)' (2021) 4(4) J 589-603, https://www.mdpi.com/2571-8800/4/4/43, accessed 19 February 2024 .

177 ibid.

178 ibid.

179 *Galvangna*, 'Discussion paper: Inclusive AI governance Civil society participation in standards development' (30 March 2023), https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/?utm_source=substack&utm_medium=email, accessed 19 February 2024.

180 ibid.

opportunities, to the establishment of a central hub with the purpose of facilitating civil society participation.[181]

The societal dimension of the Act's enforcement mechanism also needs to be enhanced. Civil society organisations can also strengthen procedural societal rights if granted the possibility to lodge complaints with competent authorities.[182] It is therefore recommended to invest in civil society organisations the right to directly report violations even outside the mandate of an affected individual.[183] Articles 85 to 87 show promising grounds, yet a right to societal participation in public decision-making on AI projects and a societal 'access to justice' right against non-compliant public decisions regardless of individual or collective harm could have been added.[184] Moreover, the suggestion of entrusting the enforcement of the regulation to DPAs as national competent authorities, given their familiarity with AI technologies and fundamental rights risk assessments, should have been upheld in the negotiations.[185]

With a sturdier basis for complainant rights secured, it would be possible to take another go at bottom-up enforcement through the EU central database. Specifically, in view of making the most of Article 71's potential, individuals should have the possibility to submit feedback on AI systems registered in the database through a review window. This would significantly address the need for public feedback while strengthening transparency and providing developers with direct, valuable consumer feedback on the adequacy of their products. A further welcome addition apart from the AI Liability Directive would be a clarification ensuring that affected individuals can still pursue claims for damages under national law in case of a breach.[186]

Finally, what seems to be the most difficult challenge of all, the 'fit for the future test' must tackle both the 'humpty dumpty fallacy' and the imminent massive job displacements to pass the fundamental rights protection check. As far as the former is concerned, *Jerome De Cooman* has reflected on HLEG's Ethics Guidelines and proposed human oversight and agency as solutions.[187] Human agency stands for the possibility of operators to 'make informed autonomous decisions', while human oversight consists of three dimensions: (a) human-in-the-loop, meaning human intervention, (b) human-on-the-loop, standing for human supervision and (c) human-in-command, meaning human command of the process.[188] Using recommender systems as a point of reference, *De Cooman* further argues that sys-

181 ibid.

182 *Smuha* (n 95) 16-20.

183 *European Center for Not-for-Profit Law* (n 84).

184 *Smuha* (n 95) 16-20.

185 *Access Now*, 'Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act' (August 2021), https://www.accessnow.org/wp-content/uploads/2021/08/Submission-to-the-European-Commissions-Consultation-on-the-Artificial-Intelligence-Act.pdf, accessed 19 February 2024, 27-30.

186 *Ebers* et al. (n 176).

187 *De Cooman* (n 143) 74-77.

188 ibid.

tems should be designed to allow operators to have a full understanding of their capacities and limitations, so as to mitigate possible automation bias.[189]

However, human monitoring of AI is not always effective.[190] *Johannes Walter* and his associates have identified that humans are likely to rely on inaccurate algorithmic outputs, as a result of multiple psychological factors spanning from absolution of responsibility to varying degrees of trust towards the system's capability to carry out a specific task.[191] The suggested recommendations to counter this include acknowledging that human oversight is fallible, conducting mandatory assessments of the capacity of human oversight to prevent harm from high-risk AI systems and, if oversight is found inadequate, abandoning the use of the system altogether or modifying it.[192] In addition to considering these recommendations, the prospects of general principles in tackling the 'humpty dumpty fallacy' should be further explored so as to avoid missing the tree before the forest by putting 'too much emphasis on specificities rather than generalities in the decision-making process', as *De Cooman* has noted.[193]

As for mitigating the impact of AI on the fundamental rights of workers, a by design approach of aligning AI used in the workplace, such as management software, with EU labour laws related to workers' fundamental rights, such as the right to decent working conditions and working time restrictions of Article 31 of the Charter, seems to be in order.[194] Workers' first line of defence against the AI revolution should be making sure that the existing social acquis that is embedded in EU labour laws, such as non-discrimination and occupational safety and health laws, is effectively imprinted in the development of AI systems that impact the employment relationship.[195] Future guidelines could seize the opportunity to stipulate a clear, worker friendly by design approach for high-risk AI-systems used in the workplace.

