European Journal of Risk Regulation (2023), 1–40
doi:10.1017/err.2023.72

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Encoding the Enforcement of Safety Standards into Smart Robots to Harness Their Computing Sophistication and Collaborative Potential: A Legal Risk Assessment for European Union Policymakers

Riccardo Vecellio Segate and Angela Daly

School of Science and Engineering and School of Law, University of Dundee, Dundee, UK
Corresponding author: Riccardo Vecellio Segate; Email: rvecelliosegate001@dundee.ac.uk

Abstract

Until robots and humans mostly worked in fast-paced and yet separate environments, occupational health and safety (OHS) rules could address workers' safety largely independently from robotic conduct. This is no longer the case: collaborative robots (cobots) working alongside humans warrant the design of policies ensuring the safety of both humans and robots at once, within shared spaces and upon delivery of cooperative workflows. Within the European Union (EU), the applicable regulatory framework stands at the intersection between international industry standards and legislation at the EU as well as Member State level. Not only do current standards and laws fail to satisfactorily attend to the physical and mental health challenges prompted by human–robot interaction (HRI), but they exhibit important gaps in relation to smart cobots ("SmaCobs") more specifically. In fact, SmaCobs combine the black-box unforeseeability afforded by machine learning with more general HRI-associated risks, towards increasingly complex, mobile and interconnected operational interfaces and production chains. Against this backdrop, based on productivity and health motivations, we urge the encoding of the enforcement of OHS policies directly into SmaCobs. First, SmaCobs could harness the sophistication of quantum computing to adapt a tangled normative architecture in a responsive manner to the contingent needs of each situation. Second, entrusting them with OHS enforcement vis-à-vis both themselves and humans may paradoxically prove safer as well as more cost-effective than for humans to do so. This scenario raises profound legal, ethical and somewhat philosophical concerns around SmaCobs' legal personality, the apportionment of liability and algorithmic explainability. The first systematic proposal to tackle such questions is henceforth formulated. For the EU, we propose that this is achieved through a new binding OHS Regulation aimed at the SmaCobs age.

Keywords: Encoding of safety rules enforcement; quantum computing; smart collaborative robots

> *Robotics and AI have become one of the most prominent technological trends of our century.*[1]

[1] EA Macías, "Legal challenges for robots and autonomous artificial intelligence systems in the healthcare context with special reference to Covid-19 health crisis" (2021) 7(1) Ius and Scientia 119–34, 127.

I. Introduction

Ensuring the safety of human operators in human–robotic interactions is a key requirement for the adoption and implementation of robots in a variety of settings. Robotics currently seeks synergies that can optimise the sustainable circularity of standard robotic outcomes while adding quality inputs to more customised small-scale projects, without discounting renewed calls for promoting ethical robotics to frame the quality of working life within the so-called "Industry 4.0" or "Fourth Industrial Revolution".[2] Against this backdrop, and despite outstanding barriers,[3] robots are starting to replace or accompany humans in performing job tasks that, far from being repetitive and/or standardised only, might even touch upon higher spheres of operational and conceptual complexity – if not yet "creativity".[4] It is no longer the case that machines simply act upon human instruction[5]: as robots become "suitable to act in unstructured scenarios and [interact] with unskilled and undertrained users",[6] the robots and humans might be considered as peers, or it might even be the other way round, with humans acting in accordance with robotic direction[7] – in any case performing tasks working closely together. Think, for instance, of lightweight surgical and caregiving robots, robot-enabled aerospace repairing missions, rehabilitation and recovery robots that help patients overcome their psychological resistance to, for example, walking[8] or automated defence applications where human attendees serve subordinate functions or whose safety and survival anyway depend upon robots whose indications they could not double-check rapidly enough.[9] As a growing market,[10] collaborative robots can be used in different sectors and applications.[11] In manufacturing, a collaborative robot "bridges the gap

[2] See D Ramos et al, "Frontiers in Occupational Health and Safety Management" (2022) 19 International Journal of Environmental Research and Public Health 10759, 4; L Gazzaneo et al, "Designing Smart Operator 4.0 for Human Values: A Value Sensitive Design Approach" (2020) 42 Procedia Manufacturing 219–26, 220; S Vezzoso, "Competition by design" in B Lundqvist and MS Gal (eds), *Competition Law for the Digital Economy* (Cheltenham, Edward Elgar 2019) pp 93–123, 101. Some authors have already started to explore the ethically informed human-centrism of interconnected living that should feature as a core tenet of Industry *5.0* (!); check, for instance, A Adel, "Future of Industry 5.0 in society: human-centric solutions, challenges and prospective research areas" (2022) 11 Journal of Cloud Computing 40.

[3] Refer, eg, to AG Burden et al, "Towards human–robot collaboration in construction: current cobot trends and forecasts" (2022) 6 Construction Robotics 209–20; R Galin and M Mamchenko, "Human–Robot Collaboration in the Society of the Future: A Survey on the Challenges and the Barriers" in PK Singh et al (eds), *Proceedings of the Futuristic Trends in Network and Communication Technologies (FTNCT): Third International Conference held in Taganrog (Russia)* (New York, Springer 2021) pp 111–22, 115–16.

[4] See, most recently, M Javaid et al, "Substantial capabilities of robotics in enhancing industry 4.0 implementation" (2021) 1 Cognitive Robotics 58–75, 62–63; A Castro et al, "Trends of Human–Robot Collaboration in Industry Contexts: Handover, Learning, and Metrics" (2021) 21(12) Sensors 4113.

[5] This seems to be still taken for granted in the literature, though. Refer, eg, to S Bragança et al, "A Brief Overview of the Use of Collaborative Robots in Industry 4.0: Human Role and Safety" in PM Arezes et al (eds), *Occupational and Environmental Safety and Health* (New York, Springer 2019) 641–50, 647.

[6] M Valori et al, "Validating Safety in Human–Robot Collaboration: Standards and New Perspectives" (2021) 10(2) Robotics 1. Cf. A Pauliková et al, "Analysis of the Impact of Human–Cobot Collaborative Manufacturing Implementation on the Occupational Health and Safety and the Quality Requirements" (2021) 18 International Journal of Environmental Research and Public Health 1927, 3.

[7] Read further Y Lai, *Towards Personalised Robotic Assessment and Response during Physical Human Robot Interactions* (2022) PhD thesis in Information Engineering at University of Technology Sydney, 13.

[8] See G Bingjing et al, "Human–Robot Interactive Control Based on Reinforcement Learning for Gait Rehabilitation Training Robot" (2019) 16(2) International Journal of Advanced Robotic Systems 7.

[9] Refer further to Valori et al, supra, note 6, 1–2.

[10] Refer, eg, to Microsoft Dynamics 365, "2019 Manufacturing Trends Report" (2018) <https://info.microsoft.com/rs/157-GQE-382/images/EN-US-CNTNT-Report-2019-Manufacturing-Trends.pdf>9.

[11] See also E Matheson et al, "Human–Robot Collaboration in Manufacturing Applications: A Review" (2019) 8(4) Robotics 100; R Calvo and P Gil; "Evaluation of Collaborative Robot Sustainable Integration in Manufacturing

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

between totally manual and fully automated production lines".[12] Accordingly, working-space seclusion and compartmentalisation are long gone, too: the idea that (massively similar) robots are allocated a siloed working space in which to intensively perform their tasks while humans work in other company environments or departments is falling into obsolescence.

The aim of this paper is to situate the issue of occupational health and safety (OHS) and smart collaborative robotics (cobotics) within the current and prospective European Union (EU) regulatory landscape and to formulate a proposal that could steer meaningful debate among EU policymakers so as to improve the regulatory dialogue between the (existing) robotics and (prospective) artificial intelligence (AI) policy frameworks,[13] with specific reference to the OHS implications of smart cobotics. The current OHS frameworks applicable to robotics within the EU derive from a bundle of Directives as well as from an extensive network of international industry standards understood as quasi-binding. This will be accompanied by the emerging AI regulatory framework in the form of a Regulation (AI Act) and a Directive (AI Liability Directive) – both of which are still drafts at the time of writing. We aim to demonstrate that the dialogue between these two frameworks (robotics OHS on the one hand and AI on the other) is not yet fertile, grounded and sophisticated enough – both terminologically and substantially – to accommodate the specific challenges arising from smart cobots (SmaCobs) in ways which would be helpful to engineers and adequate for the EU's position in a global "regulatory race" in this area vis-à-vis China and other technologically pioneering jurisdictions.[14] Upon inspecting the shortcomings of the current and forthcoming rules, we will advance a proposal for a comprehensive EU Regulation to combine the two fields into a unitary piece of legislation, with the purpose of covering smart cobotics and taking account of possible future developments such as quantum computing (QC) technology. We intend to assist EU policymakers in pursuing a smart "Brussels Effect"[15] in this domain of OHS and SmaCobs and in turn contribute to ongoing global discussions as to how to secure operational safety in AI-driven cobotics through enhanced interdisciplinary integration, informed policy effectiveness and balanced regulatory scrutiny across applicable industries, markets and regions.

Our work is situated against the backdrop of challenges arising from automation as the regulatory chessboard of our time. A remarkable deal of scholarly and policy work has been produced in recent years with regards to *inter alia* robots and taxation, robot-assisted alternative dispute resolution, intellectual property (IP) adaptation to new forms of intangibles, consumer protection from subliminal cues,[16] the relationship between AI-aided neuroscientific inquiry and emotional lawyering,[17] privacy and data protection issues arising from algorithmic interference with one's thoughts, choices, preferences,

Assembly by Using Process Time Savings" (2022) 15 Materials 611, 15; A Dzedzickis et al, "Advanced Applications of Industrial Robotics: New Trends and Possibilities" (2022) 12 Applied Sciences 135.

[12] M Javaid et al, "Significant applications of cobots in the field of manufacturing" (2022) 2 Cognitive Robotics 222–33, 226.

[13] Within the "European legal space", the Council of Europe (CoE), too, has recently stepped into AI regulation, most notably through the efforts reported at <https://www.coe.int/en/web/artificial-intelligence/work-in-progress>. We will not explicitly include these most recent non-EU developments, which might, however, prove impactful in the near future if the interfaces between the Court of Justice of the European Union and the CoE's European Court of Human Rights are strengthened and formalised.

[14] See also L Monsees and D Lambach, "Digital sovereignty, geopolitical imaginaries, and the reproduction of European identity" (2020) 31(3) European Security 377.

[15] Refer to A Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford, Oxford University Press 2020).

[16] See, most recently, RJ Neuwirth, *The EU Artificial Intelligence Act: Regulating Subliminal AI Systems* (London, Routledge 2023).

[17] See R Vecellio Segate, "Navigating Lawyering in the Age of Neuroscience: Why Lawyers Can No Longer Do Without Emotions (Nor Could They Ever)" (2022) 40(1) Nordic Journal of Human Rights 268–83.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

aspirations and behaviour[18] and even the legal personality of robots themselves.[19] Nonetheless, one area that has received very little attention is the regulation of automation vis-à-vis OHS in smart robot-populated workspaces. In fact, as humans are featuring back into the picture, human–robot collaborative environments are in need of a safety overhaul – most crucially in so-called "developing countries"[20] – in order to overcome safety-dependent barriers to wider cobotic adoption.[21] It is this issue that represents the focus of this article: how the EU regulatory framework currently addresses OHS issues in smart cobotic use and the need for a reform of the legislation in this area.

Furthermore, underexplored legal areas concern *inter alia* the application of AI and QC to collaborative robotics – or integration with them. Indeed, robotics is increasingly intertwined with AI applications, towards "a future that will increasingly be invented by humans and [AI systems] in collaboration".[22] Extensive literature does exist on the topic, but almost exclusively from an engineering perspective and mostly focusing on medical surgery[23]: there is some safety discussion within this literature.[24] Yet, no comprehensive *cross-sector* analysis of safety legal–ethical dilemmas raised by *autonomous* robots *working with humans* has ever been accomplished. The present work is not about whether "good protection of workers' health in the performance of their duties using robots, AI"[25] is afforded per se. Rather, it appraises the extent to which safety rules can be encoded into machines, in such a way that workers' safety could be entrusted upon robots regardless of human supervision, especially in contexts of human–robot collaboration (HRC) – not least through embodied (and yet removable) applications. Indeed, HRC has been questioning "the traditional paradigm of physical barriers separating machines and workers", and it was enabled by "the use of multi-modal interfaces for more intuitive, aware and safer human–robot interaction".[26] How should (EU) lawmakers envision the regulation of these collaborations from an OHS perspective? How are we to account for the complexity of "human factors" in collaborative robotics[27] and encode SmaCobs with safety rules that consider all such factors? To what extent does this require a new OHS framework?

[18] See A McStay and LD Urquhart, "'This time with feeling?' Assessing EU data governance implications of out of home appraisal based emotional AI" (2019) 24(10) First Monday.

[19] See B Bennett and A Daly, "Recognising rights for robots: Can we? Will we? Should we?" (2020) 12(1) Law, Innovation and Technology 60–80.

[20] Refer, eg, to S Yang et al, "Robot application and occupational injuries: are robots necessarily safer?" (2022) 147 Safety Science 105623.

[21] Check, eg, N Berx et al, "Examining the Role of Safety in the Low Adoption Rate of Collaborative Robots" (2022) 106 Procedia CIRP 51–57.

[22] A Gerdes, "An Inclusive Ethical Design Perspective for a Flourishing Future with Artificial Intelligent Systems" (2018) 9(4) European Journal of Risk Regulation 677–89, 689.

[23] Illustrative examples are: M Wagner et al, "The importance of machine learning in autonomous actions for surgical decision making" (2022) 2(1) Artificial Intelligence Surgery 64–79; M Runzhuo et al, "Machine learning in the optimization of robotics in the operative field" (2020) 30(6) Current Opinion in Urology 808–16; E Battaglia et al, "Rethinking Autonomous Surgery: Focusing on Enhancement over Autonomy" (2021) 7(4) European Urology Focus 696–705; AA Gumbs et al, "Artificial Intelligence Surgery: How Do We Get to Autonomous Actions in Surgery?" (2021) 21(16) Sensors 5526; Y Kassahun et al, "Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions" (2016) 11 International Journal of Computer Assisted Radiology and Surgery 553–568.

[24] Refer extensively to G Guerra, "Evolving artificial intelligence and robotics in medicine, evolving European law: comparative remarks based on the surgery litigation" (2021) 28(6) Maastricht Journal of European and Comparative Law 805–33, 808–12.

[25] M Jarota, "Artificial intelligence and robotisation in the EU – should we change OHS law?" (2021) 16(18) Journal of Occupational Medicine and Toxicology 3.

[26] Valori et al, supra, note 6, 1. See also ibid, 18.

[27] In this respect, the most comprehensive literature review to date seems to be: M Faccio et al, "Human factors in cobot era: a review of modern production systems features" (2023) 34(1) Journal of Intelligent Manufacturing 85–106.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

To be clear, our aim is not limited to addressing "the integration of safety-related features in the robot architectures and control systems" (which is concerned with solutions to enhance robots' operational safety),[28] but to expound the legal–ethical implications of having safety rules – be they in the form of industry standards, technical protocols, company policies, soft regulations or binding laws – algorithmically *encoded into (and thus enforced by)* robotic machines, as if they were working environments' safety guardians as regards their own conduct,[29] and that of their human collaborators as well. Would it be legal to entrust them with supervisory OHS duties, as is already being experimented with in Europe and around the world?[30] And even so, would it prove ethically sound, in the absence of validation by humans? In a scramble towards robotic morals-free "efficiency", "[w]e are increasingly faced with ethical issues[,] particularly w[ith] self-modifying AI that might change its optimisation goals without humans being able to interfere or even discover that this was the case".[31] This issue stands beyond that of deliberative safety, described as "the capability of the system to optimize speed, degrees of freedom, and safety to correspond to the particular task at hand".[32] It is about robots being programmed to "smartly" take initiative about safety *vis-à-vis both themselves and those humans they collaborate or share a working environment with*, with the purpose of enforcing safety rules adaptively, responsively, sophisticatedly and comprehensively, in a horizontal and yet customised manner.

The managerial turn is hardly new in the robotics literature. Entrusting robots with management tasks has been suggested by scientists for quite a few years already.[33] Yet, far from the mundane responsibility of organising schedules and arranging deliverables, the more demanding *safety* managerialism as applied to robots has not been addressed in the literature so far, even though it has been identified as a need. Indeed, scholars have already noted that "[t]he problem of charging AI-assisted robots with specific OHS duties, and then enforcing such obligations from them, [. . .] requires additional research".[34] It is one thing for managers to identify workplace safety using risk-assessment software, to then decide where to allocate resources and refine strategies; but it is completely another to posit that robots themselves should run such software and make choices based on its appraisal. This work intends to fill this gap.

Considering that robots will increasingly perform better than humans in certain workstations, and taking stock of the distancing "new normal" inaugurated with the

[28] Valori et al, supra, note 6, 1.

[29] To exemplify, ML has been resorted to in order to remotely feed robots with data from which they could rapidly learn how to monitor their system parameters and enhance safety accordingly (eg by isolating the most robust reliability factors or predicting outages and the need of safety interruptions for robots); refer to K Aliev and D Antonelli, "Proposal of a Monitoring System for Collaborative Robots to Predict Outages and to Assess Reliability Factors Exploiting Machine Learning" (2021) 11 Applied Sciences 1621, 17.

[30] Check, eg, J Schepers and S Streukens, "To serve and protect: a typology of service robots and their role in physically safe services" (2022) 33(2) Journal of Service Management 197–209, 203; S Sarker et al, "Robotics and artificial intelligence in healthcare during COVID-19 pandemic: a systematic review" (2021) 146 Robotics and Autonomous Systems 103902, 7.

[31] Gerdes, supra, note 22, 681. Read also R Alkhatib and R Lebdy, "Digital Manufacturing and the Fifth Industrial Revolution" in AD Cheok and TH Musiolik (eds), *Analyzing Future Applications of AI, Sensors, and Robotics in Society* (Hershey, PA, IGI Global 2020) pp 69–86, 83.

[32] A Hanna et al, "Deliberative safety for industrial intelligent human–robot collaboration: regulatory challenges and solutions for taking the next step towards Industry 4.0" (2022) 78 Robotics and Computer-Integrated Manufacturing 102386, 11.

[33] Refer, eg, to V Murashov et al, "Working safely with robot workers: recommendations for the new workplace" (2016) 13(3) Journal of Occupational and Environmental Hygiene 61–71, 66; ME Gladden, "Managerial Robotics: A Model of Sociality and Autonomy for Robots Managing Human Beings and Machines" (2014) 13(3) International Journal of Contemporary Management 67–76.

[34] Jarota, supra, note 25, 6.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

COVID-19 pandemic[35] (and unlikely to completely backtrack, not least subconsciously, especially at certain latitudes), the identification of appropriate rules for robots to be entrusted with OHS procedures and decision-making is a useful and increasingly urgent endeavour. Where should thresholds be set in order to mark the boundaries between human supervisors' liability and automated (thus "autonomous") decision-making? Can the two be reconciled? Should we allow machines *to even reach the point* of taking independent decisions that can impact workers' safety? On the other hand, has this not always already been the case, that robots do make these decisions or are designed in a way that does encode certain normative OHS values, although in less "smart" a way? And were we to accept that machines can independently decide on safety issues, should vicarious liability rules for relevant corporate officers be envisioned? How to apportion them between programmers, managers, owners, shareholders, operators, supervisors and other relevant functions? These, and more, are all interdependent issues we explore in the present article.

