*European Journal of Risk Regulation* (2021), 1–29
doi:10.1017/err.2021.52

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

# A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications

Jon Truby*, Rafael Dean Brown, Imad Antoine Ibrahim and Oriol Caudevilla Parellada

Center for Law & Development, College of Law, Qatar University, Doha, Qatar
*Corresponding author. Email: jon.truby@qu.edu.qa

## Abstract

This paper argues for a sandbox approach to regulating artificial intelligence (AI) to complement a strict liability regime. The authors argue that sandbox regulation is an appropriate complement to a strict liability approach, given the need to maintain a balance between a regulatory approach that aims to protect people and society on the one hand and to foster innovation due to the constant and rapid developments in the AI field on the other. The authors analyse the benefits of sandbox regulation when used as a supplement to a strict liability regime, which by itself creates a chilling effect on AI innovation, especially for small and medium-sized enterprises. The authors propose a regulatory safe space in the AI sector through sandbox regulation, an idea already embraced by European Union regulators and where AI products and services can be tested within safeguards.

**Keywords:** artificial intelligence; sandbox regulation; strict liability; fault-based liability; EU regulation

## I. Introduction

While governments have been slow to respond to the increasingly urgent demand[1] to govern artificial intelligence (AI),[2] recent legislative activity signals a growing effort to mitigate fears with a myriad of regulations intended to rein in the potential for risks and uncertainties posed by AI. The USA has a pending bill named the Algorithmic Accountability Act of 2019, which would require companies to assess and "reasonably address" risks posed by automatic decision systems that are related

[1] Elon Musk gave a dramatic and widely publicised apocalyptic warning of AI's existential risk to humans. C Clifford, "Elon Musk: 'Mark my words – A.I. is far more dangerous than nukes'" (*CNBC*, 13 May 2018) <https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html> (last accessed 13 December 2020).

[2] Y Chae, "U.S. AI regulation guide: legislative overview and practical considerations" (2020) 3 The Journal of Robotics, Artificial Intelligence & Law 17.

to "privacy and security of personal information" and that lead to "inaccurate, unfair, biased, or discriminatory decisions".[3] The European Union (EU), however, has arguably been leading the march towards regulating AI, with notable examples such as the European Parliament's (EP) Resolution on Civil Law Rules on Robotics in 2017 and the EU's Ethics Guidelines for Trustworthy AI in April 2019.[4] The EU's proactive, innovative and visionary strategy has encouraged the EU to be "a front-runner in AI development".[5] [6] Regardless, governments continue to grapple with how to best regulate AI, with most broadly waiting for guidance from the EU and the USA.

The EU's approach to AI regulation, however, has been *ad hoc*, relying on piecemeal legislation from Member States and proposals or Resolutions from EP committees.[7] The EU is aware of the need for harmonisation among Member States, especially as the European Commission (EC) acknowledged the challenges posed to the EU's and national liability frameworks that may impact their effectiveness because of emerging technologies such as AI.[8] The EU initiatives and policies have been subject to great scrutiny in recent years given the central role that the EU is playing in the regulation of AI.[9] Its ambitious agenda[10] includes suggestions for how AI and AI regulation may affect other sectors such as health, manufacturing and mobility.[11] This debate is taking place in the context of a growing discussion concerning AI and high-risk activities that require regulation.[12] Currently, the discussion is revolving

[3] M MacCarthy, "An examination of the Algorithmic Accountability Act of 2019" (2019) Georgetown University, Transatlantic Working Group <http://dx.doi.org/10.2139/ssrn.3615731> (last accessed 13 December 2020).

[4] Chae, supra, note 2, 17.

[5] G Carriço, "The EU and artificial intelligence: a human-centred perspective" (2018) 17 European View 29, 33.

[6] C Stix, "A survey of the European Union's artificial intelligence ecosystem" (2019) Leverhulme Centre for the Future of Intelligence, University of Cambridge <http://lcfi.ac.uk/resources/survey-european-unions-artificial-intelligence-eco/3615731> (last accessed 11 January 2020).

[7] A Bertolini, "Study on artificial intelligence and civil liability" (2020) European Parliament's Committee on Legal Affairs, Policy Department for Citizens' Rights and Constitutional Affairs, Study Requested by the JURI Committee <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf> (last accessed 13 December 2020; noting that Member States have not had a unified regulation of advanced technology but have emerging piecemeal interventions).

[8] Commission, "Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics" COM (2020) 64 final.

[9] M Hildebrandt, "The artificial intelligence of European Union law" (2020) 21 German Law Journal 74; J Niklas and L Dencik, "Working Paper: European Artificial Intelligence Policy: mapping the institutional landscape" (2020) Data Justice Lab.

[10] R Csernatoni, "An ambitious agenda or big words?: developing a European approach to AI" (*Egmont Institute*, 2019) <https://www.jstor.org/stable/resrep21397?seq=1#metadata_info_tab_contents> (last accessed 13 December 2020).

[11] Eit, Climate-KIC, "How should Europe approach AI in the strategic areas of climate, health, manufacturing and mobility?" (*Climate-KIC*, 30 September 2020), <https://www.climate-kic.org/in-detail/how-should-europe-approach-ai/> (last accessed 13 December 2020).

[12] See, for instance, VC Mülle, "Risks of general artificial intelligence" (2014) 26 Journal of Experimental & Theoretical Artificial Intelligence 297; MU Scherer, "Regulating artificial intelligence systems: risks, challenges, competences, and strategies" (2016) 29 Harvard Journal of Law & Technology 353; A Turchin and D Denkenberger, "Classification of global catastrophic risks connected with artificial intelligence" (2020) 35 AI & Society 147.

around the type of liability to be imposed considering various variables.[13] In particular, the main question is whether and how to impose strict liability on high-risk AI activities.[14] The EU clarified its legislation of high-risk AI activities, and additionally recognised the importance of including sandbox regulation for AI, in the recent EC 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 (EC Proposal).[15]

This paper argues for a robust sandbox approach to regulating high-risk AI applications as a necessary complement to strict liability regulation.[16] This topic, for the time being, is not well examined in the literature due to the recency of the EC Proposal and because very few studies have addressed the application of a regulatory sandbox to AI. Hence, this article seeks to fill this gap in the literature.[17] The article focuses on the period between sandbox regulation and placement on the market of high-risk AI applications. The EC Proposal takes a similar approach when it allows for the use of an AI regulatory sandbox "for a limited time before their placement in the market or putting into service".[18] The EC Proposal, however, imposes the same strict liability regime for both sandboxed and non-sandboxed high-risk AI applications.

The authors here argue that the use of a more unified and robust regulatory sandbox framework for AI, rather than the limited-duration and liability-exposed regulatory sandbox proposed by the EC, is more appropriate when balancing the liability risks of AI, the cost of regulatory compliance and the need to encourage AI innovation. With a complementary regulatory sandbox, the question of AI liability should only arise after a high-risk AI application enters the market, assuming successful testing in the sandbox environment. In other words, a regulatory sandbox should be a safe space for both discovery and application, or for both innovation and regulation.

Furthermore, EU regulators should not leave sandbox regulation implementation to each Member State, as doing so would encourage sandbox shopping and the exploitation of lax sandbox environments. AI sandbox regulation, like AI regulation in

[13] See, for instance, N Osmani, "The complexity of criminal liability of AI systems" (2020) 14 Masaryk University Journal of Law and Technology 53; P Cerka, J Grigiene and G Sirbikyte, "Liability for damages caused by artificial intelligence" (2015) 31 Computer Law & Security Review 376.

[14] C Wendehorst, "Strict liability for AI and other emerging technologies" (2020) 11 Journal of European Tort Law 150.

[15] European Commission Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (2021/0106) (COD) COM (2021) 206 Final ("EC Proposal").

[16] Strict liability "means that a party can be held liable despite the absence of fault". European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)) (2020) P9_TA-PROV(2020)0276. Page 3, Paragraph C.

[17] See, for instance, W-G Ringe and C Ruof, "A regulatory sandbox for robo advice" (2018) European Banking Institute Working Paper Series no. 26; BK Olsen, "Sandbox for responsible artificial intelligence" (*DataEthics*, 14 December 2020) <https://dataethics.eu/sandbox-for-responsible-artificial-intelligence/> (last accessed 6 March 2021); H Kruyne, "Letter: regulators must get in the sandbox with AI innovators" (*Financial Times*, 22 January 2020) <https://www.ft.com/content/7f0c6eb2-3bb3-11ea-a01a-bae547046735> (last accessed 6 March 2021).

[18] EC Proposal, supra, note 15, Art 53.

general, should aim for uniformity since the use and effect of the AI technology would ultimately extend beyond each individual Member State.

This paper argues that a strict liability regime would be difficult and costly to implement given the chilling effect that a strict liability regime would impose on AI innovation. Of particular concern are the compliance costs and barriers to entry that a strict liability regime for AI creates for small and medium-sized enterprises (SMEs). To make this argument, the authors analyse the reasons for which the strict liability rules established by the EU in the various proposed AI regulations are not appropriate for the regulation of AI. While the EU recognises the role of sandbox regulation to support AI innovation, the authors argue for a more robust and unified regulatory sandbox model that can suitably complement the EU's strict liability regime in order to mitigate the stifling of innovation and the costs of a strict liability regime. The UK Information Commissioner's Office (ICO) regulatory sandbox for AI technologies is selected as a case study to illustrate the use of sandbox regulation in the AI context.