Yet when it comes to the last line of defence, the Act should have been equipped with a mechanism to cater for the landslide of AI induced displacements. Instead of particular sectoral provisions which would contradict the horizontal nature of the regulation, the focus could instead be shifted once again to general principles. Article 4a (1) of the Parliament's mandate introduced a rich palette of principles with paragraph (f) standing out as a possible tool in intercepting the upcoming changes in the labour market. A more liberal reading of the proposed principle of 'social and environmental well-being' could lead to understanding that harnessing AI should not entail mass unemployment. Yet, were the Act to limit itself

189 ibid.

190 *Walter*, 'The AI Act should use humans to monitor AI only when effective' (15 February 2023), https://www.euractiv.com/section/digital/opinion/the-ai-act-should-use-humans-to-monitor-ai-only-when-effective/?utm_source=substack&utm_medium=email, accessed 19 February 2024.

191 ibid.

192 ibid.

193 *De Cooman* (n 143) 74-77.

194 *Cefaliello & Kullmann*, 'Offering False Security: How the Draft Artificial Intelligence Act Undermines Fundamental Workers Rights' (2022) 13(4) European Labour Law Journal, 542, 548-550.

195 ibid, 552-550.

to a horizontal application of this principle, under the suggested interpretation, it would forfeit a great deal of the perks of automation, even when it comes to mundane tasks. The level of automation that should be allowed under this principle should take into account the livelihood and fulfilment of workers without neutering the prospects for innovation. Striking this balance becomes less difficult when viewing workers not solely as recipients of rights, but also as driving factors of the economy. In order to equitably avoid massive layoffs, guidelines could, on the one hand, rely on general principles to prevent them, and on the other, encourage employers to invest in retraining their human resources to better fit the new digital landscape.

3. Balancing the Scales: In principia veritas

We have gone through numerous suggestions to improve the AI Act's capacity to protect fundamental rights without jeopardising innovation. Enhancing regulatory sandboxes and making them appealing in the international stage while financially supporting SMEs to prevent brain drain are measures that could support innovation. Additional innovation-oriented suggestions include fostering an innovation-friendly overall ecosystem through interactive dialogue between stakeholders and laying down an indicative enumeration of socially and environmentally beneficial outcomes so as to provide clarity and delineate the Act's preemptive effect.

In the realm of fundamental rights protection, inspiration can be drawn from other regulatory initiatives around the world, such as the US AAA, to tilt its scope towards a more technology neutral and future-proof direction. The adoption of voluntary codes of conduct should be incentivised in order to make up for the timelapse till the AI Act becomes fully effective. Both these initiatives can be carried out in tandem with enhanced AI Office guidance for a more detailed approach for each high-risk category so as to not compromise innovation. The powers of private standardisation bodies need to be kept in check while encouraging the participation of civil society organisations to the standardisation process. The Act's enforcement mechanism can be further improved by granting civil society organisations the right to directly report violations and by upgrading the role of DPAs in enforcement. Furthermore, the EU Database can be used as a tool for the collection of public feedback to the benefit of both enforcers and developers. Effective coverage of possible harms can be achieved through the principles of human agency and oversight provided their capacity to reliably prevent harm is effectively assessed. As for addressing the upcoming changes in the labour market, AI systems should be designed to the benefit of workers as per the principles of social and environmental well-being and incentives towards reinvesting in human capital should be given.

The discussion yields a need for flexibility and temporal adaptability in view of balancing fundamental rights protection and innovation in the age of AI. AI general principles show the most promise in effectively achieving this endeavour. Indeed, given how dynamic the field of AI is, precise rights such as those envisaged in the GDPR seem more absent in the AI Act. The proposed Parliamentary amendments instead offered general principles in

Article 4a and directions through recitals, laying the groundwork for what could prove to become a playing field for future litigation with a lot of room for adjudicating which way the scales should shift in each particular occasion. Although Article 4a was not upheld in the end, its quasi-counterpart can be found in Recital 27, which refers to the AI HLEG's AI principles upon which Article 4a's general principles were based. The Recital mentions these as 'important to recall' and then goes on to enumerate them describing each as per the wording for the general principles of the proposed Article 4a. According to the Recital, they should be translated in the design and use of AI models and serve as the basis for voluntary codes of conduct, while stakeholders are encouraged to consider them for the development of voluntary best practises and standards.