This article narrowly addresses choices *on and about* OHS, not those broadly *impacting* OHS, which could encompass virtually any robotic action within a professional environment. At the same time, this work fits with the most general scholarly discourse on legal automation and robotised rule enforcement, which has mostly taken a transactional path inspired by smart contracts and the blockchain, while – it seems to us – neglecting certain fundamental aspects of measure-enforcement automation in the workplace, from a labour law and health law standpoint. Other relevant bodies of literature are: "disaster literature", concerned with the potentially catastrophic consequences of experimenting too ambitiously with robots' ability to transform themselves through self-learning[36]; and the debate on algorithmic governmentality.[37]

From here, we offer some definitions and background on SmaCobs, before turning to entrusting robots with enforcing safety rules. From there, we provide an overview of the current (and forthcoming) regulatory framework on this issue in the EU and set forth our proposal for reform. Overall, we find that the long-standing EU framework exhibited profound gaps when it comes to ensuring OHS in SmaCobs; the most recent binding Regulation in this area,[38] approved by the EU Council in May 2023, did endeavour to systematise the regulatory landscape and bridge some of those gaps, but we will argue that it still lies far from satisfactorily filling those lacunae. We outline certain key points and topics that the Regulation should have covered but failed to do so. Summarised in Fig. 1, we present a contrasted overview of: (1) the traditional, long-standing EU framework applicable to robotics safety; (2) the most recent regulatory effort by the EU in this area (ie the aforementioned Regulation – hereinafter "New Machinery Regulation"); and (3) our arguments as to why the Regulation is still far from achieving

[35] Read, for instance, P Agarwal et al, "Artificial Intelligence Adoption in the Post COVID-19 New-Normal and Role of Smart Technologies in Transforming Business: A Review" (2022) Journal of Science and Technology Policy Management 12–13.

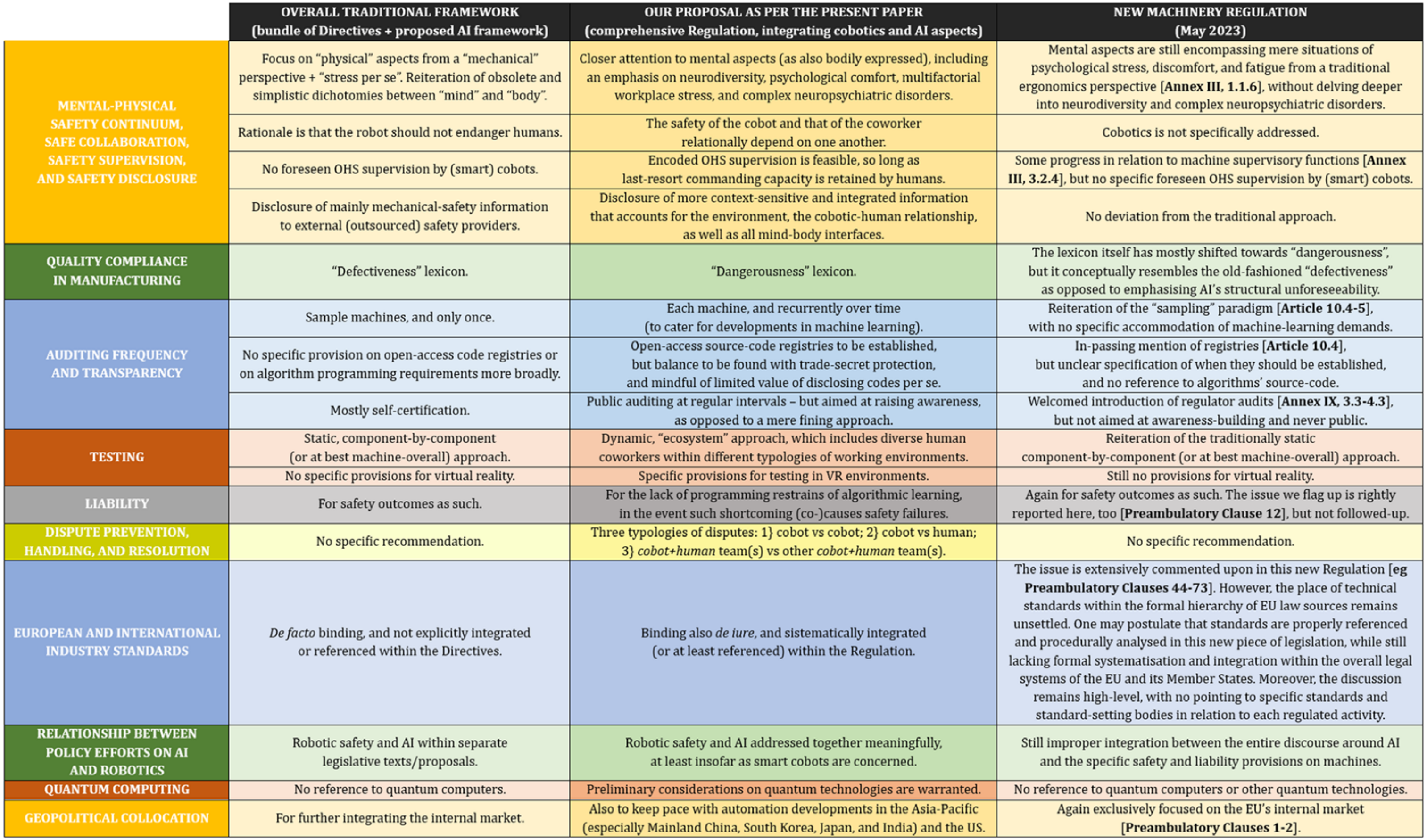
[36] See, eg, S Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (London, Penguin 2019); R White, *Unintended Dystopia* (Eugene, OR, Wipf and Stock 2021); CF Naff, "Approach to the Singularity: The Road to Ruin, or the Path to Salvation?" in AB Pinn (ed.), *Humanism and Technology: Opportunities and Challenges* (London, Palgrave 2016) pp 75–94; RJ Neuwirth, "The 'letter' and the 'spirit' of comparative law in the time of 'artificial intelligence' and other oxymora" (2020) 26(1) Canterbury Law Review 1–31, 3–4; E Yudkowsky, "Artificial intelligence as a positive and negative factor in global risk" in MM Ćirković and N Bostrom (eds), *Global Catastrophic Risks* (Oxford, Oxford University Press 2012) pp 308–45.

[37] See, eg, A Rouvroy, "Governing Without Norms: Algorithmic Governmentality" (2018) 32 Psychoanalytical Notebooks 99; see also S Roberts, "Big Data, Algorithmic Governmentality and the Regulation of Pandemic Risk" (2019) 10 European Journal of Risk Regulation 94.

[38] Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC (Text with EEA relevance), PE/6/2023/REV/1, Document 32023R1230.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press



	OVERALL TRADITIONAL FRAMEWORK (bundle of Directives + proposed AI framework)	OUR PROPOSAL AS PER THE PRESENT PAPER (comprehensive Regulation, integrating cobotics and AI aspects)	NEW MACHINERY REGULATION (May 2023)
MENTAL-PHYSICAL SAFETY CONTINUUM, SAFE COLLABORATION, SAFETY SUPERVISION, AND SAFETY DISCLOSURE	Focus on "physical" aspects from a "mechanical" perspective + "stress per se". Reiteration of obsolete and simplistic dichotomies between "mind" and "body".	Closer attention to mental aspects (as also bodily expressed), including an emphasis on neurodiversity, psychological comfort, multifactorial workplace stress, and complex neuropsychiatric disorders.	Mental aspects are still encompassing mere situations of psychological stress, discomfort, and fatigue from a traditional ergonomics perspective [**Annex III, 1.1.6**], without delving deeper into neurodiversity and complex neuropsychiatric disorders.
	Rationale is that the robot should not endanger humans.	The safety of the cobot and that of the coworker relationally depend on one another.	Cobotics is not specifically addressed.
	No foreseen OHS supervision by (smart) cobots.	Encoded OHS supervision is feasible, so long as last-resort commanding capacity is retained by humans.	Some progress in relation to machine supervisory functions [**Annex III, 3.2.4**], but no specific foreseen OHS supervision by (smart) cobots.
	Disclosure of mainly mechanical-safety information to external (outsourced) safety providers.	Disclosure of more context-sensitive and integrated information that accounts for the environment, the cobotic-human relationship, as well as all mind-body interfaces.	No deviation from the traditional approach.
QUALITY COMPLIANCE IN MANUFACTURING	"Defectiveness" lexicon.	"Dangerousness" lexicon.	The lexicon itself has mostly shifted towards "dangerousness", but it conceptually resembles the old-fashioned "defectiveness" as opposed to emphasising AI's structural unforeseeability.
AUDITING FREQUENCY AND TRANSPARENCY	Sample machines, and only once.	Each machine, and recurrently over time (to cater for developments in machine learning).	Reiteration of the "sampling" paradigm [**Article 10.4-5**], with no specific accommodation of machine-learning demands.
	No specific provision on open-access code registries or on algorithm programming requirements more broadly.	Open-access source-code registries to be established, but balance to be found with trade-secret protection, and mindful of limited value of disclosing codes per se.	In-passing mention of registries [**Article 10.4**], but unclear specification of when they should be established, and no reference to algorithms' source-code.
	Mostly self-certification.	Public auditing at regular intervals – but aimed at raising awareness, as opposed to a mere fining approach.	Welcomed introduction of regulator audits [**Annex IX, 3.3-4.3**], but not aimed at awareness-building and never public.
TESTING	Static, component-by-component (or at best machine-overall) approach.	Dynamic, "ecosystem" approach, which includes diverse human coworkers within different typologies of working environments.	Reiteration of the traditionally static component-by-component (or at best machine-overall) approach.
	No specific provisions for virtual reality.	Specific provisions for testing in VR environments.	Still no provisions for virtual reality.
LIABILITY	For safety outcomes as such.	For the lack of programming restrains of algorithmic learning, in the event such shortcoming (co-)causes safety failures.	Again for safety outcomes as such. The issue we flag up is rightly reported here, too [**Preambulatory Clause 12**], but not followed-up.
DISPUTE PREVENTION, HANDLING, AND RESOLUTION	No specific recommendation.	Three typologies of disputes: 1} cobot vs cobot; 2} cobot vs human; 3} *cobot+human* team(s) vs other *cobot+human* team(s).	No specific recommendation.
EUROPEAN AND INTERNATIONAL INDUSTRY STANDARDS	*De facto* binding, and not explicitly integrated or referenced within the Directives.	Binding also *de iure*, and sistematically integrated (or at least referenced) within the Regulation.	The issue is extensively commented upon in this new Regulation [**eg Preambulatory Clauses 44-73**]. However, the place of technical standards within the formal hierarchy of EU law sources remains unsettled. One may postulate that standards are properly referenced and procedurally analysed in this new piece of legislation, while still lacking formal systematisation and integration within the overall legal systems of the EU and its Member States. Moreover, the discussion remains high-level, with no pointing to specific standards and standard-setting bodies in relation to each regulated activity.
RELATIONSHIP BETWEEN POLICY EFFORTS ON AI AND ROBOTICS	Robotic safety and AI within separate legislative texts/proposals.	Robotic safety and AI addressed together meaningfully, at least insofar as smart cobots are concerned.	Still improper integration between the entire discourse around AI and the specific safety and liability provisions on machines.
QUANTUM COMPUTING	No reference to quantum computers.	Preliminary considerations on quantum technologies are warranted.	No reference to quantum computers or other quantum technologies.
GEOPOLITICAL COLLOCATION	For further integrating the internal market.	Also to keep pace with automation developments in the Asia-Pacific (especially Mainland China, South Korea, Japan, and India) and the US.	Again exclusively focused on the EU's internal market [**Preambulatory Clauses 1-2**].

Figure 1. A comparison between the traditional "bundle" framework, the New Machinery Regulation and our own proposal with regards to the regulatory requirements and policy outlook for smart collaborative robots in Europe. AI = artificial intelligence; cobot = collaborative robot; EU = European Union; OHS = occupational health and safety; VR = virtual reality.

its stated objectives and cannot be satisfactorily applied to operations with *smart collaborative* robots more specifically.

II. Defining robots, collaboration and related safety risks

Before delving into the substance of our argument, it is necessary to make some clarifications and provide definitions. To begin with, what do we mean by safety rules? Those are prescriptions that address potential threats to *inter alia* the physical health of workers due to accidents (eg collisions or undue contact with hazardous materials) involving robots.

As for robots, they may be machines or software, but either way, in this paper we consider only smart (or AI-assisted) robots and particularly SmaCobs. SmaCobs are robots that are both "powered" by AI (and thus endowed with a certain degree of decision-making autonomy through machine learning (ML),[39] such as deep learning's improving by trial and error) and collaborative with humans. Defining "AI" itself is a long-running source of disagreement.[40] ML is generally considered a core characteristic of AI. ML routes and techniques do vary[41] (eg supervised, unsupervised, enhanced, etc.,[42] as well as interactive[43]). Here, we consider all ML paths whereby a robot can autonomously take decisions that are not directly retrievable from instructions and possibly learn from patterns of mistakes – but without necessarily adapting *comprehensively* to each human being they collaborate with, which is plausibly a behavioural ability retained by humans only. In keeping with the European Parliament (EP), intelligent robots are able to exchange positional and sensorial data with the environment, learn from environmental responses to their behavioural patterns and analyse changes in the environment – but all of this without biologically delivering on vital functions.[44] Our working definition broadly subscribes to the EP's formulation, with a focus on the decision-making follow-up to environmental learning, as well as on the collaborative features of SmaCobs.

As for the "collaboration" aspect,[45] it might take a variety of shapes, "spanning from teleoperation to workspace sharing and synergistic co-control".[46] Here we focus on OHS

[39] For a discussion on autonomy in algorithmic learning, see R Vecellio Segate, "Shifting Privacy Rights from the Individual to the Group: A Re-adaptation of Algorithms Regulation to Address the Gestaltian Configuration of Groups" (2022) 8 Loyola University Chicago Journal of Regulatory Compliance 55–114.

[40] Within the EU, refer, eg, to European Commission, "White Paper on Artificial Intelligence: A European approach to excellence and trust" COM(2020) 65 final, 19 February 2020, 16(ftns.45–46). Read more generally MC Buiten, "Towards Intelligent Regulation of Artificial Intelligence" (2019) 10(1) European Journal of Risk Regulation 41–59, 43–45.

[41] Refer, for instance, to HL Janssen, "An approach for a fundamental rights impact assessment to automated decision-making" (2020) 10(1) International Data Privacy Law 76–106, 84; DP Losey et al, "A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction" (2018) 70 Applied Mechanics Reviews 1–19, 9; D Perri et al, "High-performance computing and computational intelligence applications with a multi-chaos perspective" in Y Karaca et al (eds), *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems* (Amsterdam, Elsevier 2022) pp 55–75, 69–70; NM Gomes et al, "Reinforcement Learning for Collaborative Robots Pick-and-Place Applications: A Case Study" (2022) 3 Automation 223–41, 224–27.

[42] Read further at Jarota, supra, note 25, 2.

[43] Check, eg, S Amershi et al, "Power to the People: The Role of Humans in Interactive Machine Learning" (2014) AI Magazine 105–20, 117.

[44] See Jarota, supra, note 25, 2.

[45] On potential distinctions between human–robot "collaboration", "cooperation", "interaction" and so forth, check, eg, S El Zaatari et al, "Cobot programming for collaborative industrial tasks: an overview" (2019) 116 Robotics and Autonomous Systems 162–80, 163.

[46] Valori et al, supra, note 6, 2.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

procedures over whose compliance a robot could have tangible control. These might also encompass aspects of quality control and repairing of indoor environments as well as optimisation of production chains so as to minimise human distraction (and risks emanating therefrom) caused by over-reiteration of elementary tasks.

There are a myriad of threats and risks to health and safety involving robots, which exhibit several causes, sources and effects on the health and safety of humans. Here we offer an overview of them. The precise incidence of such risks and threats will depend upon various factors, including how the cyber-physical robotic system is set up, what other technologies it may integrate and the kind of interaction it pursues with humans in a particular workplace.

One cause of accidents may be due to substantive miscommunication between robots about their "peers" or about humans[47]: for example, human intervention is often assumed as corrective, but for it to be received as indeed correction rather than noise, the right parameters for robots to decode human intentions must be in place.[48] Miscommunication might equally occur between humans about robots, or between robots and humans[49] (individually or group-wise[50]), coming from time-related, space-related or goal-related misunderstandings,[51] but also from, for example, the misdetection, misappreciation or misinterpretation of humans' chronic or extemporary pain.[52] While very experimental for now, the deployment of brain–machine (or machine-mediated brain-to-brain) interfaces not only for remote control as today,[53] but also to encode rules via direct transfer, might represent a miscommunication-intensive risk factor.

Other incident causes may well involve concurrent action over shared working tools (fallacious instrumental interchangeability) or technical dysfunction of the machine on either the software or hardware side – and most frequently both. On the software side, one may face programming errors, including those leading to miscalculations, misplaced manoeuvring schemes, and mis-calibrated human-detection or real-time data-capture systems, but also misconfigured learning from assumed-safe test-aimed digital twins,[54] overdemanding interfaces to other subsystems, dependencies upon operating modes, ineffective response times and suboptimal selection of objects' model flows within the applicable environment. On the hardware side, frequent causes of concern are defective appliances, including sensors as well as measuring and

[47] Refer, eg, to P Jansen et al, "D4.4: Ethical Analysis of AI and Robotics Technologies" (2019) A deliverable by WP4: "AI & robotics – Ethical, legal and social analysis" for the EU's H2020-funded "SIENNA project – Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact", 143.

[48] Read, eg, DP Losey et al, "Physical interaction as communication: learning robot objectives online from human corrections" (2022) 41(1) The International Journal of Robotics Research 20–44, 32–33.

[49] For an exemplification, check A Tutt, "An FDA for Algorithms" (2017) 69(1) Administrative Law Review 83–123, 102.

[50] See further AC Simões et al, "Designing human–robot collaboration (HRC) workspaces in industrial settings: a systematic literature review" (2022) 62 Journal of Manufacturing Systems 28–43, 38.

[51] Consult the proposed taxonomy at M Askarpour et al, "Formal model of human erroneous behavior for safety analysis in collaborative robotics" (2019) 57 Robotics and Computer-Integrated Manufacturing 465–76, 468.

[52] See also S Heydaryan, *Human-Robot Collaboration in Automotive Industry* (2018) PhD thesis in Mechanical Engineering at Politecnico di Torino, pp 100–01; R Behrens et al, "A Statistical Model to Determine Biomechanical Limits for Physically Safe Interactions with Collaborative Robots" (2021) 8 Frontiers in Robotics and AI 667818.

[53] Refer, eg, to AO Onososen and I Musonda, "Research focus for construction robotics and human–robot teams towards resilience in construction: scientometric review" (2023) 21(2) Journal of Engineering, Design and Technology 502–26.