After introducing this paper's proposal for the use of a sandbox framework in AI regulation in Section I, Section II of this paper highlights the complexity of AI liability regulation. It provides an overview of approaches to regulating AI, including the competing approaches of fault-based liability and strict liability. Section II also discusses the use of a sandbox approach to AI regulation to complement a strict liability regime. Section III provides the necessary background on the EU's proposed AI regulations, which are currently a set of *ad hoc* EP committee reports and resolutions, EC proposed regulation and Council conclusions that propose a series of AI regulations, namely regulations on AI ethics and liability. Section III explores the regulation of high-risk AI applications under the EU's strict liability approach with a sandbox regulation to demonstrate the unique challenges of AI regulation. Finally, Section IV applies a sandbox approach to AI regulation using the UK ICO's regulatory sandbox for AI technologies as a case study. Section V concludes.

## II. Liability and sandbox regimes for AI regulation

This section highlights the need for AI regulation and the complex challenges regulators face. One particular challenge is the choice of approach to AI regulation, the two dominant approaches being fault-based liability championed by scholars in the USA and the strict liability of the EU across the Atlantic. Regardless of the approach, regulators must remain mindful of stifling innovation in the midst of the race for AI supremacy. For the sake of balancing the dual interests of regulation and innovation, AI regulators could borrow from the financial technology (FinTech) playbook, which was the first to apply a sandbox approach in the finance industry to regulate new technology. This allows commercial technology to be tested within an experimental phase without the normal regulatory requirements, under the supervision of regulators. This reduces the barriers to entry, allows the technology to prove its capabilities and enables the regulator to better understand the technology and business it is regulating. AI regulators could also apply the FinTech sandbox regulatory approach to AI regulation.

## 1. The necessity and complexity of AI liability regimes

While AI brings promises of economic and social benefits in a vast range of industries and social contexts, its regulation is necessary because of the equally compelling risks posed by AI to individuals and society. Commonly recognised AI risks include the threat to fundamental rights, such as privacy, and to individual and public safety and interests.[19]

Regulating AI remains complex, however, because of the importance of balancing the fostering of AI use and innovation while protecting safety and fundamental rights. In part, the source of this complexity is in the very nature of AI, which relies on the use of big data to train and develop algorithms that function at tremendous speeds, and sometimes using neural networks that operate in ways that remain incomprehensible even to the AI developer. An AI system may also be autonomous, the level of which continues to evolve, and could be used in real time, as explicitly recognised by the EC Proposal.[20] AI could also take on different roles, whether as a component within a product or larger system or as a standalone system with applications in unlimited environments and purposes. Predicting the risks that AI may pose is therefore an inherently challenging task. Finally, since the development and operation of AI involves multiple and sometimes overlapping actors, attributing the source or cause of liability could be a challenge, especially in complex neural network AI systems. Information asymmetry arises relative to the control of and knowledge about the AI.

EU regulators, explicitly recognising the inherently unique challenges of regulating AI, approach AI regulation not solely by focusing on the characteristics of AI, but by categorising the level of risk that an AI may pose. The EC Proposal includes a risk-based approach to AI in order to determine how AI should be regulated in different cases. The proposal is for four categories of risk, namely:

1) Unacceptable risk, which is banned in the EU since it contravenes basic fundamental rights.[21]
2) High risk, which may adversely affect human health and safety or fundamental rights[22] protected by the EU Charter of Fundamental Rights (as classified in Article 6(1) and Annex II as well as Article 6(2) and Annex III). Mandatory requirements are imposed on high-risk types of AI and they are assessed to ensure they comply, in which case they are deemed permissible.[23]
3) Limited risk, which imposes requirements for transparency in certain circumstances so users know it is a machine that they are interacting with.[24]
4) Minimal risk, which allows other types of applications to be legally developed.

Specific rules pertaining to high-risk AI are contained in Title III of the EC Proposal. High-risk AI is classified depending on the AI system's intended purpose. Annex II

[19] ibid, Preamble, para 4.
[20] ibid, Preamble, paras 8–23.
[21] ibid, Art 51.
[22] Chapter 1 of Title III provides for the two categories of safety or fundamental rights.
[23] Charter of Fundamental Rights of the European Union OJ C 326, 26.10.2012, 391–407.
[24] EC Proposal, supra, note 15, Art 52.

contains a list of typologies of AI systems categorised per area of application. Annex III's amendable list of AI that is of high risk to fundamental rights includes risks that are likely soon to materialise, or have already materialised, coupled with a risk assessment methodology. For fundamental rights to be protected, AI should be unbiased in its decision-making or data assessment in order to prevent discriminatory outcomes. This may not happen where there is limited transparency or understanding of an AI application's processes.

"Black box" AI is one non-exclusive example of such a lack of transparency and accountability, since the data sample used or decision-making processes may be unclear to human observers. Researchers have also found evidence of obscure or discriminatory outcomes.[25] Such discriminatory results would breach EU fundamental rights, and these types may be considered high risk and impermissible.

Black box AI refers to AI systems, primarily opaque neural networks, whose inputs and operations are visible neither to the user nor to any other interested party. In other words, a black box is, generally speaking, an impenetrable system. As stated by Bathaee, "modern AI systems are built on machine-learning algorithms that are in many cases functionally black boxes to humans. At present, it poses an immediate threat to intent and causation tests that appear in virtually every field of law. These tests, which assess what is foreseeable or the basis for decisions, will be ineffective when applied to black-box AI".[26] This is a problem for liability as it is hard to connect the harm to the developer.

While a black box AI does not necessarily mean that the AI is high risk, a black box AI may pose a threat in the sense that the processes behind black box development are largely self-directed and generally difficult for data scientists and users to interpret.[27] The algorithms are capable of learning from massive amounts of data. Once those data are processed, the algorithms are capable of making decisions experientially and developing biases in the decision-making process.

There are three ways to address the black box AI issue when it comes to AI liability. The first approach is to create a strict liability regime that places the onus of understanding the source of the liability on the developers and users of AI. A second approach is to prohibit black box AIs altogether. The first and second approaches are similar in that they address the information asymmetry problem by placing the burden on the developer. The drawback to these approaches is the potential for stifling innovation in the midst of an AI race. A third approach is a sandbox regulation that can target the intent, causation and mitigation of liability in specific high-risk scenarios. In this way, the sandbox approach could complement the existing strict liability regime.

[25] J Angwin et al, "Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks" (*ProPublica*, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (last accessed 1 August 2021); M Carabantes, "Black-box artificial intelligence: an epistemological and critical analysis" (2019) 35 AI & Society 309.

[26] Y Bathaee, "The artificial intelligence black box and the failure of intent and causation" (2018) 31 Harvard Journal of Law & Technology 889.

[27] J Truby, "Governing artificial intelligence to benefit the UN Sustainable Development Goals" (2020) 28 Sustainable Development 946.

Bathaee considers that the solution to the intent and causation problems should not be strict liability or a regulatory framework of granularly defined transparency standards for AI design and use. Both solutions risk stifling innovation and erecting entry barriers, especially for smaller firms. Rather, Bathaee argues that the most suitable solution would be a sliding scale system, which adapts the current regime of causation and intent by relaxing their requirements for liability when AI is allowed to operate autonomously or when AI lacks transparency. On the other hand, the sliding scale system preserves the traditional intent and causation tests when humans supervise AI or when the AI is transparent.[28]

## 2. Weighing fault-based liability and strict liability in AI regulation

Before delving into the EU's approach to high-risk AI, it is important to examine the predominant approaches to AI regulation with the hope of better understanding the EU's legal perspective. The approaches to regulating AI can be divided into two categories: fault-based liability and strict liability. These two theories of liability have competing approaches to allocating the burden and costs of risk knowledge and risk control.

Fault-based liability (negligence) focuses on assessing the level of care based on a legally recognised duty and a corresponding breach and causation.[29] Actors therefore are incentivised to act within the expected level of care, as determined by courts through policy and cost–benefit analysis, to avoid the risk and cost of liability. Zech notes that a limit in fault-based liability's regulation of AI lies in the information asymmetry between courts and producers.[30] In other words, fault-based liability places the risk knowledge on courts, which are not likely to have the same technological and risk knowledge and resources as developers and manufacturers. Zech's analysis, however, disregards the role of parties and lawyers in an adversarial process. Courts could leverage their judicial authority and ability to shift the burden of proof and presumptions to the parties in the litigation and mitigate the information asymmetry. Zech also notes that the incentive factor to adhere to the level of care may fail in novel technologies such as AI when the actor has no risk knowledge that would allow them to meet the level of care. In other words, information asymmetry could arise between the user and the producer and therefore not allow the user to meet its level of care. Finally, fault-based liability would not cover the risks posed by new technologies that are unforeseeable and therefore could not meet the requirements of legal causation.[31]

An alternative to fault-based liability that aims to address the risk posed by information asymmetry is strict liability. When applied to high-risk AI, strict liability imposes the risk, knowledge and costs on the producer and operator of the AI regardless of its conduct. Strict liability essentially internalises the economic risks of AI, which Zech sees as a means of creating public acceptance of AI by eliminating its economic risks.[32] Zech makes his public acceptance argument under the assumption that

[28] Bathaee, supra, note 26, 932–38.

[29] Fault-based liability is the default approach in the EU Member States. ibid; H Zech, "Liability for AI: public policy considerations" (2021) ERA Forum 147.

[30] ibid, 150.

[31] ibid.

[32] ibid, 150.