These principles could, through future guidelines and jurisprudence, serve as a common reference point for stakeholders, setting the direction that AI innovation is meant to follow, namely building and operating systems in a manner that is controllable by humans, robust against misuse, privacy aware, transparent, inclusive, and socially and environmentally beneficial. By setting a level playing field and creating demand for a particular quality for AI that is defined in accordance with this fundamental rights oriented common point of reference, the prospect of AI general principles has the potential to provide a flexibility for striking an effective balance the Charter alone would not be able to.

It is at this point that we can recall the Research Question the paper aimed to answer: How can the AI Act promise sufficient fundamental rights protection without compromising AI innovation and overburdening enterprises?

In light of all that has been discussed and in answering the research question, AI general principles are the key to unlocking the AI Act's true potential to sufficiently protect fundamental rights while fostering innovation. A dynamic field requires an equally dynamic regulatory ecosystem which is based on a solid direction imprinted on horizontal black-letter legislation that grants litigants on either side of the scales a fair arsenal of arguments. This should be spearheaded by the Court of Justice of the EU's living, temporally relevant interpretation to find which way the scales should tilt in each particular case.

V. Conclusion

This paper has presented the historical and political background that led to the adoption of the AI Act in the first place. It has traversed the legislative journey of the Act identifying strengths and remaining weaknesses. It finally suggested solutions towards an equitable balance of innovation and fundamental rights protection.

Starting from the shores of ancient Crete, we embarked on a quest to answer how the AI Act can protect fundamental rights without compromising innovation and overburdening businesses. Having tracked the historical evolution of Artificial Intelligence from the legend of *Talos* to *ChatGPT*, we then continued to trail the key policy documents that acted as precursors to the proposal of the Act.

With the legislative process now having reached an approved final draft, we described the state of the art concerning the AI Act. The strong and wavering points of the amend-

ments, those proposed and the ones agreed, were explored and lingering weaknesses regarding the Act's scope, enforcement and future proofness were identified.

In view of the AI fever spreading to both legislators and innovators worldwide, we further explored suggestions for a balanced regulatory approach in the EU's AI Act concerning innovation and fundamental rights protection. After elaborating on specific suggestions to foster innovation and protect fundamental rights, the answer to the research question emerged in the form of AI general principles, which can constitute the basis for constructing an equitable ecosystem of AI regulation, fleshed out through soft law and the Court's case-law.

Studying the genesis of AI regulation in Europe is as exciting as will be experiencing it in practice soon. The AI revolution has indeed spawned wonders, yet in what seems to paradoxically be simultaneously its most evident success and the greatest threat it poses, it made those wonders ordinary. With Pandora's box now open to everyone and the field progressing at an alarming rate, it appears that the possibilities and limitations of AI regulation will be a key topic for academic contemplation in the years to come. This modest contribution hopes to guide future discussions towards studying the past, evaluating the present and planning the future in order to better understand and improve this novel field of law.

Bibliography

Book Chapters

Walters, R., & Novak, M. (2021). Artificial Intelligence and Law. In *Cyber Security, Artificial Intelligence, Data Protection & the Law*. Springer Singapore.

Journal Articles

Cefaliello, A., & Kullmann, M. (2022). Offering False Security: How the Draft Artificial Intelligence Act Undermines Fundamental Workers Rights. *European Labour Law Journal, 13*(4).

Cole, M. D. (2024). AI Regulation and Governance on a Global Scale: An Overview of International, Regional and National Instruments, *AIRe (Journal of AI Law and Regulation), 1*(1), 126-142.

De Cooman, J. (2022). Humpty Dumpty and High-Risk AI Systems: The Ratione Materiae Dimension of the Proposal for an EU Artificial Intelligence Act. *Mkt & Competition L Rev, 6*.

Fink, M. (2021). The EU Artificial Intelligence Act and Access to Justice. *EU Law Live*, 10 May.

Fink, M., & Finck, M. (2022). Reasoned A(I)dministration: Explanation Requirements in EU Law and the Automation of Public Administration. *European Law Review, 47*(3).

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review, 61*.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Business Horizons, 62*(1).

Kop, M. (2021). EU Artificial Intelligence Act: The European Approach to AI. *Issue Stanford-Vienna Transatlantic Technology Law Forum*, Transatlantic Antitrust and IPR Developments, Stanford University.