[54] Check, eg, M Gleirscher et al, "Challenges in the Safety-Security Co-Assurance of Collaborative Industrial Robots" in MIA Ferreira and SR Fletcher (eds), *The 21st Century Industrial Robot: When Tools Become Collaborators* (New York, Springer 2022) pp 191–214, 200–01. Interestingly, it was suggested to "display a cobot's 'thought process' and plan real time in a digital-twin user-interface[, so as to] help the operator implicitly understand the perception abilities of the cobot and its limitation[, thus eventually] increase[ing] the industry's confidence in non-traditional cobot programs, i.e. those incorporating probabilistic command outcomes" – El Zaatari et al, supra, note 45, 177.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

protective components and force calibration tools. There is, however, a software aspect to this, too: the harmful action taken by cobots as a result of misleading interaction and poor integration between sensing/awareness/perception algorithms and self-learning OHS-encoding algorithms. This is all the more relevant today "given the recent rise in interest in sensing technology for improving operational safety by tracking parameters such as *mental* workload, situation awareness, and fatigue".[55] The process through which multiple algorithms are combined into one single algorithmic outcome through opaque interactions remains high on the legal and regulatory agenda.[56] Still in the hardware domain, corrupted or unexpectedly encrypted data, or partly missing/altered/inaccessible data due to data breaches[57] (including breached cloud-stored data)[58] and cyber-misappropriation,[59] which may also be accomplished through encryption-disabling QCs themselves,[60] are further risks. They may implicate misreadings in machine-to-machine interfaces, and they are primed or aggravated by a wide range of external circumstances, including poor Internet connectivity, environmental noise, toxic fumes, non-standard heat or radiation, power outages, inappropriate lighting, misleading and/or contradictory and/or asynchronous feedback and damaged environmental sensors. All these "physical" factors – be they hardware- or software-disrupting – might be exacerbated by remoteness and virtuality; for example, by interaction through immersive virtual reality (VR) and augmented reality (AR)[61] interfaces, due to both representational inaccuracies in the VR/AR itself and sensorial detachment that leads to accidents in the real world – without discounting cognitive interferences as well.[62] Defective or inappropriate

[55] A Marinescu et al, "The future of manufacturing: utopia or dystopia?" (2023) 33(2) *Human Factors and Ergonomics in Manufacturing & Service Industries* 184–200, 12, emphasis added, in-text references omitted.

[56] See also R Seyfert, "Algorithms as regulatory objects" (2022) 25(11) Information, Communication & Society 1542–58, 1545–46.

[57] Refer further to J-PA Yaacoub et al, "Robotics cyber security: vulnerabilities, attacks, countermeasures, and recommendations" (2022) 21 International Journal of Information Security 115–58.

[58] See E Fosch-Villaronga and T Mahler, "Cybersecurity, safety and robots: strengthening the link between cybersecurity and safety in the context of care robots" (2021) 41 Computer Law & Security Review 105528, 2. For an enlightening piece on "cloud robotics", read E Fosch-Villaronga and C Millard, "Cloud robotics law and regulation: challenges in the governance of complex and dynamic cyber–physical ecosystems" (2019) 119 Robotics and Autonomous Systems 77–91.

[59] Check, eg, F Maggi et al, "Rogue Robots: Testing the Limits of an Industrial Robot's Security" (2017) Trend Micro and Politecnico di Milano <https://documents.trendmicro.com/assets/wp/wp-industrial-robot-security.pdf> 6.

[60] For an exemplificatory case study on the "geopolitics" of the QC encryption debate, check M Wimmer and TG Moraes, "Quantum Computing, Digital Constitutionalism, and the Right to Encryption: Perspectives from Brazil" (2022) 1 Digital Society 12; read also D Heywood, "Quantum computing – the biggest threat to data privacy or the future of cybersecurity?" (2022) Taylor Wessing <https://www.taylorwessing.com/en/interface/2022/quantum-computing—the-next-really-big-thing/quantum-computing—the-biggest-threat-to-data-privacy-or-the-future-of-cybersecurity>; C Mangla et al, "Mitigating 5G security challenges for next-gen industry using quantum computing" (2023) 35(6) Journal of King Saud University – Computer and Information Sciences 101334.

[61] VR and AR are merged into extended reality (XR) or "hybridised" into mixed reality (MR), which are the comprehensive definitions adopted, eg, by the European Commission.

[62] A wealth of engineering and ergonomics literature exists on these matters, mostly from China or other East Asian jurisdictions (eg Japan and South Korea), where such frontier technology applications are already becoming daily routine across a number of industries, but also – more limitedly – from the USA and Canada. Check in particular this piece that warns that "[e]nsuring the safety of the user while they are interacting with the robot in a virtual reality environment is also a challenge": Y Lei et al, "Virtual reality in human–robot interaction: challenges and benefits" (2023) 31(5) Electronic Research Archive 2374–408, 2398. Refer more generally to M Dianatfar et al, "Review on existing VR/AR solutions in human–robot collaboration" (2021) 97 Procedia CIRP 407–11; O Liu et al, "Understanding Human–Robot Interaction in Virtual Reality" (2017) <https://graphics.cs.wisc.edu/Papers/2017/LRMG17/roman-vr-2017.pdf>; T Inamura and Y Mizuchi, "SIGVerse: A Cloud-Based VR Platform for Research on Multimodal Human–Robot Interaction" (2021) 8 Frontiers in Robotics and AI 549360; R Suzuki et al, "Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human–Robot Interaction and Robotic Interfaces"

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

external protections in working spaces may also contribute to risks turning into actual harm.

Risks to workers' mental health can also arise. One overarching cause of mental harm to workers is identified with their interaction with robots,[63] expressed, for instance, as frustration at robots' non-sentience and inability to "understand each other" and accommodate mood change. However, there are also risks from the converse, when robots are too sentient: alienation and loneliness may surface when robots understand and adapt so well that they are prematurely "humanised" but later fail to deliver the same responses of a human for more complex tasks or in emotional terms.[64] Human workers may be left with a misplaced sense of attachment, responsiveness, recognition, solidarity, affinity and trust,[65] all the way up to betrayed feelings of emotional safety and even intellectual comfort. Robots with reassuring, facially expressive, human-friendly appearances (typical of "humanoids", "androids", humanised chimeras, "cyborgs", etc.) may well mislead inexpert users into acting as if robots were committed to being sympathetic and trustworthy.[66] Confirming the importance of these debates, the European Commission (EC) has recently advised that "[e]xplicit obligations for producers could be considered also in respect of mental safety risks of users when appropriate (ex. *collaboration with humanoid robots*)",[67] and that "[e]xplicit obligations for producers of, among others, AI humanoid robots to explicitly consider the *immaterial harm* their products could cause to users [. . .] could be considered for the scope of relevant EU legislation".[68]

Addressing mental aspects of human–cobot interaction is key to ensuring their successful adoption. Indeed, the subjective experiences of fear and disorientation triggered by robots' AI-powered apparent humanisation risk offsetting the benefits of adopting SmaCobs for OHS oversight and enforcement. To draw on a (perhaps rudimentary, but fit-for-purpose) taxonomy, one may postulate that humans tend to require and seek *dispositional*, *situational* and *learnt* trust before productively engaging with robots in the workplace.[69] While situational trust obviously varies with the circumstances, dispositional trust is "installed" in the human

(2022) 553 CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems 1–33. On the whole, while XR is usually valued as a safer testing platform (motor validation, error detection, response awareness, etc.) for cobotics compared to real-world HRI/HRC simulations, we should be mindful that: (1) environmental stressors might be less intense/lasting virtually and thus might insufficiently prepare the user to accommodate incremental environmental challenges; (2) distances and other factors and metrics are not always accurately reproducible in immersive spaces; and (3) the XR user might become alienated from their actual surroundings to such an extent that even ordinarily innocuous objects in the *non-virtual* reality might turn into a safety risk (stumbling on a chair, falling down, clashing against a wall, leaving dangerous substances open and so forth).

[63] See, eg, AL Abeliansky and M Beulmann, "Are they coming for us? Industrial robots and the mental health of workers" (2018) CEGE Discussion Papers, No. 379, Center for European Governance and Economic Development Research at Georg-August-Universität Göttingen; BP Soentjens, *Examining the Effect of Perceived Stress on Fear of Autonomous Robots and Artificial Intelligence* (2021) BSc thesis in Cognitive Science and AI at Tilburg University.

[64] Read, for instance, E Broadbent et al, "Mental Schemas of Robots as More Human-Like Are Associated with Higher Blood Pressure and Negative Emotions in a Human–Robot Interaction" (2011) 3 International Journal of Social Robotics 291–97; A Borboni et al, "EEG-Based Empathic Safe Cobot" (2022) 10(8) Machines 603, 21.

[65] Refer extensively to AM Aroyo et al, "Overtrusting robots: Setting a research agenda to mitigate overtrust in automation" (2021) 12 Paladyn, Journal of Behavioral Robotics 423–36.

[66] See also Valori et al, supra, note 6, 6. Cf. E Broadbent, "Interactions with Robots: The Truths We Reveal About Ourselves" (2017) 68 Annual Review of Psychology 627–52.

[67] European Commission, supra, note 40, 15, emphasis added.

[68] Report from the European Commission to the European Parliament, the Council and the European Economic and Social Committee on the safety and liability implications of Artificial Intelligence, the Internet of Things and Robotics, COM(2020) 64 final, 19 February 2020, 8, emphasis added.

[69] Read JR Crawford et al, "Mental Health, Trust, and Robots: Towards Understanding How Mental Health Mediates Human-Automated System Trust and Reliance" in H Ayaz (ed.), *Advances in Neuroergonomics and Cognitive Engineering: Proceedings of the AHFE 2019 International Conference on Neuroergonomics and Cognitive Engineering and the AHFE International Conference on Industrial Cognitive Ergonomics and Engineering Psychology* (New York, Springer 2020) pp 119–28, 122.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

individual and resistant to change. Enforcing robot-encoded rules is thus a transient workaround to ensuring a level of trust in robots. Meaningful and fundamental change can only be attained through humans' "learnt trust" stemming from long-term effectiveness in SmaCobs' "care" and management of health risks, including common mental health disorders such as the anxiety–depression spiral. Self-evident as it may sound, "individuals who ha[ve] learned about robots in positive contexts, such as safety-critical applications, [exhibit] higher levels of positivity toward the technology",[70] particularly with ageing,[71] and this seems yet another good reason to win human coworkers' trust. This is especially the case when the human's trust is mediated by psychiatrically diagnosed conditions: over time, the trust acquired with "familiar" robots *will likely enhance the dispositional trust* towards robotic applications more broadly (we may call this a dynamic *trust transferability* variable). Trust is easier to gain – and lasts longer, disseminating to new workers as well – when robotic ethics resonates with human ethical precepts, and it is here that most trust-enhancing attempts fail miserably. The hindrance is ethically uninformed robot decision-making – ethics can only be encoded into emotionless robots up to a certain point (it might reflect programmers' moral horizon at a specific point in time but will never take a life on its own),[72] whereas it develops gradually and spontaneously in (most) healthy humans living socially, and its basic tenets are at least in theory widely shared even if they moderately adjust over one's lifetime. Hence, while the legal lexicon of liability can be employed with regards to robots as well, ethical guidance on what a "right" action is morality-wise should be conceived for the human side of the collaboration only.[73]

Continuing with the mental health discussion, certain aspects relate to the broader sociology and politics of HRC more than they impact the individual worker per se. Concern arises with the robotised monitoring of work performance, the chilling surveillance of workers, hyper-scheduled micromanagement and data analytics[74] – worse still when performed supposedly for OHS purposes. All of this occurs within the datafication of health policing and tracking in the workplace and the increasingly pervasive algorithmic management of welfare, wellness and well-being.[75] Against this "attention economy" backdrop, cognitive overload, too, threatens human workers in their attempt to situate themselves between human and robotic expectations, demands, practices and objectives.[76] Somewhat linked to this, quality competition with and the race to perfectionism against robots (either in the absolute or with *certain* robots against *some other* human–robot working units, even within the same department) are two more issues for today's policymakers. What the quest towards neoliberal "maximisation" entails for workers is an erroneous sense that no

[70] Marinescu et al, supra, note 55, 2.

[71] Check C Rossato et al, "Facing with Collaborative Robots: The Subjective Experience in Senior and Younger Workers" (2021) 24(5) Cyberpsychology, Behavior, and Social Networking 349–56, 354.

[72] Check generally T Walsh, *Machines Behaving Badly: The Morality of AI* (Melbourne, La Trobe University Press 2022). Read also Y Razmetaeva and N Satokhina, "AI-Based Decisions and Disappearance of Law" (2022) 16(2) Masaryk University Journal of Law and Technology 241–67, 253. Cf. S Umbrello, "Beneficial Artificial Intelligence Coordination by Means of a Value Sensitive Design Approach" (2019) 3(5) Big Data and Cognitive Computing 2.

[73] See M Constantinescu et al, "Understanding responsibility in Responsible AI: dianoetic virtues and the hard problem of context" (2021) 23 Ethics and Information Technology 803–14, 809.

[74] Refer further to PV Moore, "Artificial Intelligence: Occupational Safety and Health and the Future of Work" (2019) A Paper prepared for the European Agency for Safety and Health at Work (EU-OSHA), 9. See also J Adams-Prassl, *Humans as a Service: The Promise and Peril of Work in the Gig Economy* (Oxford, Oxford University Press 2019); A Aloisi, "Regulating Algorithmic Management at Work in the European Union: Data Protection, Non-Discrimination and Collective Rights" (2024) 40(1) 39 International Journal of Comparative Labour Law and Industrial Relations 37–70, 7–8.

[75] See, eg, C-F Chung et al, "Finding the Right Fit: Understanding Health Tracking in Workplace Wellness Programs" in G Mark and S Fussell (eds), *CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, New York Association for Computing Machinery 2017) pp 4875–86.

[76] See extensively A Alessio et al, "Multicriteria task classification in human-robot collaborative assembly through fuzzy inference" (2022) Journal of Intelligent Manufacturing 2.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

final objective or potential end exists on the "efficiency" side.[77] An exhausting pace of working, mechanisation, sense of alienation[78] (and inattentiveness stemming therefrom), isolation, exclusion, disconnection and interchangeability (replaceability, permutability, anonymousness) are all manifestations or consequences of such a misbelief. For instance, "psychosocial risk factors [arise] if people are driven to work at a cobot's pace (rather than the cobot working at a person's pace)",[79] worsened by rhythmic noise pollution and other discomforting, interfering and/or distressing factors. Conversely, disappointment may be unleashed at robots whose efficiency has been severely constrained. Indeed, unlike traditional industrial robots, cobots' performance is "continually in a state of becoming [. . . ,] in conflict with the planned day-to-day efficiencies and predictability of automated production systems".[80] They might be just as efficient, but performance will occasionally drop, readjust over time and in any event convey a disrupted sense of absolute reliability efficiency-wise.[81] On another note, neoliberal ideas of "disruptive innovation"[82] guiding SmaCobs' implementation may result in negative social repercussions from robots' takeover of certain tasks (ie uncertainty, precarity, unpredictability, contractual casualisation, skill obsolescence, deskilling,[83] tax injustice,[84] etc.), which need to be accounted for, scrutinised and avoided.

III. Entrusting robots with enforcing safety requirement for human benefit

Despite the aforementioned health and safety risks that robots may cause, "[w]hen working with robots equipped with artificial intelligence, the topic of health protection has not yet been comprehensively described in the literature".[85] While in automation engineering and industrial occupational psychology the "main clusters of interest [are] contact avoidance, contact detection and mitigation, physical ergonomics, cognitive and organizational ergonomics",[86] regulatory scholarship on health and safety has failed to keep pace with developments related to the challenges of ML in redefining what a cobot "is" and what it "does" *over time* and *through experience* – that is, *upon learning*. This calls into question the apportionment of liabilities, vis-à-vis doubts as to the possibility to assess robots' behaviour *ex ante* and design accurate *legal* mitigations of labour risks.

Notwithstanding this paucity of literature, we are not the first to identify challenges with the existing (EU) framework:

[77] Check, eg, A Arslan et al, "Artificial intelligence and human workers interaction at team level: a conceptual assessment of the challenges and potential HRM strategies" (2022) 43(1) International Journal of Manpower 75–88, 82.

[78] See also K Vredenburgh, "Freedom at Work: Understanding, Alienation, and the AI-Driven Workplace" (2022) 52(1) Canadian Journal of Philosophy 78–92.

[79] Moore, supra, note 74, 9.

[80] J Wallace, "Getting collaborative robots to work: a study of ethics emerging during the implementation of cobots" (2021) 12 Paladyn, Journal of Behavioral Robotics 299–309, 302.

[81] Check also Jansen et al, supra, note 47, 92.

[82] Read, eg, J Levich, "Disrupting global health: The Gates Foundation and the vaccine business" in R Parker and J García (eds), *Routledge Handbook on the Politics of Global Health* (London, Routledge 2019) pp 207–18, 207.

[83] Refer, eg, to Y Saint James Aquino et al, "Utopia versus dystopia: professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills" (2023) 169 International Journal of Medical Informatics 104903.

[84] Refer also to R Vecellio Segate, *The Distributive Surveillant Contract: Reforming "Surveillance Capitalism through Taxation" into a Legal Teleology of Global Economic Justice* (2022) Talent Program PhD thesis in International Law at the Department of Global Legal Studies of the University of Macau <https://library2.um.edu.mo/etheses/991010238079006306_ft.pdf> 625(ftn.1719).

[85] Jarota, supra, note 25, 3.

[86] Valori et al, supra, note 6, 3.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

> [g]iven the new psychological and physical threats related to the use of AI robots, it is necessary to expand the EU legislation with general guidelines for the use of intelligent robots in the work environment. Indeed, such robots must be defined in the applicable legal framework. Employers should also define, as part of their internal regulations, the *procedures for employee communication with artificial intelligence*, and relevantly update their training in the OHS area.[87]

Other authors have explored how to apportion liability in the event of robot failures, most often in the medical context.[88] In other words, those papers are concerned with liability arising from robots' failure to act in conformity to *external* guidance or expectations (ie explicit commands activated to deal with specific situations and not activated or deactivated *autonomously* by the robot beyond very elementary and straightforward options that are arguably not characterisable as "smart").

Here, instead, we are concerned with robotic failures to assess, validate, instruct, maintain and act upon safety rules *whose "enforcement" has been entrusted to them through design and coding*. We will thus ponder whether robots can be sophisticated enough to do this, and who exercises responsibility for ensuring that they are duly designed and encoded in this way. Another debate that is already explored in the literature is whether robots can account for their own safety.[89] Here, however, we will focus on the safety of *human* workers exclusively – though unsafe robots are unlikely to protect robot co-workers and make them safe.

If we do seek to design and encode SmaCobs with the ability to enforce OHS rules, several issues arise.

One issue is the extent to which robots can "understand" us, especially our mental health: can awareness of mental states of dysfunction or well-being be algorithmically encoded into SmaCobs for the SmaCob to recognise them as they manifest? Progress might be made towards robotic ability capturing biochemical and signalling dysfunctions subsumed under somatisation symptoms, but robots will likely remain unable to interpret the experiential components that factor into complex mental disabilities. Worse still, prospected risks may negatively impact rather than mitigate the emergence or relapse of such symptoms; this is because robots are unlikely to alleviate disorders that they themselves contribute to engendering (eg anxiety) or whose shades and contaminations they struggle to appreciate; therefore, it can be expected that their response will be depersonalised and stereotyped. Yet again, though, are humans any better in treating and addressing these conditions from an occupational health perspective?