AI development and innovation would not flourish without risk controls in place.[33] In practice, however, much of the AI innovation thus far has been made without such risk controls and public assurances. Instead, the public has been far too eager to accept the risks, primarily consenting to privacy risks in exchange for free access. Rather, strict liability acts primarily as a means of social risk distribution. As Zech notes, producers will likely conduct a cost-based analysis in a strict liability regime.[34] This analysis could either limit AI innovation when the liability far exceeds the benefits of the invention or encourage risk-taking when the benefits far exceeds the liability. If the risk and benefit balance is set correctly, strict liability could incentivise producers and operators to make AI safer, assuming of course that they have the capacity to make it so.[35]

Strict liability, however, has significant drawbacks, the most important of which is stifling innovation by deterring companies, especially small start-ups, from taking the risk of liability exposure. What may be a small sum to a large company, which can tolerate such economic costs, could end a small start-up company, which, on the other hand, must rely more on innovation and experimentation with a lower prototype-to-market risk threshold.

Additionally, while strict liability can be justified due to the information asymmetry problem, regulators may overlook the fact that the development of AI requires the cooperation of multiple parties that may separately have information asymmetry amongst themselves concerning data, algorithms or people. Additionally, AI may pose high risks that are unforeseeable to the producers and operators. An example of this is the development of neural networks that essentially act like black boxes, where their processes remain virtually unknown to the developer and operator in terms of various factors. In such circumstances, the strict liability approach is unable to address the information asymmetry dilemma.

The multiple-producer or multiple-operator scenario and the black box scenario could also trigger issues relating to causation. The EU, for example, approaches such a situation in the EC Proposal by shifting the burden of proof onto the producer or operator. This approach, however, will likely favour large companies that already dominate the AI industry since they have leverage over small companies regarding the AI development process and could alternatively shield themselves from liability.

That the EU opted for a strict liability regime exemplifies the information asymmetry concern of EU regulators when it comes to AI. In our view, however, the strict liability approach could mitigate the primary concern of stifling innovation by allowing high-risk AI activities to be tested through a sandbox, which can test the risks of the technology while also allowing for the consideration and adoption of appropriate laws tailored to the specific activities. Such a sandbox approach could complement an existing strict liability approach such as that in the EU. The following sections will argue in favour of this view.

[33] ibid, 152–53.
[34] ibid, 154–60.
[35] ibid.

## 3. What is a sandbox regulation?

A sandbox both enables entrepreneurial development and informs regulatory policy. The "sandbox" concept is particularly utilised in the areas of financial innovation and FinTech, where a regulator enables experimental innovation within a framework of controlled risks and supervision.[36] This allows a " . . . form of 'beta testing' for financial technology start-ups, where firms may test their financial services technology and other financial products under supervision of the financial services authorities".[37] The UK Financial Conduct Authority (FCA) describes it as "a 'safe space' in which businesses can test innovative products, services, business models and delivery mechanisms without immediately incurring all the normal regulatory consequences of engaging in the activity in question".[38] A sandbox is a "controlled environment or safe space in which FinTech start-ups or other entities at the initial stages of developing innovative projects can launch their businesses under the 'exemption' regime in the case of activities that would fall under the umbrella of existing regulations or the 'not subject' regime in the case of activities that are not expressly regulated on account of their innovative nature, such as initial coin offerings, crypto currency transactions, asset tokenisation, etc.".[39]

Sandboxes offer several advantages to technology developers, including the ability to verify and demonstrate an innovative technology by testing it in a live environment with real consumers.[40] Direct communication between developers and regulators creates a more cohesive and supportive industry. Successive trial-and-error testing within a controlled environment mitigates the risks and unintended consequences such as unseen security flaws when a new technology gains market adoption. It is in this context, for instance, that the financial sector implemented regulatory sandboxes to avoid such flaws given the importance of this sector to any global economy.[41] Hence, similar reasoning should follow when it comes to AI activities.

A supplementary goal of the sandbox is for regulators themselves to learn and understand the product or service better.[42] This allows regulators to develop policy

[36] D Zetzsche et al, "Regulating a revolution: from regulatory sandboxes to smart regulation" (2017) 23 Fordham Journal of Corporate & Financial Law 31.

[37] J Truby, "FinTech and the city: sandbox 2.0 policy and regulatory reform proposals" (2018) 34 International Review of Law, Computers & Technology 277.

[38] Financial Conduct Authority (FCA), "Regulatory sandbox" (November 2015) <https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf> (last accessed 11 January 2020).

[39] RG de la Cruz, "The regulatory sandbox and potential opportunities for Spanish FinTechs" (2018) 7 Spanish and International Economic & Financial Outlook 19.

[40] D Arner, "Financial regulation, technology and the future of finance" in J Walker, A Pekmezovic and G Walker (eds), *Sustainable Development Goals: Harnessing Business to Achieve the Sustainable Development Goals through Technology, Innovation and Financing* (Hoboken, NJ, Wiley 2019).

[41] JJ Goo & J-Y Heo, "The impact of the regulatory sandbox on the FinTech industry, with a discussion on the relation between regulatory sandboxes and open innovation" (2020) 6 Journal of Open Innovation Technology Market and Complexity 43; RP Buckley et al, "Building FinTech ecosystems: regulatory sandboxes, innovation hubs and beyond" (2020) 61 Washington University Journal of Law & Policy 55.

[42] R Parenti, "Regulatory sandboxes and innovation hubs for FinTech", Study for the Committee on Economic and Monetary Affairs, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020, at 24 <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652752/IPOL_STU(2020)652752_EN.pdf> (last accessed 22 March 2021).

and regulations to accommodate, supervise and control sectoral innovation within and outside of the sandbox. Regulations tested within the sandbox can determine the most appropriate framework for real-world regulations outside of the sandbox. Updating regulations and ending regulatory uncertainty would make the jurisdiction a more attractive destination for technology developers and investors. Quan warns that "regulatory uncertainty is the result of outdated regulations unable to catch up with innovation. Regulatory fear, on the other hand, is caused by risk-averse regulators unwilling or unable to green-light novel products that may be perfectly compliant with regulations".[43] Sound innovation can be promoted through a flexible regulatory regime, where regulators can provide guidance. Truby explains that "[r]ather than only regulating financial technology to keep it in line with other types of financial activities, legitimising sound and legally compatible business and financing structures and activities involved to facilitate their growth, may offer substantial opportunities for economic development. This can be done whilst retaining both supervisory control, and investor and client protections".[44] Deloitte, in 2018, in collaboration with Innovate Finance, interviewed several firms that have been through or were still going through the FCA's sandbox to seek their views on their journey.[45] The main conclusions are that regulation is no longer a barrier to innovation given the various benefits of the sandbox, but there remains room for improvement.

Multiple jurisdictions have experimented with sandboxing, typically in the financial innovation sector.[46] The FCA is credited with creating the first formal regulatory sandbox and propagating the concept throughout the world.[47] Since the FCA launched its FCA sandbox in 2016, it has supported more than 700 firms and increased their average speed to market by 40% compared with the regulator's standard authorisation time. The FCA's sandbox recently opened applications for cohort 7 (a clear proof of its success), and it is currently open to authorised and unauthorised firms that require authorisation and technology businesses that want to deliver innovation in the UK financial services market.

A number of other major jurisdictions are seeking to follow suit in order to capitalise on the fast-growing FinTech sector. In the USA, the Consumer Financial Protection Bureau (CFPB) was the first regulatory agency to set up a dedicated FinTech office to study FinTech and to help promote consumer-friendly innovation. Another very successful regulatory sandbox is that of Singapore, which takes a more innovator-centred approach than the UK prototype in terms of lower entry barriers and a greater emphasis on industry benefits.[48] The Monetary Authority of Singapore

[43] D Quan, "A few thoughts on regulatory sandboxes" (*Stanford University, Stanford PACS*, 2019) <https://pacscenter.stanford.edu/a-few-thoughts-on-regulatory-sandboxes/> (last accessed 11 January 2020).

[44] Truby, supra, note 37, 8–9.

[45] Deloitte, "A journey through the FCA regulatory sandbox. The benefits, challenges and next steps" (2018) <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/financial-services/deloitte-uk-fca-regulatory-sandbox-project-innovate-finance-journey.pdf> (last accessed 13 December 2020).

[46] DW Arner, J Barberis and RP Buckley, "FinTech, RegTech, and the reconceptualization of financial regulation" (2017) 37 Northwestern Journal of International Law & Business 371.

[47] J Truby, "Decarbonizing Bitcoin: law and policy choices for reducing the energy consumption of blockchain technologies and digital currencies" (2018) 44 Energy Research & Social Science 399.

[48] L Lin, "Regulating FinTech: the case of Singapore" (2019) 35 Banking and Finance Law Review 93.

(MAS) published its guidelines for the financial regulatory sandbox in June 2016.[49] The regulatory sandbox of MAS aims to transform Singapore into the centre of the smart finance industry. Sandbox entities are freed from the administrative and financial burdens imposed under ordinary compliance processes, and they are also entitled to a broader testing ground (whereas licensed operators may reach out only to a limited group of clients), an element that is crucial for refining their core technologies.

As jurisdictions rush to develop their sandboxes, Quan cautions on the importance of properly implementing a sandbox, as "too often sandboxes are misunderstood, misused, or mismanaged. Regulatory agencies should use sandboxes to keep up to date with fast-paced innovation and promote market competition without sacrificing consumer protection. Real innovation-minded regulatory agencies see sandboxes as means, not ends".[50]

This is of the utmost importance as scholars and international bodies have raised concerns regarding the use of regulatory sandboxes. For instance, Allen argues in the context of financial regulations that as "competition among countries for fintech business intensifies, the phenomena of regulatory arbitrage and race to the bottom are likely to drive the regulatory sandbox model toward further deregulation and disincentivise vital information sharing among financial regulators about new technologies".[51] Meanwhile Ahern argues that "pressure on regulators to produce sandbox successes and to compete with other sandboxes may influence the exercise of regulatory discretion and produce regulatory distortions that affect competition in FinTech markets".[52] In contrast to those supporting regulatory sandboxes, many experts strongly oppose regulatory sandboxes, seeing them as actually slowing down and halting innovation.[53] Indeed, practical challenges facing regulatory sandboxes have been noticed. For instance, consumers may perceive products tested in the context of a sandbox as if authorities have endorsed them, which in turn has negative legal consequences.[54]

Additionally, competing sandboxes treat the issue of liability in different ways, and some are silent on the matter. Generally, sandboxes only exclude businesses from enforcement action by the financial regulator and not from consumer liability.[55] National laws on liability usually still apply.[56]

[49] Truby, supra, note 37, 5.