Mökander, J., et al. (2022). The US Algorithmic Accountability Act of 2022 vs. the EU Artificial Intelligence Act: What Can They Learn from Each Other? *Minds and Machines, 32*.

Raposo, V. L. (2022). Ex Machina: Preliminary Critical Assessment of the European Draft Act on Artificial Intelligence. *International Journal of Law and Information Technology, 30*(1).

Ruschemeier, H. (2023). AI as a Challenge for Legal Regulation – The Scope of Application of the Artificial Intelligence Act Proposal. *ERA Forum, 23*(3).

Smuha, N. (2021). Beyond the Individual: Governing AI's Societal Harm. *Internet Policy Review, 10*(3). https://doi.org/10.14763/2021.3.1574 accessed 19 February 2024.

Stix, C. (2021). The Ghost of AI Governance Past, Present and Future: AI Governance in the European Union. *Pre-print draft*, v.6.1., last updated June 18th

Van Dijk, N., et al. (2018). Right Engineering? The Redesign of Privacy and Personal Data Protection. *International Review of Law, Computers & Technology, 32*(2-3).

Veale, M., & Borgesius, F. Z. (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).

Primary Legislatio

Charter of Fundamental Rights of the European Union [2000] OJ C 364/1

Secondary Legislation

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

Regulation (EU) 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council OJ L 316/12

Directive (EU) 2020/1828 of the European Parliament and of the Council of 25 November 2020 on representative actions for the protection of the collective interests of consumers and repealing Directive 2009/22/EC [2020] OJ L 409

Directive 2014/40/EU of the European Parliament and of the Council of 3 April 2014 on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products and repealing Directive 2001/37/EC [2020] OJ L 127/1

European Union Documents

DRAFT Amendments by the European Parliament to the Commission Proposal, Amendment 808, Report - A9-0188/2023, 'Regulation (EU) 2024/... of the European Parliament and of the Council of ... 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)' (6 March 2024), as submitted to the European Parliament in view of the final vote. Available at https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_EN.html?redirect, accessed 20 March 2024;

DRAFT AI liability Directive - Brussels, 28.9.2022 COM(2022) 496 final 2022/0303 (COD) Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) {SEC(2022) 344 final} - {SWD(2022) 318 final} - {SWD(2022) 319 final} - {SWD(2022) 320 final}

DRAFT Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD))

DRAFT Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Analysis of the final compromise text with a view to agreement' (Interinstitutional File: 2021/0106(COD) OR. en 5662/24) 26.1.2024

DRAFT European Parliament and Council, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (2021/0106(COD), Final draft as updated on 21/01) 21.1.2024

DRAFT Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

European Commission, Directorate-General for Research and Innovation, European Group on Ethics in Science and New Technologies, Statement on artificial intelligence, robotics and 'autonomous' systems – Brussels, 9 March 2018, Publications Office, 2018, <https://data.europa.eu/doi/10.2777/531856 > accessed 19 February 2024

European Commission, Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions, 'Artificial Intelligence for Europe' COM/2018/237 final

European Commission, Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions, 'Coordinated Plan on Artificial Intelligence' COM/2018/795 final

European Commission, Communication to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, 'Building Trust in Human-Centric Artificial Intelligence' COM/2019/168 final

European Commission, White Paper on Artificial Intelligence - A European Approach to Excellence and Trust, COM(2020) 65 final (19 February 2020)

European Commission, 'COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT Accompanying the proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts' (22 April 2021) <https://ec.europa.eu/transparency/documents-register/detail?ref=SWD(2021)84&lang=en > accessed 19 February 2024.

European Economic and Social Committee, Opinion on 'Artificial Intelligence - The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society' (Own-initiative Opinion) [2017] OJ C 288, 31.8.2017

Speeches

Blumenthal, R., et al. (16 May 2023). OpenAI CEO Sam Altman testifies at Senate artificial intelligence hearing | full video'. *Youtube.com*. Retrieved from: https://www.youtube.com/watch?v=TO0J2Yw7usM

Case-law of the European Court of Justice

Case C-547/14 *Phillip Morris* v *Secretary of State for Health* [2016] ECLI:EU:C:2016:325

Newspaper Articles and Press Releases

Scott, M. (Brussels, 20 April 2023). Digital Bridge: Global AI rulebook – US digital policymaking – Data rules. *Politico.eu*. Retrieved from: https://www.politico.eu/newsletter/digital-bridge/global-ai-rulebook-us-digital-policymaking-data-rules/