A related issue, where perceptions play a key role, concerns whether the robot itself could be the one to decide whether safety measures are correctly implemented and rules/procedures properly satisfied; in other words, whether it could serve not only as an enforcer, but also as the equivalent of an internal auditor or health inspector. In the affirmative, how would human–robot or robot–robot conflicts of views on the

[87] Jarota, supra, note 25, 1, emphasis added.

[88] See KR Ludvigsen and S Nagaraja, "Dissecting liabilities in adversarial surgical robot failures: a national (Danish) and European law perspective" (2022) second-draft paper available at <https://doi.org/10.48550/arXiv.2008.07381>.

[89] See, for instance, J Tørresen, "A Review of Future and Ethical Perspectives of Robotics and AI" (2018) 4(75) Frontiers in Robotics and AI 6. Importantly, "[s]afety cares about the robot not harming the environment (or humans) whereas security deals with the opposite, aims to ensure the environment does not conflict with the robot's programmed behavior. There's an intrinsic connection between safety and security" – VM Vilches, "Safety requires security in robotics" (2021) Cybersecurity Robotics <https://cybersecurityrobotics.net/safety-and-security-standards-for-robots/>. Those robots that are capable of moving in complex environments and detecting and reporting anomalies that could make them (and their operations as a consequence) unsafe are indeed called "autonomous security robots".

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

appropriateness of certain enforcement outcomes/strategies be handled – for instance, in the event of humans maintaining they are being treated unfairly (eg overstrictly or overdemandingly), or of multiple robots entrusted with similar/complementary functions that respond to the environment slightly differently despite similar encoding? Should we arrange workstations so to ensure that workers are matched with "their" machines in the long run so as to accommodate mutual long-term customisation and encourage adaptation?[90] Notably, the feelings of belonging and "expertise" thereby created do not need to (and would probably never) be scientifically tested: it is about cognitive adjustment to recurring action-schemes, the creation of interactional habits whose safeguarding would plausibly enhance trust and reduce friction while delivering on production targets. And if things do not work out eventually, could a mirrored "right to repair" be exercised by robots themselves? Could robots self-repair, or would they be required to cease operations to have a human intervene in the conflict?

Conflicts are conceivably more frequent if gestures, behaviours, physical expressions, commands, directions and individuals entirely are misallocated into boxes that will then function as addressees of group-customised sets of actions by cobots. If a worker adhered to a recurrent behavioural pattern and the robot categorises them within a certain category for rule-enforcement purposes, but the worker one day refrains from walking the same path (for any mental, physical and/or environmental reason) and such deviation causes OHS incidents, who is liable for them? We may look to engineers, for them to just "find a way" to adjust said robot's categorisations. But the conundrum remains that *we are unaware* of what those boxes look like and who lies therein. Even assuming we could know what the categories were, we would still lack cognisance of how (and why?) robots opted for them (along robotic pathways towards information accumulation, at some point a cumulation effect manifests) over alternative options, and with that we would ignore what elements from the relevant pattern of behaviour truly contributed to its overall machine assessment or to what extent/percentage. Even upon guessing, we would be left with granularity of decision-making that does not return the majority of shades ("weighing options") plausibly composed by the algorithm to formulate its sorting-into-boxes response.[91]

Let us refrain from further viability comments for now and instead turn to matters of purpose. Besides the feasibility issue, *why should we* encode smart robots with safety rules, for them to enforce those rules on our behalf? To begin with, having robots enforce rules and supervise their application on our behalf seems important because "human oversight may be used detrimentally to assign guilt and responsibility to humans"[92] even when the issue lies with machines' unpredictability. And while additional potential reasons abound, which would require more comprehensive research beyond the limits of this paper, we shall outline a few of them here. One reason is that in extensive collaborative environments, the reverse (humans enforcing safety rules on robots) might paradoxically prove more challenging logistically, as well as time-consuming and economically inconvenient. Human response might prove slower and will definitely exhibit elements of instinctiveness that make it suboptimal as either over-responsive (and thus costly due to overcautious work interruption) or under-responsive (and

[90] Read further S Bankins and P Formosa, "When AI meets PC: Exploring the implications of workplace social robots and a human–robot psychological contract" (2020) 29(2) European Journal of Work and Organizational Psychology 215–29.

[91] Read also M Brkan and G Bonnet, "Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas" (2019) 11(1) European Journal of Risk Regulation 18–50, 29, 37, 40, 49; A Deeks, "The judicial demand for explainable artificial intelligence" (2019) 119(7) Columbia Law Review 1829–50, 1837; Y Bathaee, "The artificial intelligence black box and the failure of intent and causation" (2018) 31(2) Harvard Journal of Law & Technology 889–938, 899–917, 926.

[92] R Koulu, "Human Control over Automation: EU Policy and AI Ethics" (2019) 12(1) European Journal of Legal Studies 9–46, 22.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

thus costly given the human and non-human losses potentially materialising). Economic considerations aside, an even more compelling and forward-looking reason draws from QC, and smart robots' forecasted ability to be powered by it:

> Imagine the progress of AI if the power of Quantum Computing was readily available with potential processing power a million times more powerful than today's classical computers with each manufacturer striving to reach not just Quantum Supremacy but far and beyond this revolutionary breakthrough. Humanity can harness this amazing technology with AI and [brain–computer interfaces (BCIs),] generating a new technology revolution [. . .]. The possibility for super strength and super intelligent humans to match the huge advancement in AI intelligence is attainable. [. . . H]umans and AI will strive for Artificial General Intelligence (AGI). This potential for Humans and AI to grow and develop together is staggering but the question of Security, Regulations and Ethics alongside AI and future human BCI advancements highlights the need for standard security and regulations to be put in place.[93]

Indeed, in endorsing the idea that "[a] legal–ethical framework for quantum technology should build on existing rules and requirements for AI",[94] including relevant risk awareness from algorithmic regulation,[95] we will explain in the paragraphs below that the QC revolution matters for smart cobotics in all possible respects: operationally, conceptually, logistically, energetically and even intellectually. The key takeaway here, though, is that *QC may well matter safety-wise as well*. If successfully implemented, this innovation could revolutionise the accuracy, adaptation and contextual responsiveness of OHS monitoring to the actual needs of each and every specific task and operator, be they humans or machines – whose cooperative work has never been as intertwined. Indeed, "[i]f one wanted to say what the law is in a particular domain[, . . . q]uantum computing will enhance the accuracy of the [ML] model by locating *the relevant optimal mix* of values and weights".[96] From a conceptual standpoint, this links to complex-systems physics, in that it networks the complexity of large-scale human–non-human group behavioural responses to hazardous events into mathematical formulas, to then model machine responses based

[93] A Miller, "The intrinsically linked future for human and Artificial Intelligence interaction" (2019) 6(38) Journal of Big Data 2.

[94] M Kop, "Establishing a Legal–Ethical Framework for Quantum Technology" (2021) Blog of the Yale Journal of Law & Technology <https://yjolt.org/blog/establishing-legal-ethical-framework-quantum-technology>. Read further E Perrier, "The Quantum Governance Stack: Models of Governance for Quantum Information Technologies" (2022) 1(22) Digital Society 37–38; M Kop, "Abundance and Equality" (2022) 7 Frontiers in Research Metrics and Analytics 977684, 13, 16–17, 20–21; M Forti, "Un quadro giuridico per il calcolo quantistico: un obiettivo possibile?" (2021) CyberLaws <https://www.cyberlaws.it/en/2021/quadro-giuridico-calcolo-quantistico/>; T Ménissier, "L'éthique des technologies quantiques : une formulation problématique" (2022) Conférence pour les QuantAlps Days à l'Institut de Philosophie de Grenoble <https://shs.hal.science/halshs-03796477/> 10; M Krelina and J Altmann, "Quantum Technologies – A New Field That Needs Assessment" (2022) 95(3) Die Friedens-Warte – Journal of International Peace and Organization 267–89; BC Stahl, "From computer ethics and the ethics of AI towards an ethics of digital ecosystems" (2022) 2(1) AI and Ethics 65–77, 72–74; L Rainie et al, "Could a quantum leap someday aid ethical AI?" (2021) Pew Research Center <https://www.pewresearch.org/internet/2021/06/16/4-could-a-quantum-leap-someday-aid-ethical-ai/>.

[95] See, eg, V Krishnamurthy, "Quantum technology and human rights: an agenda for collaboration" (2022) 7(4) Quantum Science and Technology 044003, 4.

[96] J Atik and V Jeutner, "Quantum computing and computational law" (2021) 13(2) Law, Innovation and Technology 302–24, 321–22, emphasis added. This is particularly helpful in common law jurisdictions such as England and Wales, where judgments' precedential value needs to be factored in as well, adding to the complexity of timely adapting vast tangles of soft laws, hard laws, corporate guidelines, industry standards and case law to the variety of on-the-ground safety challenges in large collaborative environments. From a slightly different angle, see also BJ Evans, "Rules for robots, and why medical AI breaks them" (2023) 10(1) Journal of Law and the Biosciences 1–35, 2–3:

thereupon by successive approximations.[97] This advantage may be mentally pictured through the concept of quantum entanglement: a physical state where each particle can be neither located nor described independently from all others.[98] When it comes to technology-intensive safety policing, our complex legal systems may start to resemble just as intricate tangles, and QC-powered SmaCobs could enjoy an edge in sophisticatedly interpreting such complexity in real time.

Trusting scientists' own enthusiastic forecasts on QC and AI,[99] we foresee QC-powered SmaCobs as the most powerful expression of autonomic computing, whereby "the ability to monitor system condition and provide real-time feedback"[100] is exponentially scaled up compared to non-quantum devices: precisely what is needed in automated/encoded safety enforcement! Needless to say, QC is not going to solve or circumvent the black-box conundrum – if anything, it will worsen it. Yet, on balance, the risk–benefit analysis might make it competitive over non-QC but still black-boxed solutions: the output, still unexplainable, will be exponentially more sophisticated, responsive and adaptive, catering for the interactional complexity of safety needs in contemporary (and future) collaborative workstations. Enhanced computed skillsets will be deployed not only to identify the optimal "regulatory cocktail" to apply to specific contingencies as they arise, as posited above, but also to enhance bodily detection and health monitoring capacity.[101] Even more profoundly, because quantum bits (abbreviated as "qubits") are the paradigmatic expression of entities whose probabilistically determined coordinates[102] are altered as soon as they are measured,[103] from a philosophical (but soon also practical) perspective they convey (and indeed capture) the idea that collaborative systems are those in which cobots and humans, by observing and "measuring" one another, alter each other's spatiotemporal coordinates and action-outcomes. What could be better than QC-powered AI-driven cobots one day appreciating and making the most of these precision-

> Complexity and diversity are hard for humans. The human mind can only balance several factors – perhaps two to five – simultaneously. There is an inherent "mismatch between the mathematical optimization in high dimensionality, characteristic of machine learning, and the [more limited] demands of human scale reasoning". Machine brains can weigh thousands of factors at once and cope with diversity and nuance unemotionally. If nuanced, context-appropriate rules are what humans need to survive against smart machines, then humans may lose [. . .].

[97] On the potential deployment of quantum computing to scrutinise complex phenomena in systems physics (or phenomena in complex systems physics, depending on how one prefers to term it), refer also to Entropy's introductory description of its Special Issue on "Quantum Computing for Complex Dynamics". On the physics of complexity and the reduction of high-order behaviours into formulas, check, eg, F Battiston et al, "The physics of higher-order interactions in complex systems" (2021) 17 Nature Physics 1093–98. On the importance of this for HRI, read, for instance, M San Miguel, "Frontiers in Complex Systems" (2023) 1 Frontiers in Complex Systems 1080801; I Giardina, "Collective behavior in animal groups: theoretical models and empirical studies" (2008) 2(4) HFSP Journal 205–19, 206.

[98] Read, eg, Perri et al, supra, note 41, 62.

[99] Refer, for instance, to GD Paparo et al, "Quantum Speedup for Active Learning Agents" (2014) 4 Physical Review X 031002; M Sweeney and C Gauthier, "Quantum Delegation" in G Viggiano (ed.), *Convergence: Artificial Intelligence and Quantum Computing* (New York, Wiley 2023) pp 11–22, 12.

[100] SS Gill et al, "AI for Next Generation Computing: Emerging Trends and Future Directions" (2022) 19 Internet of Things 100514, 3.

[101] Refer also to CJ Hoofnagle and SL Garfinkel, *Law and Policy for the Quantum Age* (Cambridge, Cambridge University Press 2021) p 361. For instance, bodily detection systems should be engineered sophisticatedly enough to dispel common assumptions on how "standards" or "normal" (ie "within range") human bodies would be; in fact, such assumptions fail to accommodate diversity and especially disability.

[102] See also R Girasa and GJ Scalabrini, *Regulation of Innovative Technologies: Blockchain, Artificial Intelligence and Quantum Computing* (London, Palgrave 2022) p 7.

[103] Refer to MF Crommie and UV Vazirani, "Lecture Notes: An Introduction to Quantum Computation" (2019) teaching notes for a course on Quantum Mechanics at UC Berkeley Quantum Computation Center (BQIC) <https://people.eecs.berkeley.edu/~vazirani/f19quantum/notes/191.pdf> 7.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

shaping perceptive geometries to collaborate *safely* on multiple and entangled cognitive and physical planes? Scholars in different domains expect similarly momentous ramifications; for example, psychologists and quantum physicists expect QC-powered smart robots to capture the complexity of humans' emotional shades and transpose it into quantifiable, encodable oscillations,[104] factually transfiguring the communication between human beings and computers and recasting it at a superior level of seamlessness and sophistication. The reverse is just as true: however encoded and artificial, robots too have "emotions" to communicate to their human counterparts,[105] and QC is premised to aid their "decoding" remarkably.[106] Other researchers have gone so far as to hypothesise quantum creativity,[107] whose implications are of course unparalleled.

On balance, while staying aware of the unjustified hype surrounding QC as a commercial technology *of general use*,[108] we can, however, foresee that, specifically within professional applications in cobotics, the incipient quantum ML revolution[109] indeed holds a reasonable degree of potential to radically enhance the convenience of having robots supervise OHS rules over both themselves and humans – rather than the other way round. This will be even more the case if the Penrose–Hameroff hypothesis, suggesting that human consciousness arises from quantum phenomena such as entanglement and superimposition,[110] will ever be empirically validated.

IV. The current EU regulatory framework relevant to SmaCobs

We have identified the debates and explained both their significance and what is probably awaiting ahead of us in terms of technological development. Yet how is the regulatory

[104] See, eg, G Gayathri et al, "Conjectural schema using quantum mechanics-AI to express and interpret emotional intellect in a social robot" (2021) 2115 Journal of Physics: Conference Series 012040.

[105] Read, eg, HY Ling and EA Björling, "Sharing Stress with a Robot: What Would a Robot Say?" (2020) 1 Human–Machine Communication 133–59.

[106] Refer extensively to J Viertel, *Quantum Computing for Designing Behavioral Model and Quantum Machine Learning on a Humanoid Robot* (2020) MEng thesis in Mechatronics at the University of Applied Sciences Technikum Wien.

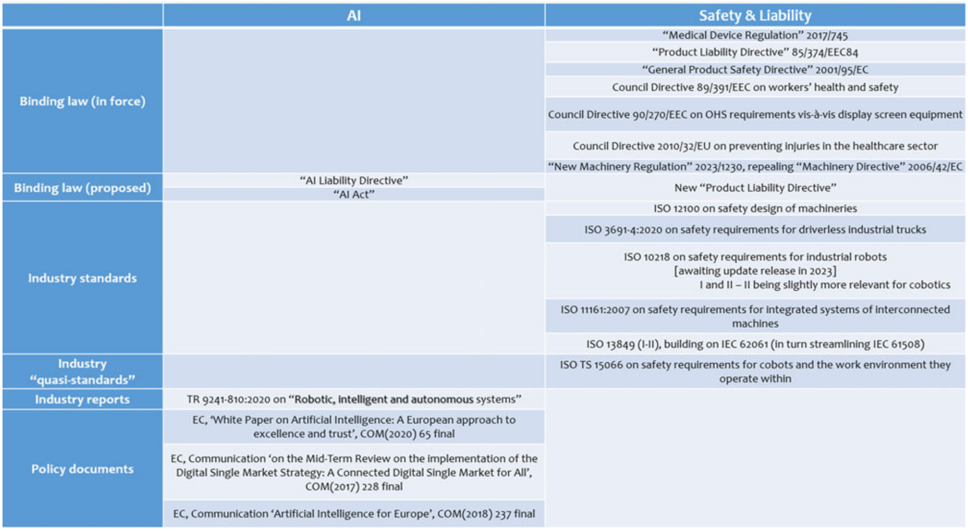
[107] See MC Mannone et al, "Does Creativity Help Us Survive? A Possible Approach with Quantum-Driven Robots" (2022) Proceedings of the CREAI-CEUR Workshop on Artificial Intelligence and Creativity in Udine (Italy) <https://ceur-ws.org/Vol-3278/short2.pdf>

[108] Read extensively E de Jong, "Own the Unknown: An Anticipatory Approach to Prepare Society for the Quantum Age" (2022) 1 Digital Society 15; O Ezratty, "Mitigating the quantum hype" (2022) preprint available at <https://doi.org/10.48550/arXiv.2202.01925>; J Vogt, "Where is the human got to go? Artificial intelligence, machine learning, big data, digitalisation, and human–robot interaction in Industry 4.0 and 5.0" (2021) 36 AI & Society 1083–87, 1085.

[109] Explore further at M Schuld and F Petruccione, *Machine Learning with Quantum Computers* (2nd edition, New York, Springer 2021); G Acampora, "Quantum machine intelligence: launching the first journal in the area of quantum artificial intelligence" (2019) 1 Quantum Machine Intelligence 1–3; N Mishra et al, "Quantum Machine Learning: A Review and Current Status" in N Sharma et al (eds), *Data Management, Analytics and Innovation Proceedings of ICDMAI 2020, Volume 2* (New York, Springer 2021) pp 101–45; S Bhattacharyya et al (eds), *Quantum Machine Learning* (Berlin, De Gruyter 2020); K Najafi et al, "The Development of Quantum Machine Learning" (2022) Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/cgmjzm3c/release/3>; O Ayoade et al, "Artificial Intelligence Computing at the Quantum Level" (2022) 7(3) Data 28; F Valdez and P Melin, "A review on quantum computing and deep learning algorithms and their applications" (2023) 27 Soft Computing 13217–36; EH Houssein, "Machine learning in the quantum realm: the state-of-the-art, challenges, and future vision" (2022) 194 Expert Systems with Applications 116512.