[50] Quan, supra, note 43.

[51] HJ Allen, "Sandbox boundaries" (2020) 22 Vanderbilt Journal of Entertainment & Technology Law 299, 299.

[52] D Ahern, "Regulators nurturing FinTech innovation: global evolution of the regulatory sandbox as opportunity-based regulation" (2019) 15 Indian Journal of Law and Technology 345, 345.

[53] Quan, supra, note 43.

[54] M Nikolova, "EU supervisory authorities outline challenges and risks for innovation hubs, regulatory sandboxes" (*Finance Feeds*, 7 January 2019) <https://financefeeds.com/eu-supervisory-authorities-outline-challenges-risks-innovation-hubs-regulatory-sandboxes/> (last accessed 6 March 2021).

[55] See, for example, the situation in India: P Advani, "Regulating to escape regulation: the sandbox approach" (*University of Oxford*, 6 August 2020), <https://www.law.ox.ac.uk/business-law-blog/blog/2020/08/regulating-escape-regulation-sandbox-approach> (last accessed 26 May 2021).

[56] I Jenik and K Lauer, "Regulatory sandboxes and financial inclusion" (Working Paper. Washington, DC, CGAP, 2016) at 4 <https://www.cgap.org/sites/default/files/Working-Paper-Regulatory-Sandboxes-Oct-2017.pdf> (last accessed 26 May 2021).

These are some of the shortcomings mentioned in the literature, while many more exist. AI innovation presents many opportunities and uncertainties, especially with uses of AI that are considered to be more dangerous, such as AI algorithmic experimentation (automated and black box AI).[57]

## III. The EU's approach to high-risk AI applications

This section explores the EU's efforts at proposing AI regulations by first providing an overview of the proposals. The section then discusses the EU's strict liability approach through a regulatory sandbox and considers competing approaches to AI liability, including the limits of a strict liability regime.

### 1. Overview

The European Parliamentary Research Service published a study establishing the case for regulating the civil liability of AI at the EU level, reporting that not doing so would "potentially discourage innovation, increase prices for consumers, substantially increase administrative costs for public administrations and judicial bodies and ultimately even challenge the social desirability of the overall liability system".[58] The Council of the European Union (Council) further set out the risks of AI applications to fundamental rights in the EU, establishing the need for regulatory intervention.[59]

The Council has adopted conclusions on Regulatory Sandboxes and Experimentation Clauses as tools for an innovation-friendly, future-proof and resilient regulatory framework that masters disruptive challenges in the digital age.[60] Having examined potential regulatory frameworks for AI, the EP proposed new rules for the Commission to adopt.[61] The first two EP Resolutions related to the creation of a framework on ethical aspects of AI, robotics and related technologies,[62] as well as a

[57] J Truby, R Brown and A Dahdal, "Banking on AI: mandating a proactive approach to AI regulation in the financial sector" (2020) 14 Law and Financial Markets Review 110.

[58] T Evas, European Parliamentary Research Service, "Civil liability regime for artificial intelligence: European added value assessment" (PE 654.178 – September 2020) at 37 <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654178/EPRS_STU(2020)654178_EN.pdf> (last accessed 26 May 2021).

[59] A Renda et al, "Study to support an impact assessment of regulatory requirements for artificial intelligence in Europe" (Final Report, European Commission, Publication Office of the European Union, April 2021) at 7–9 and 149 <https://op.europa.eu/en/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1/language-en/format-PDF/source-204305195> (last accessed 25 May 2021).

[60] Council Conclusions of 16 November 2020 on regulatory sandboxes and experimentation clauses as tools for an innovation-friendly, future-proof and resilient regulatory framework that masters disruptive challenges in the digital age 12683/1/20 REV 1 (2020) 13026/20.

[61] European Parliament, "Parliament leads the way on first set of EU rules for artificial intelligence" (*Press Releases*, 20 October 2020) <https://www.europarl.europa.eu/news/en/press-room/20201016IPR89544/parliament-leads-the-way-on-first-set-of-eu-rules-for-artificial-intelligence> (last accessed 13 December 2020).

[62] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)) (2020) P9_TA-PROV(2020)0275.

civil liability regime.[63] A further Resolution on intellectual property rights (IPRs)[64] and the Resolution of the Council on regulatory sandboxes and experimentation clauses[65] call for the adoption of new regulations.

The AI Resolution offers a long list of required ethical principles such as "human dignity, autonomy and safety ... social inclusion, democracy, plurality, solidarity, fairness, equality and cooperation ...".[66] The Resolution on civil liability regime for AI provides a brief overview of the concept of civil liability and focuses on imposing strict liability for high-risk AI systems, determining the amount and extent of compensation and the limitation period. It also includes provisions on fault-based liability for other AI systems, national provisions on compensation and limitation periods, contributory negligence, joint and several liability, etc.[67] Assessing the interplay between IPRs and AI, the Council Conclusions calls the EC to encourage the exchange of "information and good practices regarding regulatory sandboxes" between Member States,[68] for various purposes. These include determining the current use of regulatory sandboxes in the EU[69] and identifying "experiences regarding the legal basis, implementation and evaluation of regulatory sandboxes".[70] Based on these efforts, the EC ultimately issued its Artificial Intelligence Act and Amending Certain Union Legislative Acts in April 2021.[71]

The most recent EC Proposal creates new harmonised rules for the regulation of AI while prohibiting the use of certain AI practices. It also imposes specific requirements related to high-risk AI systems and those operating them and lays down harmonised rules for transparency, market monitoring and surveillance.[72] The proposal presents a holistic and comprehensive regulation that covers, in addition to the issues mentioned above, governance structure, implementation of the regulation and the creation of codes of conduct.[73] Of particular interest for purposes of this paper is the addition of the AI regulatory sandbox.

## 2. The EU's strict liability regime for AI

The EU does not have a specific liability regime applicable to AI, as the main applicable legislation is the EU Product Liability Directive (PLD), while national liability and damage rules apply in case of accidents. Despite its importance, the PLD is seen as a limited instrument when applied to advanced technologies such as AI systems. The main issues of contention are related to its scope, the definition of the notion of defect, the types of damages covered, the scope of liable persons and the exemptions

[63] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)) (2020) P9_TA-PROV(2020)0276.

[64] European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies (2020/2015(INI)) (2020) P9_TA-PROV(2020)0277.

[65] EU Council, supra, note 60.

[66] European Parliament, supra, note 62, Art 5(1).

[67] European Parliament, supra, note 16.

[68] EU Council, supra, note 60, 6(14).

[69] ibid, 6(14)(a).

[70] ibid, 6(14)(b).

[71] EC Proposal, supra, note 15.

[72] ibid, Art 1.

[73] ibid, 15–16.

and defences. Consequently, it was concluded that current EU secondary law related to liability is insufficient when it comes to its application to AI, be it for covering already existing areas within EU law or emerging risks not covered by European regulations.[74] Moreover, the national liability rules vary across states and jurisdictions where fault-based liability and strict liability are applicable. In this context, strict liability applies to "damages caused by things, dangerous activities, animals and vicarious liability".[75] States either adopted general and flexible strict liability provisions or "exhaustive, closed lists, or no provisions on strict liability for the analysed group of situations", which is considered narrow. Given this reality, there was a need to adopt new liability rules applicable to AI.[76] These rules were adopted recently by the EC, though as proposals that require the approval of the EP. As such, they represent a first step towards a legally binding regulation on liability of AI.

The EP Resolution on a civil liability regime for AI proposes strict liability for high-risk AI systems, including damage or harm resulting from a "physical or virtual activity, device or process driven" by a high-risk AI system. The list of these systems and critical sectors is annexed to the Resolution and can be amended by the EC. Operators will still be held liable even when acting with due diligence or when the damage occurred by "an autonomous activity, device or process driven by their AI-system", although force majeure is excluded.[77] The frontend operator has the responsibility to cover the operations with an adequate liability insurance and the backend operator must cover its services with an adequate business liability or product liability insurance in accordance with the compensation articles of this regulation. The requirement of purchasing insurance shall be considered as satisfied upon the condition that compulsory insurance regimes and voluntary corporate insurance funds "cover the amounts and the extent of compensation" stipulated within the regulation. This law takes primacy over national laws in case of "conflicting strict liability classification of AI-systems".[78] The draft regulation made a distinction between high-risk AI systems, where strict liability applies, and other low-risk AI systems, where fault-based liability applies.[79]

The EP justified the use of strict liability rules by noting the need to protect the general public from high-risk autonomous AI systems.[80] It also emphasised the need for a "common strict liability regime for those high-risk autonomous AI-systems".[81] The EP further underlined the need for clear criteria and definition of the term "high risk" given the existence of various levels of risks.[82] The EP considered that the default

[74] European Parliament, "Civil liability regime for artificial intelligence" 5–10 (European Added Value Assessment, European Parliamentary Research Service, September 2020) at 5–10.

[75] ibid, 32.

[76] ibid, 32.