Volpicalli, G. (30 May 2023). ChatGPT boss wants HQ in Europe. *Politico.eu*. Retrieved from: https://www.politico.eu/article/open-ai-chatgpt-sam-altman-kicks-off-eu-charm-offensive-artifical-intelligence/?utm_source=POLITICO.EU&utm_campaign=62846722ea-EMAIL_CAMPAIGN_2023_05_25_07_46_COPY_02&utm_medium=email&utm_term=0_10959edeb5-2a25e84426-%5BLIST_EMAIL_ID%5D

Websites

Access Now (August 2021). Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act. Retrieved from: https://www.accessnow.org/wp-content/uploads/2021/08/Submission-to-the-European-Commissions-Consultation-on-the-Artificial-Intelligence-Act.pdf.

Access Now (11 May 2023). Big wins, but gaps remain: European Parliament Committees vote to secure key rights protections in AI Act. Retrieved from: https://www.accessnow.org/press-release/european-parliament-committees-vote-ai-act/.

Bastian, M. (27 May 2023). OpenAI has "no plans" to withdraw from EU, says Sam Altman. *The Decoder*. Retrieved from: https://the-decoder.com/ai-overregulation-openai-ceo-sam-altman-sees-eu-exit-an-option/.

Bertuzzi, L. (15 November 2023). EU's AI Act negotiations hit the brakes over foundation models. Retrieved from: https://www.euractiv.com/section/artificial-intelligence/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/.

Bogucki, A., Engler, A., Perarnaud, C., & Renda, A. (14 September 2022). The AI Act and the Emerging EU Digital Acquis. *CEPS*. Retrieved from: https://www.ceps.eu/ceps-publications/the-ai-act-and-emerging-eu-digital-acquis/.

Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. Center for Research on Foundation Models (CRFM), Stanford Institute for Human-Centered Artificial Intelligence (HAI), Stanford University. Retrieved from: https://arxiv.org/abs/2108.07258.

Center for AI Safety. Statement on AI Risk. Retrieved from: https://www.safe.ai/statement-on-ai-risk#open-letter.

Confino, P., & Burton, A. (25 April 2023). A.I. might not replace you, but a person who uses A.I. could. *Fortune*. Retrieved from: https://fortune.com/2023/04/25/artificial-intelligence-ai-replace-humans-prompt-engineers-chatgpt/.

Council of the EU (9 December 2023). Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world. Retrieved from: https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/.

Digital SME (6 August 2021). DIGITAL SME reply to the AI Act consultation. Retrieved from: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665574_en.

Ebers, M., et al. (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(4), 589-603. Retrieved from: https://www.mdpi.com/2571-8800/4/4/43.

European Center for Not-for-Profit Law (11 May 2023). Hope on the horizon for digital civic space, as EU parliament advances protection of rights. Retrieved from: https://ecnl.org/news/hope-horizon-digital-civic-space-eu-parliament-advances-protection-rights.

European Commission (9 April 2018). Digital Day 2018: EU countries to commit to doing more together on the digital front. Retrieved from: https://ec.europa.eu/commission/presscorner/detail/en/IP_18_2902.

European Commission (8 April 2019). Ethics guidelines for trustworthy AI - High-Level Expert Group on AI. Retrieved from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

European Commission (17 July 2020). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Retrieved from: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

European Commission. Funding opportunities for small businesses. Retrieved from: https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/key-enabling-technologies/artificial-intelligence-ai_en.

European Parliament (27 January 2017). Report - A8-0005/2017. Retrieved from: https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html.

European Parliament (30 January 2019). Report - A8-0019/2019. Retrieved from: https://www.europarl.europa.eu/doceo/document/A-8-2019-0019_EN.html.

European Parliament (11 May 2023). AI Act: a step closer to the first rules on Artificial Intelligence. Retrieved from: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence.

Fox News (11 May 2023). Here is how Europe is pushing to regulate artificial intelligence as ChatGPT rapidly emerges. Retrieved from: https://www.foxnews.com/world/europe-pushing-regulate-artificial-intelligence-chatgpt-rapidly-emerges.

Frank, C. (12 May 2023). Draft of the AI Act gets on the home stretch. Retrieved from: https://www.taylorwessing.com/en/insights-and-events/insights/2023/05/draft-of-the-ai-act.