[110] For an introductory and fairly accurate short piece on the matter, read C de Morais Smith, "Can consciousness be explained by quantum physics? My research takes us a step closer to finding out" (2021) The Conversation <https://theconversation.com/can-consciousness-be-explained-by-quantum-physics-my-research-takes-us-a-step-closer-to-finding-out-164582>. For a recent scientific account of quantum consciousness, see CM Kerskens and DL Pérez, "Experimental indications of non-classical brain functions" (2022) 6(10) Journal of Physics Communications 105001.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press



	AI	Safety & Liability
Binding law (in force)		"Medical Device Regulation" 2017/745 "Product Liability Directive" 85/374/EEC84 "General Product Safety Directive" 2001/95/EC Council Directive 89/391/EEC on workers' health and safety Council Directive 90/270/EEC on OHS requirements vis-à-vis display screen equipment Council Directive 2010/32/EU on preventing injuries in the healthcare sector "New Machinery Regulation" 2023/1230, repealing "Machinery Directive" 2006/42/EC
Binding law (proposed)	"AI Liability Directive" "AI Act"	New "Product Liability Directive"
Industry standards		ISO 12100 on safety design of machineries ISO 3691-4:2020 on safety requirements for driverless industrial trucks ISO 10218 on safety requirements for industrial robots [awaiting update release in 2023] I and II – II being slightly more relevant for cobotics ISO 11161:2007 on safety requirements for integrated systems of interconnected machines ISO 13849 (I-II), building on IEC 62061 (in turn streamlining IEC 61508)
Industry "quasi-standards"		ISO TS 15066 on safety requirements for cobots and the work environment they operate within
Industry reports	TR 9241-810:2020 on "Robotic, intelligent and autonomous systems"	
Policy documents	EC, 'White Paper on Artificial Intelligence: A European approach to excellence and trust', COM(2020) 65 final EC, Communication 'on the Mid-Term Review on the implementation of the Digital Single Market Strategy: A Connected Digital Single Market for All', COM(2017) 228 final EC, Communication 'Artificial Intelligence for Europe', COM(2018) 237 final	

Figure 2. The current legal–policy framework on safety, liability and artificial intelligence (AI) as potentially applicable to smart collaborative robots under European Union (EU) law. EC = European Commission; OHS = occupational health and safety.

framework in the EU currently catering for such a complexity of techno-policy inputs and scenarios? We are going to offer a review of the state of the art hereinafter, and we have devised Fig. 2 to assist the reader in navigating it.

The main regulatory documents applicable to robotics and/or AI are included in Fig. 2. A few of them, like the Medical Device Regulation (EU 2017/745), do not warrant further elaboration here as they do not help elucidate the specific challenges raised by SmaCobs. From here we consider the Product Liability Directive, the Machinery Directive, the Framework Directive, the Product Safety Directive and selected policies and technical standards in this area. At the time of writing, the Machinery *Directive* has been just repealed by the aforementioned "New Machinery *Regulation*", but we will nonetheless analyse the Directive to support our argument that the transition from the "old-school" safety framework, of which the Directive was part, to the much welcome but still somewhat "static" regulatory mode of the Regulation insufficiently caters for the necessities of smart cobotics. Later in the paper, we will further emphasise the reasons why we contend that the Regulation, which definitely represents progress on certain issues, should be further refined in important ways.

1. Product Liability Directive

Council Directive 85/374/EEC[111] (Product Liability Directive) is relevant for SmaCobs *as products* rather than as manufacturers or production supervisors. The preambulatory recital clauses[112] mention that

[111] Council Directive 85/374/EEC of 25 July 1985 "on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products", Document 01985L0374-19990604 (consolidated text), available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A01985L0374-19990604>.

[112] Recitals' legal value is notoriously disputed under EU law. Doctrine assumes they are not binding, but EU regulators generously entrust them with meta-normative functions that are integral to the binding text, to such an extent that interdependence and even complementarity between the two are often established. Read further

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

> liability without fault on the part of the producer is the sole means of adequately solving the problem, peculiar to our age of increasing technicality, of a fair apportionment of the risks inherent in modern technological production [... although] the contributory negligence of the injured person may be taken into account to reduce or disallow such liability[.]

Since we are scrutinising *collaborative* robots, this is especially applicable here. The preamble moves on to specify that "the defectiveness of the product should be determined by reference not to its fitness for use but to the lack of the safety which the *public at large* is entitled to expect".[113] The obvious dilemma here concerns who counts as the general public and how it could elaborate an informed opinion based on reasonable expectations if the technicality of cobot-applied ML is somehow obscure even to the brightest engineering minds. This leads to the third consideration we can make based on this Directive. The just-mentioned lack of foreseeability might induce one to simplistically conclude that whatever wrong OHS decision SmaCobs adopt based on ML, the manufacturer would not be exempted from liability solely because ML makes SmaCobs' actions unforeseeable. Indeed,

> the possibility offered to a producer to free himself from liability if he proves that the state of scientific and technical knowledge at the time when he put the product into circulation was not such as to enable the existence of a defect to be discovered may be felt in certain Member States to restrict unduly the protection of the consumer; whereas it should therefore be possible for a Member State to maintain in its legislation or to provide by new legislation that this exonerating circumstance is not admitted[.][114]

Nevertheless, this clause refers to *unknowable defects* rather than to intrinsic technological limitations whose perimeter may still change over time but to which manufacturers are already accustomed. One may thus observe that if the clause does not discount liability for unknown defects, it would accordingly not solicit exceptions in the event of incident-causing foreseeable technology limitations. And yet, these choices pertain to the realm of policy and probably fall well beyond the manufacturing stage, meaning that it is probably policymakers who are those in charge of such high-level choices as to whether as a society we should accept the inherent unforeseeability in ML. From a legal perspective, the difference is that this is unforeseeability *by design*, meaning that it does not depend on subsequent technology development (eg new hazard discoveries and techno-scientific unreadiness at "the time when the product was put into circulation"[115]) as was typical of the pre-SmaCobs era. Instead, the unforeseeability in SmaCobs is inherent in the design of the machines themselves, whose behaviour is *by definition* – albeit within certain boundaries – unforeseeable to programmers when they designed and encoded the SmaCobs at the outset. In light of this, the entire rationale for liability clauses warrants rethinking so as to approximate them to the needs of a ML-powered future,[116] shifting at least part of the burden onto those regulators that assess technological products' fitness for marketisation.

M den Heijer et al, "On the Use and Misuse of Recitals in European Union Law" (2019) Amsterdam Law School Research Paper No. 2019-31.

[113] Emphasis added.

[114] See also Art 7(e).

[115] Art 6(1)(c).

[116] See also J Schuett, "Risk Management in the Artificial Intelligence Act" (2023) European Journal of Risk Regulation, 10.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

Analogous rethinking is due with regards to the causality nexus, whereby current legislation provides that "[t]he injured person shall be required to prove the damage, the defect and the causal relationship between defect and damage".[117] In fact, consider inappropriate situational evaluations by SmaCobs, leading to an underestimation of health hazards that eventually materialise and harm workers: how can the causal relationship between ML-powered decisional outcomes and workers' harm be ascertained and demonstrated? If the process leading to that cannot be explained (beyond common sense),[118] neither can it be satisfactorily proven.[119] Should we accept court proceedings grounded in *prima facie* cases for harm? The EU is attempting to address some of these issues in its draft AI Liability Directive:

> when AI is interposed between the act or omission of a person and the damage, *the specific characteristics of certain AI systems, such as opacity, autonomous behaviour and complexity*, may make it excessively difficult, if not impossible, for the injured person to meet [the standard causality-grounded] burden of proof. In particular, it may be excessively difficult to prove that a specific input for which the potentially liable person is responsible had caused a specific AI system output that led to the damage at stake. In such cases, the level of redress afforded by national civil liability rules may be lower than in cases where technologies other than AI are involved in causing damage. Such compensation gaps may contribute to a lower level of societal acceptance of AI and trust in AI-enabled products and services. To reap the economic and societal benefits of AI [. . .], it is necessary to adapt in a targeted manner certain national civil liability rules to those specific characteristics of certain AI systems [. . .], by ensuring that victims of damage caused with the involvement of AI have the same effective compensation as victims of damage caused by other technologies.[120]

[117] Art 4.

[118] This is referred to as lack of "retrospective explicability". See, eg, Buiten, supra, note 40, 50.

[119] And indeed, by analogy from the data protection realm,

> [h]umans are not seeking for a model but rather for contrastive explanations (eg not why a decision D was made, but rather why D was made instead of another decision Q) as well as causal attribution instead of likelihood-based explanations. [. . . C]ausality is also one of the main characteristics of the GDPR-based right to explanation because it seems that *it is precisely the understanding of causality that would enable the data subject to effectively challenge the grounds for an algorithmic decision*. However, while symbolic and Bayesian artificial intelligence proposed many approaches to model and reason on causality, the recent advance in machine learning [. . .] is more about learning correlations.
>
> – Brkan and Bonnet, supra, note 91, 28, emphasis added.

The excerpt above poignantly rebuts the argument – made in another strand of literature – that policymakers and especially scholars would be biased against robotic (compared to human) explainability; according to them, double standards would be at play insofar as we expect robots to be "transparent" while human action itself is not – see, eg, J Zerilli et al, "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" (2019) 32 Philosophy & Technology 661–83. It is true that the consequentialism and rationality underpinning human action can never be taken for granted, but each human is expected to reflect upon their reasons to act and to try to illustrate them through language, whereas the problem with robots is that we endeavour to extract truth from them without realising that the pursuit of "humanly relevant" explanations is else from grasping the procedure subsumed under an algorithm's outcome. Put differently, it is not just a matter of depth and quantity: human-sought and robotic-produced "explanations" are *qualitatively* uneven; in other words, resting on alternative balances between explanation ("why an outcome is such") and interpretation ("what relates the outcome to potential causes") for both single events and patterns. Furthermore, the law is structured in such a manner that even when explanation cannot be found with regards to human action, the action-taker (or someone else on their behalf) will take accountability for such an action's outcomes, pursuant to applicable chains of command and liability doctrines.

[120] EC, Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM(2022) 496 final, 28 September 2022, Preambulatory clauses (3)–(5), emphasis added.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

Of course, this debate extends well beyond the EU; across the Atlantic, for instance, US and Canadian scholars are facing exactly the same shortcomings[121]: "the law as currently established may be useful for determining liability for mechanical defects, but not for errors resulting from the autonomous robot's 'thinking'".[122] Should we expand liability boundaries so to address *dangerous* (as opposed to "defective") algorithms?[123] But are most algorithms not intrinsically dangerous? As cobots become smarter, the mechanics-intensive language of "defectiveness" falls deeper and deeper into obsolescence[124]: we advise a paradigm shift towards "dangerousness", with the caveat that any ML-driven cobot is and will remain *to some extent* dangerous. Indeed, as distinct from defectiveness, which can be "fixed", dangerousness can only be "lowered" and "coped with", but hardly "negated", both operationally and conceptionally: it is inherent to the very nature of machines that are allowed some room to learn by themselves and take autonomous decisions stemming from that learning. If dangerousness cannot be dispelled, we can at least learn how to best adapt to (and survive) its potential effects – and *this* should be the regulatory focus in smart cobotics. Implications for the industry are broad, including as regards the redefinition of insurance schemes.[125]

2. General Product Safety Directive

The General Product Safety Directive (2001/95/EC), too, may concern robots as products rather than product manufacturers. Of interest here, it provides that

> [i]nformation available to the authorities of the Member States or the Commission relating to risks to consumer health and safety posed by products shall *in general* be available to the public, in accordance with the requirements of transparency and without prejudice to the restrictions required *for monitoring and investigation activities*.[126]

Leaving aside the question of how broadly the "in general" caveat may be construed, it should be noted that restrictions apply exclusively in the interests of auditors and investigators, while trade secret ownership seems to be of no concern. If "risks to consumer health and safety" reside in the source code of cobots' AI, such code would warrant disclosure as inherently risky for consumers (ie in this case, those who purchase the robots to then feed them with relevant normative-regulatory-administrative data and redeploy them as "OHS invigilators" within production chains). "Professional secrecy" is cited as a further ground for exceptions,[127] but it merely engrains one of the sub-classes of

[121] Refer, eg, to Benny Chan, "Applying a Common Enterprise Theory of Liability to Clinical AI Systems" (2021) 47(4) American Journal of Law & Medicine 351–85, 359–60; Guerra, supra, note 24, 815–28.

[122] W Barfield, "Liability for Autonomous and Artificially Intelligent Robots" (2018) 9(1) Paladyn, Journal of Behavioral Robotics 193–203, 196.

[123] See A Beckers and G Teubner, *Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds* (London, Bloomsbury 2022) p 75.

[124] Read also M Buiten, "Product Liability for Defective AI" (2023) available on SSRN at <https://ssrn.com/abstract=4515202>; M Buiten et al, "The law and economics of AI liability" (2023) 48 Computer Law & Security Review 105794; TR de las Heras Ballell, "The revision of the product liability directive: a key piece in the artificial intelligence liability puzzle" (2023) 24 ERA Forum 247–59, 255.

[125] "Particularly with respect to 'emergent behaviours' of robotics, [. . .] the fundamental query for the insurers is how they will be placed to charge fair premiums where robots act in ways not even predictable for their programmers and trainers. This type of concern about smart robots [. . .] may gradually become an area of concern in respect of predictability of robot actions [that] would substantially make it difficult for insurance markets to offer affordable premiums" – A Bugra, "Room for Compulsory Product Liability Insurance in the European Union for Smart Robots? Reflections on the Compelling Challenges" in P Marano and K Noussia (eds), *InsurTech: A Legal and Regulatory View* (New York, Springer 2020) pp 167–98, 177.

[126] Art 16(1), emphases added.

[127] Art 16.2.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

commercially protected secrets. Moreover, disclosing algorithms' coding provides unserviceable information when it comes to tracking the entire spectrum of their potential operational outcomes. Knowing the code does little to enhance our knowledge about code-enabled learning patterns and results; disclosure just results in cosmetic, formalistic transparency, which may even be counterproductive in that it might instil a false sense of confidence in robotically encoded safety. Disclosing robot algorithms' source code may infringe copyright and violate trade secrets protection while proving of only superficial reassurance safety-wise.

Another relevant provision contains the obligation of professional upgrade and "lifelong learning". Indeed, it is provided that

> [w]here producers and distributors know *or ought to know*, on the basis of the information in their possession *and as professionals*, that a product that they have placed on the market poses risks to the consumer that are incompatible with the general safety requirement, they shall immediately inform the competent authorities of the Member States [. . .].[128]

This formulation encapsulates a cogent example of *ex post* obligation that is dependent on technology developments and baseline professionalism, as such compelling manufacturers to remain accountable for algorithmically caused harms. However, conceptual complexity intervenes here: from a technical perspective only (hence, focusing on the engineering side without entering the realm of theories of legal liability), to what extent are ML failures the fault of an algorithm as it was coded rather than of the data it was fed with? Answering this question is far from trivial. To exemplify, if ML-powered harm is believed to be mainly produced out of "bad data", then manufacturers are exempted from notifying the authorities of the risks that algorithmic learning might elicit based on deeper understandings of the inner workings of AI.

The third and final point we ought to highlight from the General Product Safety Directive regards the obsolescence of the "sampling" criterion for market authorisation: it stipulates that producers are expected to assess risks *inter alia* by "sample testing of marketed products",[129] which is irrelevant when it comes to algorithmic learning, which entails that every machine behaves uniquely. One should also bear in mind that testing ML-powered cobots in realistic conditions with humans can form part of a vicious circle, as it proves riskier (and thus regulatorily harder) precisely due to algorithmic unforeseeability, which means that the machines that most need testing (ie the smart and collaborative ones) are those that are going to be tested the least – or in the least realistic, human-participated and use-proximate conditions.[130] The debate is thus warranted on meaningful alternatives to or additional safeguards for both human participation and behavioural sampling.[131] Subtracting humans from the testing equation seems unadvisable:

[128] Art 5.3, emphases added.

[129] Art 5.1(b*bis*).

[130] See also S Bermúdez i Badia, "Virtual Reality for Safe Testing and Development in Collaborative Robotics: Challenges and Perspectives" (2022) 11 Electronics 1726, 4; A Scibilia et al, "Analysis of Interlaboratory Safety Related Tests in Power and Force Limited Collaborative Robots" (2021) 9 IEEE Access 80880.

[131] Some initial solutions engineering-wise start being offered; refer, for instance, to D Araiza-Illan et al, "Systematic and Realistic Testing in Simulation of Control Code for Robots in Collaborative Human-Robot Interactions" in L Alboul et al (eds), *TAROS 2016: Towards Autonomous Robotic Systems* (New York, Springer 2016) pp 20–32; P Maurice et al, "Human-oriented design of collaborative robots" (2017) 57(1) International Journal of Industrial Ergonomics 88–102.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

> cognition is said to be "situated". When applied to the example of collaborative robots, risk cannot only be understood from the techno-centric perspective of mere energy containment in terms of managing speed, force, and separation.[132]

At the same time, awareness must be raised of unsafe testing that could subject humans to unpredictable (and indeed untested) ML-prompted cobotic action. It is true that "difficulties [surface] in identifying the moment in time during a robot's trajectory where a specific algorithm is the least secure, requiring simulation or testing with the entire system",[133] and this is obviously going to prove resource-depleting. Yet, in the long run, it might prevent safety accidents and thus, overall, contribute to optimising costs, starting with insurance premiums.

3. Framework Directive

Council Directive 89/391/EEC (Framework Directive)[134] sets out general prevention-orientated principles on the health and safety of workers (at work). It exhibits little specificity on most matters of interest here. As with all other Directives, it is worth considering how it has been implemented in practice by Member States (MSs) and readapted to cater for the aforementioned challenges. Paradoxically, flawed or even poor implementation records[135] might turn into an opportunity for "surgical" normative reforms at the MS level without rediscussing and revising the entire EU OHS framework. Of some relevance to us here, this Directive holds that legal persons, too, may be employers,[136] which calls into question the long-standing and rather complex debate on granting legal personality to (advanced) robots.[137] Even in the affirmative, could robots take on the role as employers of other robots or even humans – thus becoming responsible for their health and safety, including civil compensation and indemnities? To borrow from the public international law (PIL) lexicon, they might even be charged directly with obligations of conduct or of result. According to the United Nations Educational, Scientific and Cultural Organization (UNESCO):

> the agency of robots is not as autonomous as human agency is, since it has its origins in the work of designers and programmers, and in the learning processes that cognitive robotic systems have gone through.[138]

[132] A Adriaensen et al, "Teaming with industrial cobots: a socio-technical perspective on safety analysis" (2022) 32(2) Human Factors and Ergonomics in Manufacturing & Service Industries 173–98, 180.

[133] K Lopez-de-Ipina et al, "HUMANISE: Human-Inspired Smart Management, towards a Healthy and Safe Industrial Collaborative Robotics" (2023) 23 Sensors 1170, 4.

[134] Council Directive of 12 June 1989 on the introduction of measures to encourage improvements in the safety and health of workers at work, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:01989L0391-20081211&from=EN>.

[135] Refer to C Colosio et al, "Workers' health surveillance: implementation of the Directive 89/391/EEC in Europe" (2017) 67(7) Occupational Medicine 574–78.

[136] Art 3.b.