[77] European Parliament, supra, note 16, Art 4. The frontend operator is defined as the "natural or legal person who exercises a degree of control over a risk connected with the operation and functioning of the AI-system and benefits from its operation". The backend operator is defined as the "natural or legal person who, on a continuous basis, defines the features of the technology, provides data and essential backend support service and therefore also exercises a degree of control over the risk connected with the operation and functioning of the AI-system" European Parliament, supra, note 16, p 7, par. 12.

[78] ibid.

[79] ibid.

[80] ibid, p 7, para 14.

[81] ibid.

[82] ibid.

position for an AI system not yet classified as high risk should be "subject to strict liability if it caused repeated incidents resulting in serious harm or damage".[83]

The Resolution on the framework of the ethical aspects of AI, robotics and related technologies examined the issue of liability and even high-risk AI systems without, however, discussing strict liability.[84] The Resolution on AI IPRs did not mention strict liability and only mentioned liability once, noting that AI and the associated technologies used for the "determination of liability for infringements of IPRs cannot be a substitute for human review carried out on a case-by-case basis ...".[85] The Council Conclusions on Regulatory sandboxes and experimentation clauses did not mention the issue of liability or strict liability at all in the document.[86]

Although the EC Proposal does not explicitly refer to strict liability, the latter mentions liability in specific provisions.[87] Moreover, the extremely detailed and numerous provisions adopted within this proposal addressing high-risk AI systems highlights the influence of the previous EU regulations, mainly the one on civil liability, as the proposal creating clear, direct and detailed rules concerning high-risk AI systems that must be complied with. These provisions, for instance, are related to classification rules for high-risk AI systems, risk management systems, transparency and provision of information to users, human oversight, obligations of providers of high-risk AI systems, obligations of product manufacturers, obligations of importers, obligations of distributors and obligations of users of high-risk AI systems.[88]

These rules should apply to physical or virtual AI that harms or damages life, health, physical integrity and/or property or that causes significant immaterial harm if it results in a verifiable economic loss. Despite admitting that high-risk AI activities are still rare, the members of the EP believe that their operators should hold insurance similar to that used for motor vehicles.

The EU's concern towards liability is not new, though, since it has been a recurring topic these last few years. In November 2019, the EC published a report on AI, new technologies and liability: the "Liability for Artificial Intelligence and Other Emerging Digital Technologies" report.[89] This report contained several recommendations, such as the default rule of strict liability for certain operators as well as producers for defects in products or digital content incorporating emerging digital technology. The possibility of giving legal personhood to AI systems was also rejected.

The most important findings of this report is on how liability regimes should be designed (and changed if necessary).[90] A person is held strictly liable for harm or damage occurring from the use of a permissible technology that poses an increasing risk of harm. In case a service provider providing technical expertise has more control than the "owner or user of an actual product or service equipped with AI", this reality

[83] ibid, p 8, para 21.
[84] European Parliament, supra, note 62, p 5, para J.
[85] European Parliament, supra, note 64, p 7, para 16.
[86] European Council, supra, note 60.
[87] EC Proposal, supra, note 15, Arts 2(5), 33(8) and 53(4).
[88] ibid, Arts 6–29.
[89] Commission, "Liability for Artificial Intelligence and Other Emerging Digital Technologies" (2019) <https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608> (last accessed 13 December 2020).
[90] ibid.

should be considered when figuring out who operates the technology. A person still needs to "properly select, operate, monitor and maintain the technology in use" even when there is no increased risk of harm where liability will occur in case of failure to fulfil these duties. The certain degree of autonomy that a technology may have does not exonerate a person from liability. Manufacturers of products or digital content involving digital technology are held liable for damage in case of product defects even when the defect occurs when the product is with the producer. The report recognises the information asymmetry and the difficulty in the victim proving liability due to the technology used. Finally, there is no need for granting legal personality to devices or autonomous systems as only persons and bodies can be held liable for any harm.[91]

The importance of liability was also highlighted in the EC's White Paper on Artificial Intelligence ("White Paper on Artificial Intelligence – A European Approach to Excellence and Trust"),[92] as well as in its associated Report on Safety and Liability ("Report on the Safety and Liability Implications of AI, the Internet of Things and Robotics").[93] The White Paper proposed the adoption of a regulatory framework for high-risk AI systems, which will have to comply with several requirements. Even though the White Paper does not address the issue of liability and AI thoroughly, it acknowledges that there is a need to improve the legal framework in order to better assign responsibilities between the actors.[94] The liability framework may be challenged by AI systems, limiting their effectiveness. Yet an equal level of protection must be provided to those suffering an injury or damage because of AI systems, similarly to other technologies. The White Paper concludes by providing an overview of the various AI benefits for "citizens, companies and society as a whole, provided it is human-centric, ethical, sustainable and respects fundamental rights and values", while emphasising the need to "develop and reinforce the necessary industrial and technological capacities" of the Union.[95]

The associated Report on Safety and Liability goes further by focusing on the following areas, mainly: (1) establishing a strict liability regime for AI systems with a "specific risk profile" and coupling it with a mandatory insurance requirement; (2) examining the question of whether or not to adapt the burden of proof regarding fault and causation for other AI systems (eg those with a low risk) – the EC thus considers a differentiated liability approach depending on the level of risk posed by AI systems; and (3) considering reversing or alleviating the burden of proof required by national rules "for damage caused by the operation of AI-systems, through an appropriate EU initiative". The Report proposes a risk-based approach to liability given the emerging challenges raised by AI and related technologies. The Report proposes "certain adjustments to the Product Liability Directive and national liability regimes through appropriate EU initiatives ... on a targeted, risk-based approach, i.e. taking into account that different AI applications pose different risks".[96]

[91] ibid.

[92] Commission, "White Paper on Artificial Intelligence – A European Approach to Excellence and Trust" COM (2020) 65 final.

[93] Commission, supra, note 8.

[94] Commission, supra, note 92.

[95] ibid.

[96] Commission, supra, note 8.

The White Paper identified the importance and need to adopt a common approach at the EU level. Consequently, the JURI Committee of the EP made available its Draft Report with Recommendations to the EC on a Civil Liability Regime for AI in April 2020.[97] The latest draft set of harmonised rules could serve as a basis for a future legislative initiative by the EC.

The Draft Report creates a twofold liability regime depending on the risk of the AI system. High-risk systems are subject to a strict liability regime in which the deployer of the system is liable without fault (Article 4.1). Low-risk systems remain subject to fault-based liability only targeting the deployer (Article 8.1). The report had other key elements. All of the elements of the report were later included within the civil liability framework initiative adopted in October 2020 and examined previously.

## 3. The role of sandboxes in the EC Proposal

Under Title V, Articles 53–55,[98] the EC Proposal explicitly adds sandbox regulation to AI liability regulation within the EU. Article 53 essentially defines a regulatory sandbox as "a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan".[99] To understand better the role of sandboxes in the EC Proposal, it is perhaps important to revisit prior proposals and recommendations made in the EU concerning the role of a regulatory sandbox in AI regulation.

In the EU, some authors have proposed the creation of an EU-level regulatory sandbox, which would make EU Member States, collectively, a more attractive destination for innovation.[100] In this sense, the Expert Group on Regulatory Obstacles to Financial Innovation recommended that the EC and the European Supervisory Authorities (ESAs) should further consider the establishment of an EU-level regulatory sandbox. The EU Digital Finance Strategy envisages the development of a "procedural framework for launching cross-border testing and other mechanisms facilitating firms' interaction with supervisors from different Member States" by mid-2021. Related to this, the 2019 ESAs Joint Report identified a set of principles for the establishment and operation of regulatory sandboxes.[101] These principles include several requirements such as having "clearly defined and published eligibility criteria for entry" and "clearly defined and published key information which needs to be submitted by the companies in support for the application to participate in the regulatory sandbox".[102]

[97] JURI Committee, European Parliament, "Draft Report with Recommendations to the Commission on a Civil Liability Regime for AI" (27 April 2020).

[98] EC Proposal, supra, note 15, Arts 53–55.

[99] ibid, Art 53.

[100] R Parenti, "Regulatory sandboxes and innovation hubs for FinTech: impact on innovation, financial stability and supervisory convergence" (2020) Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies, European Parliament <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652752/IPOL_STU(2020)652752_EN.pdf> (last accessed 13 December 2020).

[101] European Supervisory Authorities (ESAs), "Report FinTech: regulatory sandboxes and innovation hubs" (2019) <https://www.esma.europa.eu/sites/default/files/library/jc_2018_74_joint_report_on_regulatory_sandboxes_and_innovation_hubs.pdf> (last accessed 13 December 2020).

[102] ibid.

The idea of using a regulatory AI sandbox in the EU is not an entirely novel concept. Outside of the EU, both the UK (see Section III.4 below) and Norway have developed AI sandboxes. The Norwegian Data Protection Agency developed a data-focused sandbox as part of its National Strategy for AI,[103] to ensure ethical and responsible use of data. The aim is to increase supervision of data usage and to inform policy development.