Future of Life Institute (22 March 2023). Pause Giant AI Experiments: An Open Letter. Retrieved from: https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Future of Life Institute (4 August 2021). FLI position on the Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Retrieved from: https://futureoflife.org/wp-content/uploads/2021/08/FLI-Position-Paper-on-the-EU-AI-Act.pdf?x72900.

Galvangna, C. (30 March 2023). Discussion paper: Inclusive AI governance Civil society participation in standards development. Ada Lovelace Institute. Retrieved from: https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/?utm_source=substack&utm_medium=email.

Hatzius, J., et al. (26 March 2023). The Potentially Large Effects of Artificial Intelligence on Economic Growth. Goldman Sachs Research. Retrieved from: https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html.

Keller, P. (14 December 2023). A Frankenstein-Like Approach: Open Source in the AI Act. Retrieved from: https://openfuture.eu/blog/a-frankenstein-like-approach-open-source-in-the-ai-act/.

Matthews, D. (31 January 2023). US unveils light-touch strategy to deal with artificial intelligence risks. Science|Business. Retrieved from: https://sciencebusiness.net/news/AI/us-unveils-light-touch-strategy-deal-artificial-intelligence-risks?utm_source=substack&utm_medium=email.

Meeri, H. & Bryson, J. (14 March 2022). Reflections on the EU's AI Act and How We Could Make It Even Better. *Competition Policy International*. Retrieved from: https://www.competitionpolicyinternational.com/reflections-on-the-eus-ai-act-and-how-we-could-make-it-even-better/.

Meeri, H. & Bryson, J. (27 August 2021). What Costs Should We Expect from the Eu's AI Act? Retrieved from: https://osf.io/preprints/socarxiv/8nzb4.

News European Parliament (9 December 2023). Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. Retrieved from: https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

O'Shaughnessy, M., & Sheehan, M. (14 February 2023). Lessons From the World's Two Experiments in AI Governance. Carnegie Endowment for International Peace. Retrieved from: https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035?utm_source=substack&utm_medium=email.

Pinsent Masons (16 May 2023). MEPs' EU AI Act proposals focus on "foundation models". Retrieved from: https://www.pinsentmasons.com/out-law/news/meps-eu-ai-act-foundation-models.

Renda, A., & Engler, A. (22 February 2023). What's in a name? Getting the definition of Artificial Intelligence right in the EU's AI Act. CEPS. Retrieved from: https://www.ceps.eu/ceps-publications/whats-in-a-name/.

Renda, A., et al. (April 2020). Study to support an impact assessment of regulatory requirements for Artificial Intelligence in Europe. Publications Office of the European Union. Retrieved from: https://op.europa.eu/en/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1.

Rhawi, C. (11 May 2023). AI Act: EU paves the way for world's first comprehensive rulebook on AI. Renew Europe Group. Retrieved from: https://www.reneweuropegroup.eu/news/2023-05-11/ai-act-eu-paves-the-way-for-worlds-first-comprehensive-rulebook-on-ai.

Shashkevich, A. (28 February 2019). Ancient myths reveal early fantasies about artificial life. Stanford News. Retrieved from: https://news.stanford.edu/2019/02/28/ancient-myths-reveal-early-fantasies-artificial-life/.

The Future Society (August 2021). Proposal for a regulation - "Artificial intelligence – ethical and legal requirements" Trust in Excellence & Excellence in Trust Contribution by The Future Society. Retrieved from: https://thefuturesociety.org/wp-content/uploads/2021/12/The-Future-Society-response-to-AI-Act-consultation.pdf.

Walter, J. (15 February 2023). The AI Act should use humans to monitor AI only when effective. Euractiv. Retrieved from: https://www.euractiv.com/section/digital/opinion/the-ai-act-should-use-humans-to-monitor-ai-only-when-effective/?utm_source=substack&utm_medium=email.

Zahidi, S. (December 2020). The Jobs of Tomorrow. International Monetary Fund (IMF). Retrieved from: https://www.imf.org/en/Publications/fandd/issues/2020/12/WEF-future-of-jobs-report-2020-zahidi.

Zubaşcu, F. (1 June 2023). EU and US hatch transatlantic plan to rein in ChatGPT. Science|Business. Retrieved from: https://sciencebusiness.net/news/AI/eu-and-us-hatch-transatlantic-plan-rein-chatgpt?utm_source=substack&utm_medium=email.