[137] Check, eg, M Kovac, "Autonomous Artificial Intelligence and Uncontemplated Hazards: Towards the Optimal Regulatory Framework" (2022) 13(1) European Journal of Risk Regulation 94–113, 112–13; SJ Bayern, "The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems" (2016) 7(2) European Journal of Risk Regulation 297–309, 304–06; Bennett and Daly, supra, note 19; KT Mamak, "Humans, Neanderthals, robots and rights" (2022) 24 Ethics and Information Technology 33; A Bertolini and F Episcopo, "Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective (2022) 9 Frontiers in Robotics and AI 842213.

[138] UNESCO, Report of the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) on Robotics Ethics, SHS/YES/COMEST-10/17/2 REV., Paris, 14 September 2017, 42.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

Reflecting on this, we humans, acting upon our own DNA "programming", behavioural imitation by models, and our own "learning processes" acquired via growth and development, are *not* exceptionally different; and in any case, for both parties to stay safe, commitment should be reciprocal.[139]

The Directive also mandates that MSs "ensure adequate controls and supervision".[140] In light of its general preventative approach, this could plausibly be understood in our case as ensuring that robots entrusted with OHS enforcement are smart (and thus adaptable and responsive to changing circumstances) but not *too* smart (and thus capable of self-governance and resistance to external deactivation).[141] It also provides that whenever supervisory capability cannot be found within the firm, outsourcing is allowed, as long as the employer makes external services aware of factors that are actually or potentially affecting the health and safety of workers, and all relevant information thereabout, including measures and activities in the realm of first aid, fire-fighting and evacuation.[142] Since the encoding of safety rules is a profitable and technologically advanced solution, this may create issues related to algorithms and know-how as trade secrets[143] and related licensing options. In any case, it is clear that this Directive was conceived for another era:

> It does not define *levels of autonomy* for robots which in the future may play a significant role in the work process of many European employers. This makes the analysis of risks and threats posed by new areas of scientific and technological progress a justified task.[144]

4. Machinery Directive

Directive 2006/42/EC (the Machinery Directive)[145] inhabited a key position within the EU industrial safety rules architecture. As per this Directive's terminology, in the case scrutinised here, the encoded robot would oversee the application of safety policies as its normal functioning, while "safety components" would be those installed into the robot to ensure that such normal functioning is in fact carried out properly rather than endangering other robots and – most importantly – humans. Indeed, the Directive defined a safety component not merely as that "the failure and/or malfunction of which endangers the safety of persons", but also that "which is not necessary in order for the machinery to function, or for which normal components may be substituted in order for the machinery to function".[146] Tellingly, the

[139] Read also A Potthast, "Ethics and Extraterrestrials: What Obligations Will We Have to Extraterrestrial Life? What Obligations Will It Have to Us?" in KC Smith and C Mariscal (eds), *Social and Conceptual Issues in Astrobiology* (Oxford, Oxford University Press 2020) pp 197–208, 199–201; WM Schröder, "Robots and Rights: Reviewing Recent Positions in Legal Philosophy and Ethics" in J von Braun et al (eds), *Robotics, AI, and Humanity: Science, Ethics, and Policy* (New York, Springer 2021) pp 191–203.

[140] Art 4.2.

[141] For instance, this is precisely the reason why deactivation buttons should be straightforward (ie accessible enough, big enough, simple enough) and mechanical, detached from the machine's "smart" circuit, in such a manner that the human worker could always switch these machines off in the event of a safety emergency, regardless of cobotic decision-making powers. To put it simply: *this "last-resort" safety component should not be smart*. The safety-grounded importance of mechanical, easy-to-find and obvious-to-use switch-off systems is extensively reported in the literature. For a recent exemplification, refer to Z Iqbal et al, "Detachable Robotic Grippers for Human–Robot Collaboration" (2021) 8 Frontiers in Robotics and AI 644532, 5.

[142] Arts 7.4, 8.2, 10.1/2.

[143] Cf. Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast), Document 02006L0042-20190726, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006L0042-20190726> Art 18.1.

[144] Jarota, supra, note 25, 4, emphasis added.

[145] Directive 2006/42/EC, supra, note 143.

[146] Art 2(c).

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

Directive added that such a component was one "which is independently placed on the market"; hence, the applicability of this piece of legislation to safety aspects of the case under scrutiny was limited to separate components – mostly mechanical ones – and did *not* encompass built-in antagonistic algorithmic control systems that enhance the safety of OHS-encoded SmaCobs (eg by constraining the "learning" of the "main" ones).[147]

On a different note, while this Directive mandated an "iterative process of risk assessment and risk reduction"[148] based on the International Organization for Standardization (ISO)'s 12100 standard, no specific mention was made of ML risk mitigation, namely vis-à-vis strategies to identify, implement and recalibrate over time relevant limitations on robots' ability to learn from observable patterns and refine their behaviour accordingly. In fact, as per the Directive, machines could be marketed so long as they did "not endanger the health and safety of persons [. . .], when properly installed and maintained and used for [their] intended purpose or under *reasonably foreseeable conditions*",[149] with the latter hardly being identifiable when ML is at stake – unless one read the "reasonably" qualification liberally. Either way, the Commission declared that it might accommodate amendments related to the Internet of Things (IoT) and smart robotics[150] – though the contents of such amendments as well as the Commission's policy approach in integrating them remain undefined – and they definitely do not feature in the New Machinery Regulation.

5. Policies and technical standards

In touching upon the recently issued civil liability rules, we leave the legal tangle of Directives behind and approach a marginally less formal but possibly even more essential regulatory portfolio, composed of extensive policies and technical standards. EU institutions have been long framing robotics legal developments within initiatives on AI,[151] with the tandem of the proposed AI Act and AI Liability Directive representing the culmination of this policy journey of countless Recommendations, Briefs, Proposals, Communications, Reports, Surveys and so forth. In the explanatory introduction to its proposed Liability Directive aimed at repealing the 1985 Directive, the Commission asserted that "factors such as the interconnectedness or *self-learning functions* of products have been added to the non-exhaustive list of factors to be taken into account by courts

[147] The effectiveness of "antagonistic" algorithms and their exact scientific taxonomy are yet to be defined in the literature. In fact:

> some seek to develop algorithms that, by probing, can look inside the black box of a trained "self-made" neural network and reveal the logic behind a subset of results, but not the overall logic of the system. Others "embrace the darkness" and rely on neural networks to explore other neural networks.
>
> – Gerdes, supra, note 22, 682.

To exemplify, one could rebut the claim that antagonistic algorithms are themselves subject to improvement through learning, so that this learning, too, might end up amplifying the same unforeseeability (and related hazards) characterised by that of the main algorithm. Moreover, the two algorithms might learn from the same circumstances but in a "misaligned" manner (eg at a different pace), so that the working by the antagonistic one might prove inefficient in constraining (up to offsetting when necessary) the work of the main one – assuming the latter learns at a faster pace.

[148] Annex I, para 1.

[149] Art 4.1, emphasis added.

[150] See EC, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "on the Mid-Term Review on the implementation of the Digital Single Market Strategy: A Connected Digital Single Market for All", COM(2017) 228 final, 10 May 2017, 11.

[151] Refer also to EC, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: "Artificial Intelligence for Europe", COM(2018) 237 final, 25 April 2018, 3.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

when assessing defectiveness",[152] which resonates well with the general aim of adapting the EU's liability framework to smart applications.

The next step for us here is to study EU law's interaction with international industry standards.[153] Those are, for example, the ones developed by ISO Technical Committees (TCs) 299 "Robotics", 184/SC 2 "Robots and robotic devices", and 110/SC 2 "Safety of powered industrial trucks", including the aforementioned ISO 12100, as well as ISO 3691-4:2020 on safety requirements for and verification of *inter alia* driverless industrial trucks and their systems, and ISO Technical *Specification* (TS) 15066, dedicated to cobots – but not a standard (yet).

Attention should be paid to ISO 10218, originally issued in 2011 and whose updated version was set to be released by 2022 – at the time of writing, though, it is still awaiting a compliance check with the theoretically voluntary and yet quasi-binding EU harmonised standards (ENs).[154] The second part of ISO 10218 (ie ISO 10218-2) is dedicated to industrial robot systems, robot applications and robot cells, with a decisive collaborative flavour. It was reported that

> [a]s a "draft international standard" (DIS), the ISO/DIS 10218-2 was published in late 2020 and is currently under evaluation. Collaborative applications are identified as characterized by one or more of three technologies: "hand-guided controls" (HGC), "speed and separation monitoring" (SSM), and "power and force limiting" (PFL). Specific risk assessment is envisaged for potential human–robot contact conditions, as well as related passive safety design measures and active risk reduction measures.[155]

One of the key premises of this paper can be restated again here: if one examines the list of "[c]hanges between the 2011 edition and the ongoing revision",[156] numerous interesting items can be found – including "local and remote control", "communications" and "cybersecurity" – but the issue of AI is totally neglected; dossiers such as AI security and liability, algorithmic governance or ML are not tabled as deserving of inclusion. This class of challenges, in its intersection with the collaborativeness of future robots, is left off the agenda. Unfortunate as it may be, it is not entirely surprising: international safety standards are still hampered by priorities that were conceived for high-volume manufacturing. Instead, not only "today [can] the same robot manipulator [. . .] be used for manufacturing, logistics, rehabilitation, or even agricultural applications[, a versatility that] can lead to uncertainty with respect to safety and applicable standards",[157] but algorithmic learning warrants even deeper policy specialisation in order to account for the most diverse professional domains.

Another ISO TC, namely number 199 ("Safety of machinery"), has developed standard 11161:2007, specifying the safety requirements for integrated manufacturing systems that incorporate interconnected machines for specific applications. The same Committee, and

[152] EC, Proposal for a Directive of the European Parliament and of the Council "on liability for defective products", COM(2022) 495 final, 28 September 2022, sec 5, 12, emphasis added.

[153] For a non-exhaustive but well-compiled table on OHS-focused HRC-applicable standards, refer to Valori et al, supra, note 6, 5; the related recommended testing procedures are summarised in ibid, 6–8. Check also this European Parliament infographic: <https://www.europarl.europa.eu/thinktank/infographics/robotics/public/concern/safety-issues.html>. The reader will realise how these are component-focused, as if assembling complex machines would not prompt the origination of emergent features; emphasising component testing over comprehensive robotic decisional outcomes makes these machine validations obsolete, or at least not ready for the ML era.

[154] Refer to T Meany, "Changes to the Industrial Robot Safety Standard ISO 10218" (2022) Engineer Zone <https://ez.analog.com/ez-blogs/b/engineerzone-spotlight/posts/changes-to-the-industrial-robot-safety-standard-iso-10218>.

[155] Valori et al, supra, note 6, 6.

[156] Available online at <https://committee.iso.org/sites/tc299/home/projects/ongoing/iso-10218-2.html>.

[157] Valori et al, supra, note 6, 8.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

particularly its Working Group 8 ("Safe Control Systems"), has also developed the two-part standard EN/ISO 13849; this builds on the International Electrotechnical Commission (IEC)'s 62061 standard, which had simplified the original IEC 61508 standard for the machinery sector, and it is of special relevance here as it applies it to safety-related parts of control systems – that is, components of control systems that respond to safety-related input signals and generate safety-related output signals.[158] The 13849 standard adopts the definitions of safety integrity level (SIL) and performance level (PL) so as to rate probabilities of harmful events occurring due to *overall machine safety levels* through quantifiable and non-quantifiable variables.[159] In this sense it displays, we consider, the right approach.[160] However, the shortcomings of the other standards mentioned previously can also be seen here: no mention is made of QC or AI-related hazards (and opportunities), and expressions such as "artificial intelligence", "algorithms" or "machine learning" are wholly absent from both the text and the accompanying major techno-policy reports.[161] The text does incorporate parameters and procedures such as "software safety lifecycle with verification and validation", "use of suitable programming languages and computer-based tools with confidence from use" or "impact analysis and appropriate software safety lifecycle activities after modifications".[162] However, these only cover traditional electronics and software and thus "non-smart" IT programming and coding endeavours, and their simplicity would be frustrated by the intricacies of algorithmic self-"improvement".

In September 2020, yet another ISO Committee – number 159, namely Working Group 6 on "Human-centred design processes for interactive systems" from its Sub-Committee 4 on the "Ergonomics of human-system interaction" – published its Technical Report (TR) 9241-810:2020 on "Robotic, intelligent and autonomous systems". This does fulfil its promise in identifying some of tomorrow's challenges for AI-driven cobots, but it is far from being translated into policy (ie a technical standard) – let alone action. One worrying aspect is that it strives for human enhancement as much as it calls for human-centric robot design, without elaborating on the risks of expecting humans to "enhance" their performance through moral elevation and physical fitness (including integrated bodily extensions such as prosthetics, implants and powered exoskeletons) with a robot-imposed pace and stakes.

The preceding paragraphs have given a due overview (and SmaCobs-orientated commentary) of the current scenario. On a more socio-legal reading, as "robot manufacturers typically take part in standardization committees, along with integrators, end-users and stakeholders representing public health and health insurance",[163] these are important for

[158] Read further A Söderberg et al, "Safety-Related Machine Control Systems using standard EN ISO 13849-1" (2018) RISE Research Institutes of Sweden <https://www.ri.se/sites/default/files/2019-09/Safety-related%20Machine%20Control%20Systems%20using%20standard%20EN%20ISO%2013849-1.pdf> 10; F Pera and GL Amicucci, "I sistemi di comando delle macchIne secondo le norme EN ISO 13849-1 e EN ISO 13849-2" (2017) Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro (INAIL) <https://www.inail.it/cs/internet/docs/alg-pubbl-sistemi-comando-macchine-secondo-norma.pdf> 7.

[159] Read further ABB Jokab Safety, "Safety in control systems according to EN ISO 13849-1" (2011) <https://search.abb.com/library/Download.aspx?DocumentID=2TLC172003B02002&LanguageCode=en&DocumentPartId=&Action=Launch> 2–6.

[160] However, it is often stressed in the literature that the "comprehensive approach" should cover not only all components of the machine and their overall operational outcome, but also an analysis of the tasks required of such a machine, together with the working environment (workpieces and workstations) as a system – in "ecosystem" terms. Check, eg, A Siebert-Evenstone et al, "Safety First: Developing a Model of Expertise in Collaborative Robotics" in AR Ruis and SB Lee (eds), *Proceedings of the ICQE Second International Conference: Advances in Quantitative Ethnography* (New York, Springer 2021) *pp* 304–18, 310.

[161] Refer, eg, to M Hauke et al, "Functional safety of machine controls – Application of EN ISO 13849" (2017) Deutsche Gesetzliche Unfallversicherung e. V. (DGUV), IFA Report 2/2017e <https://www.dguv.de/medien/ifa/en/pub/rep/pdf/reports-2019/report0217e/rep0217e.pdf>.

[162] Check para 4.6.2.

[163] Valori et al, supra, note 6, 4.

understanding the state of the field and for building consensus around policy harmonisation. In order "to devote attention to the significant influence of human–cobot workplace ethics on the process standardization of collaborative workplaces",[164] we call for hybrid committees that could inject insights from the social sciences into these engineering-intensive quasi-normative endeavours. As for their legal authoritativeness, international standards' salience is first grounded in governmental and "peer" auditors' expectations about their *de facto* bindingness, with the case of China (but referring to *domestic* standards) being an outstanding and normatively influential exemplification of this[165] – curiously, those auditors would themselves become a subject of SmaCobs' safety decision-making. Within the EU, for instance, the aforementioned ENs are adopted by the EC, explicitly endorsed by relevant governmental and executive agencies,[166] as well as referred to by courts through judicial activism[167] from the domestic to the Union level in adjudicating technically demanding cases.[168] Furthermore, they

> provide manufacturers with means to presume conformity with the requirements, through the legally binding "presumption of conformity". If the manufacturer decides not to use ENs, it bears the burden of proof to satisfactorily demonstrate that an alternative standard or methodology provides an equivalent or [higher] level of safety than that provided by the harmonised standard.[169]

International technical standards frequently feature in binding laws as an explicit or – most often – implicit reference; this is the case, for example, with the abovementioned Directive 89/391/EEC, which commits the European Council to remain aware of and up to date about not only "the adoption of [other D]irectives in the field of technical harmonisation and standardisation", but equally on "technical progress, *changes in international regulations or specifications* and new findings".[170] In this way, standards are often "hardened" into enforceable legislation. Several courts, and most prominently the Court of Justice of the European Union (CJEU), are increasingly following suit, extending their jurisdiction over the interpretation and bindingness of technical standards issued by private industry bodies and deciding liability cases based thereupon.[171] Given their incorporation into law, the extent to which those standards

[164] F Chromjakova et al, "Human and Cobot Cooperation Ethics: The Process Management Concept of the Production Workplace" (2021) 13(3) Journal of Competitiveness 21–38, 35.

[165] Refer to R Vecellio Segate, "Horizontalizing insecurity, or securitizing privacy? Two narratives of a rule-of-law misalignment between a Special Administrative Region and its State" (2022) 10(1) The Chinese Journal of Comparative Law 56–89, 67; R Vecellio Segate, "Litigating trade secrets in China: an imminent pivot to cybersecurity?" (2020) 15(8) Journal of Intellectual Property Law & Practice 649–59, 657–58.

[166] For the mere sake of exemplification, check this brief prepared by the German firm Redeker Sellner Dahs & Widmaier and published on the official website of Germany's Federal Ministry for Economic Affairs: <https://www.bmwk.de/Redaktion/EN/Downloads/Q/q-a-european-system-of-harmonised-standards.pdf>.

[167] Cf. G Gentile, "Ensuring Effective Judicial Review of EU Soft Law via the Action for Annulment before the EU Courts: A Plea for a Liberal-Constitutional Approach" (2020) 16(3) European Constitutional Law Review 466–92, 484–90.

[168] Refer, eg, to CJEU, *James Elliott Construction Limited v Irish Asphalt Limited* (C613/14), para 40; CJEU, *Global Garden Product v European Commission* (ECLI:EU:T:2017:36), para 60.

[169] JD Ramírez-Cárdenas Díaz, "European harmonised standards: Voluntary rules or legal obligations?" (2022) European Institute of Public Administration <https://www.eipa.eu/blog/european-harmonised-standards-voluntary-rules-or-legal-obligations/>; read also <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62021CC0588> paras 61–68; Directive 2006/42/EC, supra, note 143, Art 2(l).

[170] Art 17.1(a)(b), emphasis added.