In January 2020, the EC organised a series of workshops to refine further the concept of reference Testing and Experimentation Facilities (TEFs) for AI with the help of the invited experts and national delegations. In the upcoming Digital Europe Programme 2021–2022, the EC will launch calls to create these TEFs. The idea of creating sandboxes was discussed during these workshops. Furthermore, the EC Communication "Coordinated Plan on Artificial Intelligence" from 2018[104] referred to sandboxes, stating the need to establish "world-reference testing facilities" to test and experiment with new technologies in the real world in various areas such as "mobility, healthcare, manufacturing, agro-food or security", while limiting the "number of specialised large-scale reference sites" across the EU. The suggested facilities can have regulatory sandboxes in specific areas where the law can be flexible for the duration of the sandbox. The Communication only made reference to "large-scale testing facilities" in some specific AI areas, but no mention was made of high-risk AI.[105]

Both the EC and the EP had taken matters a step further when recognising regulatory sandboxing as a highly desirable tool for coping with the regulatory challenges presented by new technologies, especially AI. The EC did so in its Coordinated Plan on Artificial Intelligence,[106] while the EP did so in its Resolution of 12 February 2019 on a comprehensive European industrial policy on AI and robotics (2018/2088(INI)).[107] The Resolution states the need for a "deeper understanding of the technology and [to] integrate it into their daily life" given the importance of social acceptance for the future of AI where effective communication is vital. Indeed, public awareness of AI affects it acceptance, emphasising the need for the EC and Member States to share credible information. To that end, the Resolution encouraged the use of AI-specific regulatory sandboxes to introduce "innovative new ideas, allowing safeguards to be built into the technology from the start, thus facilitating and encouraging its market entry" and to "test the safe and effective use of AI technologies in a real-world environment".[108]

In its current state, the EC Proposal does not mandate Member States to create sandboxes, but rather encourages the competent authorities of Member States to create regulatory sandboxes, including their basic framework of facilitating the development, experimentation and testing of AI innovations under strict supervision.[109]

[103] Ministry of Local Government and Modernisation, National Strategy for AI <https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=2> (last accessed 22 March 2021).

[104] Commission, "Coordinated Plan on Artificial Intelligence" COM (2018) 795 final.

[105] ibid.

[106] ibid.

[107] European Parliament, "European Industrial Policy on Artificial Intelligence and Robotics" (2018/2088(INI)).

[108] ibid.

[109] EC Proposal, supra, note 15, Art 53.

The framework would include the governance, supervision and liability of sandbox participants. The EU views such a sandbox as a pre-market deployment phase, but it does not exempt sandbox participants from AI liability. The EC also envisages the creation of common rules and a framework for the implementation of regulatory sandboxes across the relevant Member State authorities.[110] Notably, the conduct of sandbox participants could be taken into account when determining the imposition of fines under the General Data Protection Regulation (GDPR).[111]

## 4. Limitations of the EC Proposal's sandbox

The EC Proposal's sandbox approach suffers from three important limitations. The first two stem from its failure to address the need to balance liability protection and innovative experimentation for participants while in the sandbox and before market placement. First, the continued imposition of liability under Article 53(4) means that the sandbox only provides an exemption from regulatory compliance. While developers should not be allowed to use the sandbox as a shield to liability, imposing the same liability regime on sandbox participation could lead to limiting innovation. Second, the EC Proposal's sandbox could create a false perception of safety and compliance in the market. Third, the EC Proposal's sandbox could lead to uncertainty and confusion in the market since it is not mandatory to, and not necessarily uniform among, EU states.

The EU appears to have chosen the pure strict liability approach, which could restrict innovation in the field of AI, and it applies the same approach in the sandbox environment. As mentioned earlier, the recent EU civil liability framework for AI makes those operating high-risk AI strictly liable for any resulting damage.[112] Meanwhile, the EC's legislative proposal has even imposed liability when establishing a regulatory sandbox even though it is not clear whether such liability is strict.[113]

In other words, the EU approach gives a nod to sandboxes, recognising their utility in fostering innovation, but with limited protection from potential liability. The EC Proposal holds AI sandbox participants liable and thereby creates a sandbox regulatory environment that remains subject to AI strict liability. While liability exemption for regulatory sandboxes is not the norm even in FinTech, the lack of liability protection in AI sandboxes becomes more prominent because of the unlimited application of AI and its unforeseen risks, making liability testing even more important than regulatory compliance in AI sandboxes than in FinTech sandboxes.

As it is, the EC Proposal would offer a lowered incentive for AI developers to participate in the sandbox since doing so would only subject the developer to the exposure of trade secrets and algorithms and added regulatory compliance, without the benefit of a moratorium on strict liability. In this way, the AI sandbox may inadvertently lead to stifling innovation. In other words, the role of the sandbox in development, testing and validation is only for the purposes of regulatory compliance rather than to assess the AI innovation's exposure to potential liability.

[110] ibid, Art 53.
[111] ibid, Art 54.
[112] European Parliament, supra, note 107.
[113] EC Proposal, supra, note 15, Art 53.

In the instance that an AI innovation passes the sandbox regulatory requirements, it is given the stamp of approval to proceed to market placement. However, this stamp of approval does not mean that the AI system does not pose a liability risk. In this scenario, the AI sandbox creates a false perception of protection since it creates the appearance that an AI system that has gone through an AI sandbox no longer poses a threat to fundamental rights and safety. In reality, an AI system could meet regulatory compliance but could lead to unforeseen liability risks, or it could evolve into a high-risk AI through unanticipated applications. The EC Proposal's sandbox essentially only offers a limited timeframe in which to determine AI innovations' regulatory compliance before market placement.

According to Article 53 of the EC Proposal, one or more EU states may establish an AI regulatory sandbox, thereby allowing for different sandbox frameworks and implementations.[114] These various AI sandboxes need to coordinate and cooperate with the European Artificial Intelligence Board and set out implementing acts, which will later lead to common implementing rules and frameworks.[115]

Yordanova remains cautious about the applicability of an AI regulatory sandbox in the EU when it remains unclear "how a national regulator can fully participate in a regulatory sandbox when the area of regulation falls partly or entirely under EU's competences".[116] This issue needs to be fully resolved in the future, though the EC Proposal seems to leave the sandbox regulation to the primary domain of Member States, especially by not making it mandatory. Additionally, the EC Proposal envisages the creation of a common regulatory sandbox implementing rules and frameworks that would very likely require the participation and input of the various competent authorities of Member States. Still, Yordanova's criticism persists, as it remains to be seen whether the competent authorities of Member States can agree on a common implementation and framework that combines both EU and Member State rules. It may also be possible for some Member States not to implement a regulatory sandbox. Since the implementation of the sandbox is not uniform among and not mandatory to EU states, the EC Proposal's sandbox regulation could create confusion in the market. It could also encourage AI developers to choose EU states with less stringent sandbox regimes, at least for the purposes of regulatory compliance.

The "black box" problem in AI exemplifies one type of high-risk AI that would be ideal for analysis through sandboxing, as the results can be studied in a controlled environment. This is also an example of where sandbox regulation may be a more effective approach than a pure strict liability regime. However, the sandbox can only address the black box problem when the sandbox regulation aims to exempt AI innovation from both regulatory compliance and liability exposure while within the oversight and control of the sandbox regulators. Furthermore, since the black box AI would likely have use and application across the EU, thus its sandbox regulation should likewise be supervised at the EU level rather than through different national approaches and priorities.

[114] ibid, Art 53(1).

[115] ibid, Art 53(5–6).

[116] K Yordanova, "The shifting sands of regulatory sandboxes for AI" (*KU Leuven, Centre for IT&IP Law*, 2019) <https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/> (last accessed 13 December 2020).

## IV. Applying the sandbox approach to AI regulation

### 1. Assessing the benefits of sandboxes: the case of the UK ICO

The UK ICO developed a regulatory sandbox as a service "to support organisations who are creating products and services which utilise personal data in innovative and safe ways".[117] Those participating engage with the sandbox team to benefit from the ICO's expertise and so on. It is worth mentioning that the sandbox is a "free, professional, fully functioning service for organisations, of varying types and sizes, across a number of sectors".[118] There are many benefits to organisations taking part in this sandbox. These include: "1) access to ICO expertise and support; 2) increased confidence in the compliance of your finished product or service; 3) a better understanding of the data protection frameworks and how these affect your business; 4) being seen as accountable and proactive in your approach to data protection, by customers, other organisations and the ICO, leading to increased consumer trust in your organisation; 5) the opportunity to inform future ICO guidance; 6) supporting the UK in its ambition to be an innovative economy; and 7) contributing to the development of products and services that can be shown to be of value to the public".[119] The ICO has specific key areas of focus that are "innovations related to the Age Appropriate Design Code" and "innovations related to data sharing, particularly in the areas of health, central government, finance, higher and further education or law enforcement".[120] The ICO is currently accepting expressions of interest from organisations innovating in these key areas where substantial public benefits can be proven.[121] The importance of this regulatory sandbox is highlighted through the previous participants who benefitted from the regulatory sandbox, which include Future Flow Research, Inc., Onfido Limited, Heathrow Airport Limited, JISC – Wellbeing Code of Practice and Novartis Pharmaceuticals UK Limited.[122]

The ICO along with the Alan Turing Institute established guidance providing "organisations practical advice to help explain the processes, services and decisions delivered or assisted by AI, to the individuals affected by them". This guidance was developed as organisations are using AI for supporting or making decisions concerning individuals.[123] The guidance is necessary given the need to balance legal compliance while realising the benefits of AI.[124] In fact, there is a need to clarify "how to

[117] Information Commissioner's Office, "The Guide to the Sandbox" <https://ico.org.uk/for-organisations/regulatory-sandbox/the-guide-to-the-sandbox/> (last accessed 6 March 2021).

[118] ibid.

[119] ibid.

[120] Information Commissioner's Office, "Key Areas of Focus for the Regulatory Sandbox" <https://ico.org.uk/media/for-organisations/documents/2618112/our-key-areas-of-focus-for-regulatory-sandbox.pdf> (last accessed 6 March 2021).

[121] Information Commissioner's Office, supra, note 117.

[122] Information Commissioner's Office, "Previous Participants" <https://ico.org.uk/for-organisations/regulatory-sandbox/previous-participants/> (last accessed 6 March 2021).