[171] Refer, eg, to KP Purnhagen, "Voluntary 'New Approach' Technical Standards are Subject to Judicial Scrutiny by the CJEU! – The Remarkable CJEU judgment 'Elliott' on Private Standards" (2017) 8(3) European Journal of Risk Regulation 586–98.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

are genuinely international is worth pondering: do they reflect geo-economic power or scientifically validated best practices? Do they account for different (working) cultures, production dynamics, technical capabilities and socio-political orientations? To put it bluntly: are they *legitimate*?[172] These comparative socio-legal questions exceed the scope of this paper, but what we preliminarily observed is that jurisdictions around the world (perhaps less so in "non-Western" polities), differently from some scholarly circles,[173] tend to endorse such standards' technical validity and "policy legitimacy". For instance, Japanese scholars have called on the Japanese government to make such standards directly enforceable under Japanese law *as per the model of the EU*.[174] Furthermore, certain standards and specifications are directly addressed to jurisdictions beyond the Union.[175]

Lastly, can both rules and standards be encoded into robots, and will robots be capable of untangling normative conflicts – even situated at different degrees of bindingness? It was boldly asserted that "within a conceptual framework of algorithmic law and regulation, classical distinctions of legal theory [...] become either redundant or obsolete",[176] but this cannot discount smart robots' ability to output behaviours based on their assessment of the overall legal instruction stemming from multi-level and tangled legal documents whose instructions are situated on a sliding scale of characteristics, including indeed bindingness. This seems topical in a context such as algorithmic regulation that appears to be questioning long-standing dichotomies between, for instance, hard and soft laws.[177] This is precisely the aforementioned potential added value envisioned for QC in the (near?) future: to provide algorithms governing the robotic enforcement of safety rules with both the computing power and the sophistication to immediately respond to even the most unheard-of scenarios by recourse to the most appropriate combination of legal resources as selected among the applicable hundreds or thousands – from the softest to the hardest ones. However, whether such a future scenario for QC arises remains to be seen and will be dependent on a number of technical, social, economic, political and – of course – legal factors.

V. Key points for a tentative proposal under EU law

> *Quantum mechanics and chaos theory have demonstrated that perfect predictability is nothing more than a chimera.*[178]

[172] Read also R Neerhof, "The Use of Conformity Assessment of Construction Products by the European Union and National Governments: Legitimacy, Effectiveness and the Functioning of the Union Market" in P Rott (ed.), *Certification – Trust, Accountability, Liability* (New York, Springer 2019) pp 73–106; M Eliantonio and C Cauffman (eds), *The Legitimacy of Standardisation as a Regulatory Technique: A Cross-Disciplinary and Multi-Level Analysis* (Cheltenham, Edward Elgar 2020).

[173] Refer, eg, to M Almada, "Regulation by Design and the Governance of Technological Futures" (2023) European Journal of Risk Regulation <https://doi.org/10.1017/err.2023.37>.

[174] Refer to T Saito, "Global harmonization of safety regulations for the use of industrial robots-permission of collaborative operation and a related study by JNIOSH" (2015) 53 Industrial Health 498–504, 501–02.

[175] Check, eg, the ISO/TS 3691-8:2019 "Industrial trucks – Safety requirements and verification – Part 8: Regional requirements for countries outside the European Community".

[176] PRB Fortes et al, "Artificial Intelligence Risks and Algorithmic Regulation" (2022) 13 European Journal of Risk Regulation 357–72, 365.

[177] Refer to ibid, 365. The same debate has long entertained global and regional powers such as the EU, China, India and Russia over the negotiation of norms for the governance of other technologies as well; read, eg, R Vecellio Segate, "Fragmenting Cybersecurity Norms through the Language(s) of Subalternity: India in 'the East' and the Global Community" (2019) 32(2) Columbia Journal of Asian Law 78–138, 89–90.

[178] A Romano, *Quantum Tort Law: The Law of Torts in a Probabilistic World* (2019) PhD thesis in Law & Economics at LUISS University in Rome, p 138.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

In the preceding section, we have listed and analysed current laws, policies and standards in place internationally and especially within the EU to outline all major elements that, as they stand to date, appear to contrast with the effective policing of safe environments where SmaCobs and human workers can thrive together. Drawing on the gaps we identified in these frameworks, accounting for the relevant contexts described earlier and noting that the New Machinery Regulation has failed to deliver on its promise, here we offer some key reasoning and recommendations, with a view to developing a "core" policy set that (EU) policymakers may wish to consider towards the aim of crafting of a more coherent, comprehensive and forward-looking regulatory approach to SmaCobs.

Such an approach would be in line with the regulatory turn in technology-intensive sectors, as especially but not exclusively witnessed in the EU.[179] Over the last decade or so, the idea that market mechanisms would suffice to govern technology and innovation has become less popular in the wake of surveillance and data protection-related scandals such as the Snowden revelations, Cambridge Analytica and others. The EU has also harnessed the "Brussels Effect" of its legislation to (attempt to) assert its power and stance within the multipolar geopolitical scenario and race to dominate AI development: with China and the USA ahead of it in terms of technology development, the EU may be compensating with regulatory development, "persuasiveness" and sophistication.[180] While market mechanisms may assert a governing force on SmaCobs, these are beyond the scope of this paper and require further research. In the meantime, given the EU's regulatory turn and the deficiencies of the current legal framework identified in the previous section, we consider that legal reform is warranted and consistent with the EU's current approach to technology governance.

At the outset, let us stress that after several decades of robotics being regulated through Directives, a legal instrument in the form of a Regulation rather than a Directive, as demanded by the EP itself[181] as well as by the Commission in its then-draft New Machinery Regulation,[182] was indeed warranted. As outlined in the previous section, this field was already Directive-intensive, and it would have certainly benefitted from a bloc-wide harmonised approach. Transposition timings for Directives into MSs' domestic legal orders are lengthy, and such a strategic dossier could no longer be fragmented along national dividing lines. The New Machinery Regulation is a move in the right direction, but Fig. 1 supra delineates in detail, comparatively, why its role in potentially regulating smart cobotics will remain limited.

The time has come for a fully-fledged Regulation *conceived for the challenges of smart cobotics*, particularly if the EU aspires to outpace (or at least keep up with) China and the USA, reinforce the Brussels Effect and establish an overarching and efficient normative framework that prepares companies for the automation age by integrating data protection, IP and OHS standards into a coherent regulatory landscape. Also, this industry needs to be highly standardised because it is by definition an integrated one: SmaCobs are expected to contribute to "high manufacturing" operations, with examples including the assembly of aerospace components or the construction of cutting-edge biomedical facilities – all concerted efforts that mostly involve cooperation across domestic jurisdictions (and, indeed, across the European continent). Related to this, the Regulation should adopt an "omnibus" as opposed to

[179] See also D Broby, A Daly and D Legg, "Towards Secure and Intelligent Regulatory Technology (Regtech): A Research Agenda" (2022) Technology and Regulation, 88.

[180] Read, eg, A Daly, "Neo-Liberal Business-as-Usual or Post-Surveillance Capitalism with European Characteristics? The EU's General Data Protection Regulation in a Multi-Polar Internet" in R Hoyng and GPL Chong (eds), *Communication Innovation and Infrastructure: A Critique of the New in a Multipolar World* (East Lansing, MI, Michigan State University Press 2022).

[181] Check, eg, European Parliament resolution of 20 October 2020 with recommendations to the Commission "on a civil liability regime for artificial intelligence" (2020/2014(INL)) (2021/C 404/05), 20 October 2020, para 5.

[182] Preambulatory Clause 11 to the Proposal for a Regulation of the European Parliament and of the Council on machinery products, 21 April 2021, COM(2021) 202 final, 2021/0105 (COD), Document 52021PC0202.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

"sectoral" approach[183] – possibly leaving room to domestic legislators for some autonomy regarding how it would apply to industry sectors in which each MS is prominent, but it should nevertheless ensure compliance with the binding core of our proposed Regulation and account for SmaCobs' use across traditional industry or sectoral boundaries. By "binding core" we intend the (majority) bundle of obligations that would be immediately detailed enough to be uniformly transposed into MSs' domestic legal orders, without further specification as to the means to satisfy said Regulation's requirements. True, immediate transposition is the very essential advantage of any Regulation over a Directive, but we have already witnessed (eg with biometric data processing as per the General Data Protection Regulation (GDPR)) that when a Regulation promises to exhibit extreme complexity, MSs might negotiate its precise requirements up to a certain extent, leaving *a few* most controversial sections somewhat "open-ended" as to how MSs are to implement specific groups of provisions[184] through *a sort of* "mini-Directive" within the Regulation itself.

Such a Regulation should finally resolve and take a "future-proof" stance on a range of controversial matters (as we surveyed them earlier). The first of these should be whether robots can be granted legal personhood and, if so, be deemed as employers and thus entrusted with responsibility to oversee workers, train them, inform regulatory agencies, coordinate with authorities and appoint relevant subordinates (eg OHS managers). In this respect, it should be recalled that Directive 89/391/EEC confers on workers the duty to "cooperate [. . .] with the employer and/or workers with specific responsibility for the safety and health of workers",[185] which would establish a whole new human–robot interaction (HRI) field and redefine the parameters of notification and processing of imminent dangers. If we accept robots as legal persons (equivalent to, eg, a corporation), how are we supposed to apportion liabilities in the event of faults? The issue would deserve extensive analysis on its own, but for now, and within the limits of this paper, we will reason by analogy from data protection law. Under the EU's GDPR, the users of the data-processing device can be generally assumed to be data controllers, while manufacturers are exempted from liability as they are not those who process the data directly. However, a recent stream of scholarship advises that

> the development of "smart" devices with "local" or "edge" computing upends the assumption that device manufacturers automatically fall outside the scope of the GDPR. Device manufacturers may often have sufficient influence on the processing to be deemed "controllers", even where personal data is processed on-device only without any direct processing by the manufacturer. This is because device manufacturers may in certain cases "determine the means and purposes of the processing".[186]

Similar reasoning could apply in the OHS domain. Indeed, if robots are truly "smart" and respond to the environment largely independently of the input of their ultimate user, there would be no reason to bestow the ultimate user with civil liability for faults that are, in fact, closer to shortcomings on the programming or market approval side. If a robot is programmed to pursue unsupervised learning, programmers should not face liability for

[183] On this distinction, borrowing again from the data protection realm, read, eg, PM Schwartz, "The EU–U.S. Privacy Collision: A Turn to Institutions and Procedures" (2013) 126(5) Harvard Law Review 1966–2009, 1974.

[184] See P Quinn and G Malgieri, "The Difficulty of Defining Sensitive Data – The Concept of Sensitive Data in the EU Data Protection Framework" (2021) 22(8) German Law Journal 1583–612, 1589–90; M Monajemi, "Privacy Regulation in the Age of Biometrics That Deal with a New World Order of Information" (2018) 25(2) University of Miami International and Comparative Law Review 371–408, 382; EJ Kindt, "Having Yes, Using No? About the new legal regime for biometric data" (2018) 34(3) Computer Law & Security Review 523–38.

[185] Art 13.2(e).

[186] A Dahi and MC Compagnucci, "Device manufacturers as controllers – expanding the concept of 'controllership' in the GDPR" (2022) 47 Computer Law & Security Review 105762.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

the outcomes themselves of such ML but rather for the very fact that no appropriate limits to this learning were encoded into the robot. By "limits" we mean not mere temporal or contextual limitations, but self-restraint in the number and complexity of interconnections among pieces of data that the robot feeds itself with in order to learn from them. Admittedly, this is easier said than done; however, the technical preliminary discussion before encoding standards should indeed revolve around the conceptual identification, engineering viability, legal definition and ethical boundaries of the mentioned "limits". Once smart cobots OHS legislation in the EU becomes accurate enough to set out red lines not to be overcome and mandates programmers to grant machines learning abilities *up to a specific extent only*, then any harm ensuing from the defiance of such limits on the part of the machine should indeed be attributed to the manufacturer rather than final users.

On more operational grounds, machines' smart-to-non-smart progressions and multiphase gradations should be foreseen by law, to the extent of temporarily shutting learning endeavours down upon need (eg during the most critical assembly stages or upon potential harm materialising). No matter how smart, there should be a mechanism to switch machines off if risks of harm to humans concretise, meaning that at least one available human should always retain a last-resort commanding capacity within the firm at any one time – somewhat based on the model of flight commanders versus automatic pilots in civil aviation. This would offset the unprecedented and discouraging relationship of subordination that certain low-skilled workers may experience vis-à-vis robots if the latter take over the direction of (given chains of) field operation. It is worth emphasising that workers should not necessarily enjoy the capacity to *redirect* dysfunctional robots' operations, but just to turn them off; indeed, workers' substandard (or in any case cognitively untrained) performance in redirecting robotic operations might itself represent a source of hazard,[187] which could even come as unsupervised in this case because OHS standards would be encoded into and enforced by those same robots that exhibits signs of dysfunction – *who watches the watchers*? In this respect,

> regulators tackling the issue of human oversight in the future will need to develop clear criteria that balance the potentially competing interests of human-centrism versus utility and efficiency. Indeed, [. . .] the "robot-friendliness" of the law in different jurisdictions may lead to a new regulatory race to the bottom.[188]

Even where workers were not originally under-skilled, the protracted entrusting of standards enforcement (eg on safety) to robots could gradually lead to obsolescence: "as professionals

[187] This is because

> [h]umans, bounded by the cognitive limitations of the human brain, are unable to analyze all or even most of the information at their disposal when faced with time constraints. They therefore often settle for a satisfactory solution rather than an optimal one [. . . . A]n AI program can search through many more possibilities than a human in a given amount of time, thus permitting AI systems to analyze potential solutions that humans may not have considered, much less attempted to implement. When the universe of possibilities is sufficiently compact [. . .,] the AI system may even be able to generate an optimal solution rather than a merely satisfactory one. Even in more complex settings, [. . .] an AI system's solution may deviate substantially from the solution typically produced by human cognitive processes.

– MU Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (2016) 29(2) Harvard Journal of Law & Technology 353–400, 364–65. This is not to say that AI-driven outcomes are necessarily *better*, but they might be *different* from human-conceived outcomes. Read also Janssen, supra, note 41, 84.

[188] E Hickman and M Petrin, "Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective" (2021) 22 European Business Organization Law Review 593–625, 602.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

cease to exercise their skills and end up less proficient when the technology fails",[189] the threat of "skill fading" should be factored into any emergency plan contingent to robotic failure.

Related to this, MSs should establish a procedure to make sure that robots with encoded OHS standards remain *smart but not too smart*, perhaps combining an open-access techno-scientific registry keeping record of periodic cycles of review and auditing. This registry should evidence how, within the ethical autonomy continuum, encoded robots stop short of exceeding the point upon which their ethical landscape would become fully automated,[190] and it should be open for the public to monitor and even submit observations. As regards auditing, given that tomorrow's robots will be in the hundreds of millions but most of them highly customised, and tracking how each learns over time would be unsustainably expensive, problems would arise as to what "sample" would stand as representative of all robots from a given market batch, production line or corporate conglomerate. It is emphasised in the literature that "[w]ith the emergence of automation in audit, sampling procedures become obsolete because algorithms allow auditors to audit the entire data population, not just the sample".[191] This might hold true for all "algorithms as such" as they are initially programmed, but it will prove unhelpful with regards to both how such algorithms learn over time and how they come to interact with the specific sensing and motor apparatuses of each different cobot. Indeed, the parallel problem arises as to what class of on-the-ground situations would be representative of hazards caused by dysfunctional safety-encoded cobots. All of these matters, in turn, depend on whether safety-encoded robots encompass themselves within their policing scope or are confined to enforcing rules vis-à-vis third parties (ie humans and other robots) only – an issue that also relates to cognate debates around robots' self-perception and rudimental degrees of sentience.

The proper balance between OHS standards' encoding by-design and on-the-road readjustment should be sought. We should ensure that (most of) cobots' functions are designed to be safe from the outset, but safeguarding reprogramming flexibility can prove equally wise and salient in certain contexts where design rigidity is more of a barrier to day-to-day safety choices than an enabler, particularly with regards to the neuropsychiatric spectrum of professional health. The legislator is called upon to be pragmatic and to ensure that the "division of labour" between safety encoders at the design stage and those contributing to it subsequently is clearly delineated and mutually subscribed to, both inclusively and participatorily through testbeds attended by physically, socio-culturally and neurodiverse pools of individuals representative of expected but also future potential users. Not only are there "difficulties in identifying the point in time during a robot's trajectory where a specific algorithm is the least safe, requiring either a simulation or a test with the completed system",[192] but even once said completed system is assembled, programmed and tested, unbounded ML would make it impossible for regulators to assess its hazardousness. Programmers and inspectors alike are called upon to pursue not maintenance but reassessment and verification *from scratch* at regular intervals. Mindful of the further caveat that certain "learnings" might trigger exponential rather than linear behavioural changes vis-à-vis the time scale, regulators should also appreciate that

[189] JX Dempsey, "Artificial Intelligence: An Introduction to the Legal, Policy and Ethical Issues" (2020) A paper for the Berkeley Center for Law & Technology <https://www.law.berkeley.edu/wp-content/uploads/2020/08/Artificial-Intelligence-An-Introduction-to-the-Legal-Policy-and-Ethical-Issues_JXD.pdf> 21.

[190] Read also Z Tóth et al, "The Dawn of the AI Robots: Towards a New Framework of AI Robot Accountability" (2022) 178(4) Journal of Business Ethics 895–916, 905.

[191] A Tiron-Tudor and D Deliu, "Reflections on the human–algorithm complex duality perspectives in the auditing process" (2022) 19(3) Qualitative Research in Accounting & Management 255–85, 265.

[192] Valori et al, supra, note 6, 3.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

scenario-disrupting expectational asymmetries can arise any time, so that "regular" inspection might not mean much as regards safety.[193]

In some cases, smart adaptation to complex (actual or intended) interactional stimuli and environmental cues would lead to incidents that are, however, fewer in number (and perhaps magnitude?) compared to non-adaptation. Think of a robot that is capable of adapting its instruction delivery to the linguistic proficiency, diction (accent, tone), phraseological structure, cultural literacy, situational awareness and frequent vocabulary of relevant users. This could lead to overconfident users or confuse other relevant users (eg temporary shift replacements), but mastering the language at different levels, too, might trigger incidents caused by miscomprehension between robots and human speakers – particularly when non-natives or specially disadvantaged subgroups of blue-collar workers are at play.

The identification of context-sensitive risk-mitigation strategies is of the essence. These may depend *inter alia* on the nature (field of activity, public exposure, political susceptibility), location (surroundings, applicable jurisdiction, etc.) and size of the firm, training background of relevant officers, safety equipment and engineering conditions of building spaces such as laboratories and workstations, evacuation plans, substance toxicity, presence of workers in special need of assistance, project flows and proximity to first-aid centres.

In policing robotics, legislators have long tended to reinforce old-fashioned dichotomous divides between "mental" and "physical" health, to focus on the latter. Nonetheless, common (and increasingly prevalent all across "developed" societies) neuropsychiatric conditions such as anxiety and depression – but also panic, bipolar, schizophrenic, post-traumatic and obsessive–compulsive disorders, to mention but a few – should feature right at the core of any assessment strategy and be given due weight by engineers while designing robotic forms of interaction with humans.[194] More specifically, we recommend that engineers prioritise them within AI-driven checklists when coding SmaCobs' enforcement of OHS rules. For instance, biomedical engineers should seek the assistance of clinicians and other relevant health professionals in pondering how the understanding of, say, generalised anxiety disorders approximate to real human experiences when anxiety identification, mitigation and prevention are encoded into robots as part of safety responses. Would robots be sensitive to said disorders and their inexplicable fluctuations? Would they perceive (clinically relevant) anxiety in any "human-like" fashion – assuming (without conceding) there is any?[195] Also, the information currently to be shared with potential outsourced safety providers does not read as having been conceived for the automation age, in which hazards tend to blend physicality and psychology with cognition conundrums and mostly relate to machine autonomy, human mental health well-being and suboptimal techno-managerial expertise. EU legislation has long addressed the issue of workplace-related stress,[196] requiring supervisors to keep it monitored and intervene when necessary: this is what would be required of robots, too, were they entrusted with health supervisory functions – also vis-à-vis more demanding conditions. On a less clinical regard but still related to the human mind, mitigation strategies to reduce *psychological* (as "alternative" to *psychiatric*) *discomfort* (as "opposed" to *disorders*) should be devised as well, including through

[193] Cf. 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, COM(2021) 206 final, 21 April 2021, para 66.

[194] See JR Crawford et al, "Generalised Anxiety Disorder and Depression on Implicit and Explicit Trust Tendencies Toward Automated Systems" (2021) 9 IEEE Access 68081–92.

[195] Refer further to H Ashrafian, "Can Artificial Intelligences Suffer from Mental Illness? A Philosophical Matter to Consider" 2017 23(2) Science and Engineering Ethics 403–12.

[196] Check, eg, the Framework Directive 89/391/EEC, as well as Council Directives 90/270/EEC, 92/85/EEC and 2010/32/EU.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

questionnaires, psychological metrics and behavioural metrics developed by HRI-specialised research institutes.[197] These should feature as binding (but progressive) requirements within the Regulation we are outlining here.

Just like any other algorithmically driven machines, smart robots are prone to bias.[198] Mindful of the discriminatory allocational strategies subsumed under the vast majority of algorithmically devised groups,[199] the Legislator should decide whether SmaCobs can and should define specially protected (or "vulnerable") clusters of individuals for OHS purposes and who has ultimate legal recourse capability as a remedy if serious consequences arise from robots' misallocation of a worker in over- or under-protected categories. Whose interests will SmaCobs protect first?[200] In human-populated working environments, "[w]hat is advantageous to the firm may be viewed negatively from an individual's perspective[, as d]iffering groups need to navigate and negotiate their values with respect to the others".[201] But what if the negotiation is to be carried out between humans *and robots*? Would humans accept safety "paramount interests" to be prioritised, calibrated and acted upon by algorithmically fed robotic entities on behalf of a superior collectivity? Also with a view to "democratising" this process (to the extent possible), the most meaningful balance should be drawn between the disclosure of algorithmic codes for public accountability purposes and the protection of trade secrets that extend far beyond professional secrets. This is to ensure that innovation is not jeopardised through trade secrets' over-disclosure (and thus poor protection)[202] while little gain is achieved on the accountability side – *due to ML outcomes' inherent inexplicability*. Tentatively, specific provisions should be devised to encourage the sharing of best practices among employers and employees from different companies and business districts, without running into trade secret misappropriation, non-compliance with non-compete contractual clauses and prohibitions on collusive conduct in competition law. Furthermore, exactly because interest-maximisation strategies differ so remarkably between smart machines and humans, specific rules shall be devised to accommodate dispute prevention, handling and mitigation not only among SmaCobs or among humans, and not even among each SmaCob and its human co-worker only, but equally between different SmaCob–coworker teams within the same working group.

No matter how cohesive, our proposed EU Regulation should, however, incorporate context-sensitive provisions aimed at leaving room for further sector-specific regulatory manoeuvring (depending on the actual job profiles and tasks warranting smart robotic collaboration within Europe), and it should be informed by participatory stakeholder input and processes, especially from workers themselves. Its bindingness and direct applicability throughout the EU will come at a cost; in other words, that it should be conceived in the form of "framework legislation" whereby certain details are deferred to later legal instruments (while possibly being covered by soft arrangements in the meantime). This is also aligned with the preferences expressed by States during the negotiations towards the

[197] Read further at PA Lasota et al, "A Survey of Methods for Safe Human–Robot Interaction" (2014) 5(4) Foundations and Trends in Robotics 261–349, 321–22.

[198] See extensively L Londoño et al, "Fairness and Bias in Robot Learning" (2022) preprint available at <https://doi.org/10.48550/arXiv.2207.03444>.

[199] Refer generally to Vecellio Segate 2020, supra, note 165; Y Saint James Aquino et al, "Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives" (2023) Journal of Medical Ethics; S Alon-Barkat and M Busuioc, "Human–AI Interactions in Public Sector Decision Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice" (2023) 33(1) Journal of Public Administration Research and Theory 153–69; J Adams-Prassl et al, "Directly Discriminatory Algorithms" (2023) 86(1) Modern Law Review 144–75.

[200] Read also L Kähler and J Linderkamp, "The Legal Challenge of Robotic Assistance" in U Engel (ed.), *Robots in Care and Everyday Life: Future, Ethics, Social Acceptance* (New York, Springer 2023) pp 81–101, 92.

[201] Wallace, supra, note 80, 303.

[202] Read also R Vecellio Segate, "Securitizing Innovation to Protect Trade Secrets between 'the East' and 'the West': A Neo-Schumpeterian Public Legal Reading" (2020) 37(1) UCLA Pacific Basin Law Journal 59–126, 65.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

first ever international binding instrument on business and human rights[203] – which is somewhat relevant here as well.[204]

The Regulation we propose would also strive to more organically integrate industry standards into binding EU law, not as mere expectations of conduct but as compulsory safety requirements where applicable. Industry standards currently feature six main "skills" that machines should exhibit for them to be deemed safe (maintain safe distance; maintain dynamic stability; limit physical interaction energy; limit range of movement; maintain proper alignment; limit restraining energy),[205] but we advocate for the introduction of a seventh skill: monitor self-learning pace and outcomes and prevent harmful effects thereof on humans (and the environment) through pre-emptive switch-off. This is *what* should be tested; as for *how*, the beyond-component on-the-whole approach should be preferred whenever feasible, accounting for worst-possible-scenario types of unexpected behaviours grounded not so much in the most extreme actions that machines *are programmed* to deliver as in the most extreme actions they *could technically* (learn to) accomplish. Not least, interfaces should be explored with PIL and its jurisdictional assertions – for instance, when it comes to safety incidents stemming from inter-jurisdictional VR-mediated interactions between robots and humans, as well as in the humanitarian aid domain with the encoding of rules of engagement within automated health recovery procedures.

VI. Conclusions

Compared to non-collaborative industry robots, cobots (let alone smart ones) are still a market niche, but powering them with AI and possibly QC in the future – even accounting for sustainability challenges related to increased computing power and related energy consumption[206] – will have them strategically deployed for the forthcoming smart factories.[207] Yet, while the EU is assertively legislating across virtually the entire spectrum of policy areas invested in the digital, AI and soon quantum transformations,[208] including algorithmically intensive industries, no current law addresses the encoding of OHS standards for robots, nor does any ongoing legislative process address the specific risks stemming therefrom – or to cobotics more widely, for that matter.

The new EU liability regime for AI applications, which will be regulated through the (currently draft) AI Liability Directive, is not robotics-specific and fails to address most socio-technical complexities arising from human–cobot interactions in collaborative settings – as summarised in Fig. 1 supra. Furthermore, its harmonising momentum will

[203] See C Macchi, *Business, Human Rights and the Environment: The Evolving Agenda* (The Hague, TMC Asser Institute 2022) pp 148–49.

[204] Refer, for instance, to I Emanuilov and K Yordanova, "Business and human rights in Industry 4.0: a blueprint for collaborative human rights due diligence in the Factories of the Future" (2022) 10 Journal of Responsible Technology 100028.

[205] Check Valori et al, supra, note 6, 10.

[206] See generally D Mhlanga, "The Role of Artificial Intelligence and Machine Learning Amid the COVID-19 Pandemic: What Lessons Are We Learning on 4IR and the Sustainable Development Goals?" (2022) 19 International Journal of Environmental Research and Public Health 1879.

[207] For market forecasts and industry outlooks, check, eg, Intesa Sanpaolo Innovation Centre, "Industry Trends Report: Industrial & Mechanics – Process Automation and Digitization" (2022) <https://group.intesasanpaolo.com/content/dam/portalgroup/repository-documenti/newsroom/news/2022/ISPIC_Industrials%20and%20Mechanics_abstract.pdf> 4; Goldman Sachs, "Humanoid Robots: Sooner Than You Might Think" (2022) <https://www.goldmansachs.com/insights/pages/humanoid-robots.html>.

[208] Notably, the EC, and particularly its Directorate-General on Communication Networks, Content, and Technology (CONNECT, formerly CNECT), is supporting a decade-long "New Strategic Research Agenda on Quantum technologies" within its "Quantum Technologies Flagship" Action: <https://digital-strategy.ec.europa.eu/en/news/new-strategic-research-agenda-quantum-technologies>.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

remain limited precisely due to its status as a Directive, especially from a tort law perspective.[209] The AI Act is not resolutory in this essential respect either.[210]

As per occupational safety in robotics, the field was entirely regulated through Directives. With its most recent New Machinery Regulation, coming into effect in mid-June 2023, the EU did chart a qualitative step forward, which is nevertheless too modest and scope-limited to respond to today's needs *in this specific industry domain* as well as to withstand the related sector-specific regulatory competition from other world regions. This represents a fracture between the policy arena and industry developments, just as much as it embeds a disconnection between the lawyering world and engineers' concerns.[211] The misalignment is specular: law and technology specialists are aware of the AI revolution elicited by smart machines as regards the ethics of machine–human interactions but fail to translate them into cutting-edge, "frontier" policies; on their side, engineers mostly dismiss these ethical matters on potentially unforeseeable ML-triggered risk as peripheral, futuristic or at most improbable while progressing fast in assembling algorithmically powered robots. In fact:

> we could interpret their reactions as indicative of their lack of interest in such questions, or their lack of exposure to more metaphysical debates on the nature of AI. [. . .] From their perspective, they are only building machines that need to be safe, they are not building machines whose behaviour needs to be safe or lawful. [. . . Nonetheless,] this question of machine behaviour as *a distinct and emergent feature of robotics, which goes beyond the mere sum total safety of the components assembled* could become a relevant trope for analysis of the engineering practices.[212]

We agree: overall unpredictability of result for machine behaviour *taken as a whole* should be the driving concept behind policing efforts and culturally informed lawyer–engineer cooperation in smart collaborative robotics, with dignity for humans and robots at once as their guiding momentum,[213] and with mutual training[214] as well as contextual ethical awareness as appropriate.[215]

Hence, we surmise that the time has come for EU regulators to either embrace the techno-scientific and regulatory challenge *fully* (also via the establishment of supercomputers,

[209] This limitation is widely acknowledged and explored in the literature. See, eg, G Wagner, "Liability Rules for the Digital Age – Aiming for the Brussels Effect" (2023) 13(3) Journal of European Tort Law 198, 225, 232; ME Kaminski, "Regulating the Risks of AI" (2023) available on SSRN at <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4195066>.

[210] Refer, for instance, to C Gallese, "Suggestions for a Revision of the European Smart Robot Liability Regime" (2022) Proceedings of the 4th European Conference on the Impact of Artificial Intelligence and Robotics <https://doi.org/10.34190/eciair.4.1.851>; M Almada and N Petit, "The EU AI Act: Between Product Safety and Fundamental Rights" available on SSRN at <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4308072>.

[211] Indeed, "it is likely that the roboticists will not have done anything objectively wrong in the design of the defective product, but will also probably not have gone much further than 'good enough' state-of-the-art safety". For engineers, in most cases, safety seems to be "first and foremost a technical concern, it is a legal concern only at a secondary level [. . .]. Our perspective as lawyers is fundamentally different on those issues, where we tend to label safety regulation and their related standards as part of the legal spectrum." In the end, "[l]awyers see that the origin of safety is the legal regulation that requires specific thresholds of safety; whereas the roboticists do not see the necessity or value in mobilizing law" – L Van Rompaey et al, "Designing lawful machine behaviour: roboticists' legal concerns" (2022) 47 Computer Law & Security Review 105711, 7, 12.

[212] ibid, 8, emphasis added.

[213] Refer also to CC Goodman, "AI/Esq.: Impacts of Artificial Intelligence in Lawyer–Client Relationships" (2019) 72(1) Oklahoma Law Review 149–84, 180.

[214] By analogy, see, eg, H Drukarch et al, "An iterative regulatory process for robot governance" (2023) 5 Data & Policy e8, 12–13.

[215] Read extensively FS de Sio and G Mecacci, "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them" (2021) 34(4) Philosophy & Technology 1057–84, 1066–71.

https://doi.org/10.1017/err.2023.72 Published online by Cambridge University Press

serendipity-encouraging regulatory sandboxes[216] (including in VR[217]) that can mitigate the over-enforcement of safety compliance – as well as through knowledge support by funded projects[218] and the newly signed European Public–Private Partnership in AI, Data, and Robotics[219]) or accept that the EU will soon be outpaced by its East Asian and North American counterparts. The Commission itself acknowledged that "advanced robots and [IoT] products empowered by AI may act in ways that were not envisaged at the time when the system was first put into operation[, so much that g]iven AI's widespread uses, both horizontal and sectoral rules may need to be reviewed",[220] but in a time-shrunk policy sector where timing counts volumes,[221] it is not yet systemically acting upon such need. Without forward-looking regulatory efforts,[222] the EU will fail to secure its self-declared "world-leading position in robotics and competitive manufacturing and services sectors, from automotive to healthcare, energy, financial services and agriculture".[223] The New Machinery Regulation is "too little, too late"; and it falls short of properly serving *the specificity of smart cobotics.*

Across a wide portfolio of policy and professional domains, it seems often to be lamented that regulators' and lawyers' concerns tend to halt or delay as opposed to facilitate or enable innovation,[224] but OHS standards are so key to cobots' trustful adoption

[216] Refer generally to SH Ranchordás, "Experimental Regulations for AI: Sandboxes for Morals and Mores" (2021) 1 Morals + Machines 86–100; C Rosemberg et al, "Regulatory Sandboxes and Innovation Testbeds: A Look at International Experience and Lessons for Latin America and the Caribbean" (2020) A Final Report by the Inter-American Development Bank <https://www.technopolis-group.com/wp-content/uploads/2020/09/Regulatory-Sandboxes-and-Innovation-Testbeds-A-Look-at-International-Experience-in-Latin-America-and-the-Caribbean.pdf>; J Truby et al, "A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications" (2022) 13(2) European Journal of Risk Regulation 270–94. For EU and UK exemplifications, respectively, check, eg, J Dervishaj, "EC – NEWS: First regulatory sandbox on Artificial Intelligence to start in autumn 2022" (2022) <https://futurium.ec.europa.eu/en/european-ai-alliance/events/news-first-regulatory-sandbox-artificial-intelligence-start-autumn-2022>; LA Fahy, "Fostering regulator–innovator collaboration at the frontline: a case study of the UK's regulatory sandbox for fintech" (2022) 44(2) Law & Policy 162–84.

[217] Refer to A Arntz et al, "A Virtual Sandbox Approach to Studying the Effect of Augmented Communication on Human–Robot Collaboration" (2021) 8 Frontiers in Robotics and AI 728961.

[218] Explore further at EC, supra, note 150, 22. Check also the VOJEXT Project <https://cordis.europa.eu/project/id/952197> and its deliverable no. D1.3 on "Safety regulation and standards compliance" – work package no. 1 on "Fundamental building blocks: Framework, requirements and functional design" <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5e11e86b8&appId=PPGMS>.

[219] Read more at EUnited – European Engineering Industries Association, "The New Public Private Partnership: AI, Data, and Robotics" (2021) <https://www.eu-nited.net/eunited+aisbl/robotics/news/the-new-public-private-partnership-ai-data-and-robotics.html>.

[220] EC, supra, note 150, 15.

[221] For theoretical context around AI-driven legal time-shrinking and future-proof legislative outputs, read also RJ Neuwirth, "Future Law, the Power of Prediction, and the Disappearance of Time" (2022) 4(2) Law, Technology and Humans 38–59, 48.

[222] We shall advise against "legal solutionism" for its own sake: L Reins, "Regulating New Technologies in Uncertain Times – Challenges and Opportunities" in L Reins (ed.), *Regulating New Technologies in Uncertain Times* (The Hague, TMC Asser Press 2019) pp 19–28, 5. When it comes to smart cobotics, the law is not falling behind generally, but just in specific (and yet essential!) respects as outlined in this paper; hence, we are not suffering from "the urge to call law or regulation outdated or flawed (disconnected) and the desire to fix the problems by addressing the law, rather than using other ways to mend the assumed gaps" (ibid). It is more about considering existing techno-industrial standards, adapting current legal doctrines to specific challenges (eg the "boundaries exercise" mentioned earlier) and aligning them to societal expectations as closely as possible. As far as the EU is concerned, we believe that a Regulation is long overdue, but beyond the relatively narrow matters outlined earlier, it would be more about harmonising and systematising (and, seizing the chance, updating) all scattered Directives into one coherent corpus than about reinventing the wheel. Aside from the few express scenarios and dilemmas we addressed here, smart cobotics is not in need of legal revolution!

[223] European Commission, supra, note 40, 3.

[224] Check, for instance, Committee on Patient Safety and Health Information Technology Board on Health Care Services of the National Academies' Institute of Medicine, *Health IT and Patient Safety: Building Safer Systems for Better Care* (Washington, DC, US National Academy of Sciences 2012) p 140; T Relihan, "Will regulating big tech

and diffusion that regulatorily addressing the challenges that they bring about will only catalyse technical improvements, enhance reliability and unlock trustworthiness in this fast-paced sector over the years to come. We aspire for our European proposal to represent a frontier indication to that end, for the EU and beyond.

Acknowledgments. Earlier drafts of the present work were presented by the authors at the University of Oxford's Bonavero Institute of Human Rights ("Algorithms at Work" Reading Group, 9 March 2023), University of Aberdeen (2nd Annual SCOTLIN Conference, 27 March 2023), as well as at Belfast's Titanic Centre during the SPRITE+ Conference on 29 June 2023. We are grateful to the organisers and attendees of these three scholarly gatherings for their challenging questions and comments about our work. We acknowledge funding from the UK Engineering and Physical Sciences Research Council's "Made Smarter Innovation – Research Centre for Smart, Collaborative Industrial Robotics" Project (2021-2025, EPSRC Reference: EP/V062158/1). We also acknowledge precious inputs and insights from former and current members of the aforementioned Centre, including Professor YAN Xiu-Tian, Dr Tiziana Carmen Callari and Dr NIU Cong. Riccardo Vecellio Segate gratefully acknowledges the superlative learning environment at Politecnico di Milano (Polytechnic University of Milan), where he is currently enrolled as a BEng Candidate in Industrial Production Engineering and without whose inspiring teaching staff and library resources this paper would have never been accomplished. No humans or animals were involved in this research. While the substance of the present article was conceived for the first draft as completed and submitted in early December 2022, the authors have tried their best to keep the manuscript and its references current throughout the extensive peer-review and editorial process.

Competing interests. The authors declare none.

stifle innovation?" (2018) Ideas Made to Matter <https://mitsloan.mit.edu/ideas-made-to-matter/will-regulating-big-tech-stifle-innovation>. Read further Y Lev-Aretz and KJ Strandburg, "Regulation and Innovation: Approaching Market Failure from Both Sides" (2020) 38(1) Yale Journal on Regulation Bulletin 1–27.

Cite this article: R Vecellio Segate and A Daly, "Encoding the Enforcement of Safety Standards into Smart Robots to Harness Their Computing Sophistication and Collaborative Potential: A Legal Risk Assessment for European Union Policymakers". *European Journal of Risk Regulation*. https://doi.org/10.1017/err.2023.72