[123] Information Commissioner's Office and Alan Turing Institute, "Explaining Decisions Made with AI" <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/> (last accessed 6 March 2021).

[124] Information Commissioner's Office and Alan Turing Institute, "Part 1: The Basics of Explaining AI" <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/> (last accessed 6 March 2021).

apply data protection provisions associated with explaining AI decisions, as well as highlighting other relevant legal regimes outside the ICO's remit".[125] The guidance supports organisations with the "practicalities of explaining AI-assisted decisions and providing explanations to individuals".[126] The guidance allows organisations to "1) select the appropriate explanation for your sector and use case; 2) choose an appropriately explainable model; and 3) use certain tools to extract explanations from less interpretable models".[127] The actors that will find the guidance useful are technical and compliance teams, among others. The guidance was issued given the government's commitment with regards to AI's Sector Deal.[128] A number of tasks are set up to "design and deploy appropriately explainable AI systems and to assist in providing clarification of the results these systems produce to a range of affected individuals (from operators, implementers, and auditors to decision recipients)".[129] These are: (1) selecting "priority explanations by considering the domain, use case and impact on the individual"; (2) collecting and pre-processing "data in an explanation-aware manner"; (3) building the "system to ensure you are able to extract relevant information for a range of explanation types"; (4) translating the "rationale of your system's results into useable and easily understandable reasons"; (5) preparing "implementers to deploy your AI system"; and (6) considering "how to build and present your explanation".[130] In summary, the guidance "covers the various roles, policies, procedures and documentation that you can put in place to ensure your organisation is set up to provide meaningful explanations to affected individuals".[131] The ICO's shift in its regulatory sandbox to cover organisations looking to use AI highlights the great potential that is seen when it comes to regulating AI through sandbox regulation, especially as this move complies with the UK's commitment to invest in and regulate the AI sector. Although this guidance is still at the nascent stage and is not binding, it sends the message that sandboxes, as a mechanism, can be appropriate for the regulation of AI.

Indeed, several of those who participated in the ICO's regulatory sandbox highlighted a great level of satisfaction. Novartis' participation in the sandbox helped it to understand voice technology and its risks.[132] Similarly, the Greater London

[125] ibid.

[126] Information Commissioner's Office, "Part 2: Explaining AI in Practice" <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-2-explaining-ai-in-practice/> (last accessed 6 March 2021).

[127] ibid.

[128] ibid.

[129] Information Commissioner's Office, "Summary of the Tasks to Undertake" <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-2-explaining-ai-in-practice/summary-of-the-tasks-to-undertake/> (last accessed 6 March 2021).

[130] ibid.

[131] Information Commissioner's Office, "Part 3: What Explaining AI Means for your Organisation" <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-3-what-explaining-ai-means-for-your-organisation/> (last accessed 6 March 2021).

[132] Information Commissioner's Office, "Regulatory Sandbox Final Report: Novartis Pharmaceuticals UK Ltd: A Summary of Novartis' Participation in the ICO's Regulatory Sandbox Beta" (5 February 2021) at 1–16.

Authority was capable of reviewing the processes and documentation associated with its SafeStats data portal, demonstrating to the "public, stakeholders and data-providing organisations that they are cognisant of legal requirements in handling, processing and sharing of personal data; with the relevant and necessary procedures and requirements in place".[133] Moreover, the participation of the Ministry of Housing, Communities and Local Government enabled it to understand how public authorities can conduct complex data-sharing activities while complying with data protection law. This participation emphasised the importance of establishing key dependencies related to "clear legal powers/gateways to use data and upfront commitments from partners to permit the use of the essential data that they hold".[134] Finally, the participation of Tonic Analytics demonstrated their commitment to using innovative technology to provide "new insights and actionable intelligence to improve the way that law enforcement organisations, public sector agencies, local authorities and the private sector can collaboratively tackle crime and safety challenges on the roads in a compliant and secure manner while maintaining individual' [*sic*] rights to data protection and privacy".[135]

## 2. A proposal for regulating AI applications: strict liability complemented by sandboxes

Sandbox regulation offers the possibility of controlling AI development. This would mean applying the concept of FinTech sandbox regulations to AI algorithmic experimentation, which would solve a few issues, namely:

1) It would help avoid the stifling of innovation that may occur in a strict liability regime, since the sandbox would not restrict experimentation in high-risk areas of AI such as black box AI, allowing the technology to be tested within supervised limits to understand its impact on the market and society.
2) The concerns that most EU lawmakers have when it comes to high-risk AI would be heard, in the sense that we would not be freely allowing high-risk AI experimentation, but rather conducting it in a safeguarded and controlled regulatory environment. Risk management of disruptive technologies is one of the main reasons for the existence of a sandbox.
3) It would nevertheless be challenging to design a regulatory environment that is flexible enough to accommodate new changes to markets and that, at the same time, can create regulatory certainty for all market participants.[136]

[133] Information Commissioner's Office, "Regulatory Sandbox Final Report: Greater London Authority: A Summary of Greater London Authority's Participation in the ICO's Regulatory Sandbox Beta" (February 2021) at 20–21.

[134] Information Commissioner's Office, "Regulatory Sandbox Final Report: The Ministry of Housing, Communities and Local Government (MHCLG): A Summary of MHCLG's Participation in the ICO's Regulatory Sandbox Beta" (March 2021) at 13.

[135] Information Commissioner's Office, "Regulatory Sandbox Final Report: Tonic Analytics: A Summary of Tonic Analytics' Participation in the ICO's Regulatory Sandbox Beta" (February 2021) at 20.

[136] W-G Ringe and C Ruof, "Regulating FinTech in the EU: the case for a guided sandbox" (2020) 11 European Journal of Risk Regulation 605.

4) The sandbox would allow for the mitigation of risks and costs posed by a strict liability regime on producers and operators.

Notably, the European Council itself has determined that both regulatory sandboxes and experimentation clauses have practical uses beyond FinTech, and in technological growth generally. On 16 November 2020, the European Council published its Conclusions on Regulatory Sandboxes and Experimentation Clauses as Tools for an Innovation-Friendly, Future-Proof and Resilient Regulatory Framework that Masters Disruptive Challenges in the Digital Age.[137] Through this document, the Council encourages the EC to continue considering the use of experimentation clauses on a case-by-case basis when drafting and reviewing legislation, as well as to evaluate the use of experimentation clauses in *ex post* evaluations and fitness checks on the basis of an exchange of information with Member States. Conclusion Number 8 defines regulatory sandboxes as "concrete frameworks which, by providing a structured context for experimentation, enable where appropriate in a real-world environment the testing of innovative technologies, products, services or approaches – at the moment especially in the context of digitalisation – for a limited time and in a limited part of a sector or area under regulatory supervision ensuring that appropriate safeguards are in place".[138] On the other hand, Conclusion Number 9 defines experimentation clauses as "legal provisions which enable the authorities tasked with implementing and enforcing the legislation to exercise on a case-by-case basis a degree of flexibility in relation to testing innovative technologies, products, services or approaches".[139]

The Council demonstrates its keenness not to waste time in the pursuit of economic competition through technological innovation. It called upon the EC to present the findings of this evaluation, followed by practical recommendations for the possible future use of regulatory sandboxes and experimentation clauses at the EU level. These Conclusions seem consistent with the other two documents previously analysed – the Coordinated Plan on Artificial Intelligence[140] and the Resolution of 12 February 2019 on a comprehensive European industrial policy on AI and robotics (2018/2088(INI))[141] – since it insists on the opportunities offered by regulatory sandboxes.

In these Conclusions, the Council affirms that regulatory sandboxes can offer relevant opportunities, particularly for innovation and growth, and especially for SMEs, micro-enterprises and start-ups. This is consistent with Conclusion Number 2, in which the Council states that, in order for the EU to emerge stronger after the COVID-19 crisis, "the EU regulatory framework needs to be as competitive, effective, efficient, coherent, predictable, innovation-friendly, future-proof, sustainable and resilient as possible. It needs to be evidence-based and has to protect and support both citizens and businesses in the context of the aim of a fully functioning EU Single

[137] European Council, supra, note 60.
[138] ibid.
[139] ibid.
[140] Commission, supra, note 104.
[141] European Parliament, supra, note 107.

Market without imposing new unnecessary burdens and while reducing existing unnecessary burdens".[142] The Council clearly envisions that technological competition is essential to driving growth, and it recognises the need to eliminate barriers and burdens for firms. Nevertheless, the strict liability approach was chosen given the human-centric focus that the EU legislator is emphasising for the adoption of ethical AI and for protecting citizens. Hence, the legislator is seeking to strike a balance between innovation and regulation through strict liability but also sandbox regulations fostering innovation.

Most recently, the EC Proposal laying down harmonised rules on AI (Artificial Intelligence Act) and amending certain Union legislative acts includes provisions on the use of sandboxes for the regulation of AI. Accordingly, such regulatory sandboxes can be established by a single or several Member States or the European Data Protection Supervisor. The regulatory sandbox can only occur if it is supervised and guided by the competent authorities and in accordance with the existing regulations applicable to sandboxes at the EU and national level.[143] An emphasis on the role of national authorities is made, especially on ensuring that the supervisory and corrective powers of the competent authorities are not affected. Moreover, mitigation and even suspension actions must be taken in case of risks to health and safety and to fundamental rights.[144] Other elements related to regulatory sandboxes are also included, such as the "Further processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox"[145] and the provision of "small-scale providers and start-ups with priority access to the AI regulatory sandboxes to the extent that they fulfil the eligibility conditions".[146]

More importantly, the proposal explicitly mentions that participation in a sandbox experiment does not exempt a participant from liability, but rather remains liable under EU and Member State liability legislations.[147] This is very important, as sandbox participation cannot become a shield from liability. The EU's policy approach to AI sandboxing, while focusing on limiting any potential abuses of using the sandbox as a liability shield, could be criticised for eroding the very essence of sandbox regulation. It may be a more prudent approach to create a different set of liability regulations under the sandbox or to exempt certain types of high-risk AI technology from participating in the sandbox. Another option is to create a fault-based liability while participating within the sandbox or to lower the level of risk for experimenting on AI applications within the sandbox. One can argue that participation in a supervised sandbox justifies the lowering of the risk level from a high risk, for example, to that of limited risk. It remains unclear whether participation in the sandbox could be a defence or mitigating factor against strict liability, as conduct in the sandbox can be a factor when imposing fines under the GDPR.

[142] ibid.
[143] EC Proposal, supra, note 15, Art 53(1).
[144] ibid, Art 53(2–3).
[145] ibid, Art 54.
[146] ibid, Art 55(1)(a).
[147] ibid, Art 53(4).

The EU approach could send the wrong message and discourage some producers and operators from participating in an AI sandbox, at the very least due to the uncertainty in liability that could diminish any potential upsides. After all, sandbox participation is not without risks. As examples, it exposes the developer to compliance and setup costs, an added layer of regulatory supervision, and it exposes the AI technology to regulators and third parties. In exchange for these risks, a regulatory sandbox should minimise regulatory constraints, lower the risk of regulatory enforcement and liability and provide ongoing guidance from regulators. Despite all of the efforts made in clarifying high-risk AI liability in the EC Proposal, the EC could have further clarified the interplay between strict liability and the regulatory sandbox. Hence, even though the EU approach seems to focus on striking a balance between innovation and regulation, not creating liability protections or at least clarifying liability protection benefits within the regulatory sandbox creates uncertainty and substantially weakens the adoption of a sandbox approach. Rather, further details on the regulatory and liability benefits to sandbox participant and AI experimentation activities within the sandbox ought to be added. Additionally, the determination of an AI's classification as high risk should be made after a determination of AI experts and sandbox regulators of the results of the AI experimentations.

Strict liability in particular would be more costly to SME producers and operators as a regulatory regime in this context. The EC Proposal recognises this reality when adopting the regulatory sandbox provisions by giving priority access, organising awareness activities and creating dedicated communication channels to small-scale providers under Article 55(1).[148] The Explanatory Memorandum to the EC Proposal (Sections 1.1 and 3.2) explains that the AI regulatory sandbox could be useful for the promotion of AI and aims to reduce the regulatory burden on SMEs and start-ups.[149] The proposed regulatory sandboxes are also obligated to consider the interests of SMEs when setting compliance fees under Article 55(2).[150]

Strict liability transfers the cost of risk knowledge and risk control from the user primarily to producers and operators (whether backend or frontend operators), who are deemed to control the risk.[151] The cost of compliance, development and liability will increase significantly for producers and operators. This allocation of risks and costs will likely result in the stifling of innovation, especially for SMEs that cannot afford the cost of the risks in a strict liability regime. SMEs may also face barriers to entry, for example, for procedures on full quality compliance.[152] These SMEs are already at a disadvantage, according to Bathaee, since AI resources are already concentrated in a few large companies.[153] SMEs that lack significant funds will likely stay away from AI development, especially when taking into account that AI liability

[148] ibid, Art 55(1).
[149] ibid.
[150] ibid, Art 55(2).
[151] Zech, supra, note 29, 4–6.
[152] Renda et al, supra, note 59, 149.
[153] Bathaee, supra, note 26, 930.

total compliance costs account for an estimated 17% of total AI investment costs, although this figure is likely higher for SMEs than large companies due to economies of scale.[154] Furthermore, large companies are in a better position to absorb the costs of a strict liability regime and are better prepared for new regulations.[155] SMEs would be able to offset regulatory compliance costs only by sharing systems,[156] which is ideal in a regulatory sandbox environment. The use of strict liability will only encourage monopolies in the AI industry, which in turn represent a risk to consumers, who will ultimately bear the cost of the strict liability regime as that cost is passed on to them.[157] One could argue that insurance could mitigate the cost of strict liability AI regulation for small businesses.[158] However, insurance likely would not cover all risks, especially those that are uncontrolled and unforeseeable. Most importantly, insurance will not address the stifling of innovation since insurance will not necessarily encourage experimentation, as it can only help with risk tolerance.

To be fair, a sandbox regulatory regime also has costs for regulators, namely the administrative ones of running the sandbox. However, companies participating in the sandbox could offset such costs, which are not as significant and lasting when compared to the costs of stifling innovation and competition. Rather, sandboxing could be a means for addressing the cost imbalance created by a strict liability regime by encouraging innovation despite the strict liability regime.

To sum up, as previously stated, the strict liability approach might restrict innovation in the area of AI in the EU, which strengthens the call herein for supplementing AI development within a regulated sandbox to promote innovation. Indeed, there are various arguments against the use of strict liability rules for the regulation of AI, especially as strict liability is possible in case of individual causation. By assigning only a strict liability rule, the AI developer or operator will need to assess whether the risks exceed the expected benefits.[159] If so, such actors may not invest in the innovation. These actors cannot foresee the behaviour of AI algorithms where several variables play an important role, including databases, big data gathering and the end users themselves.[160] Indeed, this is further worsened by the fact that many parties are involved that are "AI developers; algorithm trainers; data collectors, controllers, and processors; manufacturers of the devices incorporating the AI software; owners of the software (which are not necessarily the developers); and the final users of the devices (and perhaps many more related hands in the pot)".[161] It is in this context there are new calls for new liability approaches such as giving AI systems legal

[154] Renda et al, supra, note 59, 155, 160, 166.

[155] ibid, 160.

[156] ibid.

[157] Bathaee, supra, note 26, 930. See, however, A Lior, "AI strict liability vis-à-vis AI monopolization" (2020) SSRN Electronic Journal (arguing that there is a lack of connection and that insurance can mitigate the cost of strict liability).

[158] ibid.

[159] Zech, supra, note 29, 6.

[160] E Marchisio, "In support of 'no-fault' civil liability rules for artificial intelligence" (2021) 1 SN Social Sciences 54.

[161] I Giuffrida, "Liability for AI decision-making: some legal and ethical considerations" (2019) 88 Fordham Law Review 439, 443.

personhood.[162] The concept of strict liability for assigning fault for harms will become even more obsolete as AIs become more independent.[163]

## V. Conclusion

This paper presented a different view concerning the regulation of high-risk AIs. Rather than accepting the calls for adopting pure strict liability rules, the authors suggest that the use of a sandbox approach at the stage of new high-risk AIs emerging on the market could complement a strict liability regime that could stifle innovation. The main reason for the use of this approach is the fact that a strict liability regime would be difficult and costly to implement, particularly the cost and chilling effect that a strict liability regime would impose on AI innovation. This paper highlighted the attempts made by the EU to regulate AI in the general sense, where attempts are being made for the adoption of strict liability and human-centred rules while simultaneously investigating the option of using sandbox regulation. Hence, in a way, and despite the great efforts made for the adoption of strict liability rules by the EU for high-risk AIs, it seems as though two different approaches are being tried simultaneously with the purpose of figuring out the best way to regulate high-risk AIs. It remains to be seen whether a compromise can be reached in which both approaches are applied.

Regardless of the EU approach that will prevail, the authors opine that applying a sandbox regulation to the AI sector is more appropriate than a pure strict liability regime when it comes to high-risk activities. There is a vital need to create a balance between the regulation of the sector to protect citizens and society while fostering innovation, given the constant and fast developments occurring in the AI field. The authors acknowledge the complexities surrounding the suggestion made, as sandbox regulations are mostly applied in the FinTech sector where there is a great literature on the topic.[164] In contrast, the use of sandboxes to regulate AI requires further studies. Despite these challenges, sandbox regulation remains more appropriate to complement the strict liability regime in the context of high-risk AIs because of the objective of sandbox regulation that greatly differs from the one related to strict liability rules. While strict liability rules focus on assigning strict responsibility and making sure someone is held liable in case of damages, a sandbox regulation is a form of innovation in the legal sphere whereby the objective is to "regulate before regulation even exists".[165] As such, similar to the safe space created in the FinTech sector,[166] a new safe space will be created in the AI sector for trial-and-error approaches with the end goal of adopting a regulation that has been tested and represents the most appropriate one for the regulation of high-risk AIs.

[162] ibid, 444.

[163] BW Jackson, "Artificial intelligence and the fog of innovation: a deep-dive on governance and the liability of autonomous systems" (2019) 35 Santa Clara High Technology Law Journal 35, 56.

[164] See, generally, Ringe and Ruof, supra, note 136, 605; Allen, supra, note 51.

[165] J Kálmán, "*Ex ante* 'regulation'? The legal nature of the regulatory sandboxes or how to 'regulate' before regulation even exists", in G Hulkó and R Vybíral (eds), *European Financial Law in Times of Crisis of the European Union* (Budapest, Dialóg Campus 2019) pp 215–25.

[166] Zetzsche et al, supra, note 36, 31–103; Truby, supra, note 37, 9.

**Acknowledgements.** None.

**Competing interests.** None.

**Funding statement.** This publication was made possible by the NPRP award NPRP11C-1229-170007 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.



**Cite this article:** J Truby, RD Brown, IA Ibrahim and OC Parellada, "A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications". *European Journal of Risk Regulation*. https://doi.org/10.1017/err.2021.52