



Safety by simulation: theorizing the future of robot regulation

Mika Viljanen¹

Received: 26 August 2022 / Accepted: 13 July 2023
© The Author(s) 2023

Abstract

Mobility robots may soon be among us, triggering a need for safety regulation. Robot safety regulation, however, remains underexplored, with only a few articles analyzing what regulatory approaches could be feasible. This article offers an account of the available regulatory strategies and attempts to theorize the effects of simulation-based safety regulation. The article first discusses the distinctive features of mobility robots as regulatory targets and argues that emergent behavior constitutes the key regulatory concern in designing robot safety regulation regimes. In contrast to many accounts, the article posits that emergent behavior dynamics do not arise from robot autonomy, learning capability, or code unexplainability. Instead, they emerge from the complexity of robot technological constitutions coupled with near-infinite environmental variability and non-linear performance dynamics of the machine learning components. Second, the article reviews rules-based and performance-based regulation and argues that both will fail adequately constrain emergent robot behaviors. The article claims that controlling mobility robots requires a simulation-based regulatory approach. Simulation-based regulation is a novelty with significant theoretical and practical implications. The article argues that the approach signifies a radical break in regulatory forms of knowledge and temporalities. Simulations enact virtual futures to create a new regulatory knowledge type. Practically, the novel safety knowledge type may destabilize the existing conceptual space of safety politics and liability allocation patterns.

Keywords Robots · Autonomous vehicles · Simulation · Regulation · Regulatory knowledge

1 Introduction

Mobility robots will soon be among us, maybe. Autonomous cars may roam our streets, drones zip by delivering groceries, and autonomous ships navigate the seas. If the robots arrive, they will have to be regulated, at the very least for safety reasons, to ensure that the robots do not kill or maim or inflict property losses at an unsustainable pace.

While seemingly important for our robot futures, robot safety regulation has garnered surprisingly little interest in the scholarly literature (Weng et al. 2018, pp. 614–618). Only a few attempts at analyzing the strengths and weaknesses of existing safety instruments and arrangements in robotic contexts exist (Koopman 2018; Nakamura and Kajikawa 2018; McDermid et al. 2019; Lee et al. 2022). Similarly, although scholars have recognized that robots will have multiple technological features that appear legally disruptive, accounts outlining what regulatory techniques

regulators could use to ensure robot safety essentially do not exist (Leenes et al. 2017; Koopman and Wagner 2018; Koopman et al. 2019; Cf. Kouroutakis 2020; Hansson 2020; Bäck 2021; Petit and Cooman 2021; Ivanov et al. 2022).

This article attempts to address the research gap and chart the space of near-future mobility robot safety regulation, the possible safety regulation transformations that robots may trigger, and their downstream consequences for regulatory theory and practices. In contrast to many other areas of technology regulation (Bennett Moses 2017), the rationale and objectives for robot safety regulation are relatively clear-cut. The regulators aim to ensure that accident costs—material or immaterial—are sustainable.

While the orientation of the article is practical in the first six sections, Sects. 7 and 8 move to a more theoretical level, arguing that the simulation-based regulatory approach will trigger significant changes in the production of regulatory safety knowledge, regulatory politics, and responsibility practices. These developments are vital to understanding the implications of the transformations that mobility robots may initiate in law.

✉ Mika Viljanen
mika.viljanen@utu.fi

¹ Faculty of Law, University of Turku, Turku, Finland

The article discusses robot safety regulation in a particular setting. It is concerned with exploring the methods regulators could use to ensure the safety of a single mobility robot—an autonomous car, ship, or drone—as a stand-alone object operating independently without direct human involvement or available moral crumple zones (Elish 2019). The article further does not discuss how smart or intelligent mobility systems (Vöckler and Eckart 2022)—as composites of mobility platforms, infrastructures, and traffic governance—and their safety could or should be regulated. In addition, it does not discuss the complexities of mixed traffic Cheng and Zelinsky (2001).

Section 2 reviews the state-of-the-art *mobility robot navigational cognitions*, i.e., the technological assemblages that allow mobility robots to operate without direct human involvement. The objective is to lay out an empirical grounding for identifying the technological trajectories that near-future robot safety regulation will likely have to tackle. The article proceeds as follows.

The section uses maritime autonomous surface ships (MASS) as the primary technology mooring point. It is important to note that while MASS navigational cognitions will undoubtedly have their quirks, they share the essential functions, core technologies, and architectures with the cognitions used in other mobility robots, allowing for a generalizable discussion of the possible regulatory transformations. In addition, the article is primarily concerned with the modality of future regulation, not its implementation within the different regulatory bodies such as the UN Economic Committee for Europe (and vehicle regulatory forum or the International Maritime Organization).

In Sect. 3, I review what technological features of mobility robot navigational cognitions may give rise to regulatory challenges arguing that many existing accounts are misguided. Section 4 briefly outlines the instruments in the current safety regulation playbook. Section 5 tracks the difficulties of mobility robot complexity to the existing safety regulation instruments. The discussion focuses on rules-based and performance-based regulation, the two most promising candidates.

In Sect. 6, I argue that a simulation-based regulation will be necessary for ensuring mobility robot safety and outline its key design parameters.

Section 7 attempts to theorize simulation-based regulation by focusing on how simulations produce safety knowledge and what kind of knowledge they create. Section 8 argues that the new type of safety knowledge may destabilize the established space of safety politics and liability allocation patterns.

2 MASS navigational cognitions in a nutshell

Mobility robots are *cyber-physical systems*. The systems have two functional components: a *computational sensory cognition* first acquires information on its environment, processes it, and produces outputs (Hayles 2017). The physical platform second actuates the processing outcomes into real-world changes. In contrast to traditional machines, the systems integrate computation with physical processes: "embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa" (Lee 2008).

Of the two functional components, the cognitions are the real novelties (Bennett Moses and Gollan 2015; Behere and Törngren 2016). They will replace humans as operators, fundamentally reconfiguring how cars, ships, and drones operate and shifting risk generation from human-dominated to technological processes in the cognitions. As future primary accident risk drivers, the cognitions will be the focal points when regulators design future regulatory frameworks for mobility robots.

What will the cyber-physical cognitions be like? In the following, I will use an *autonomous navigation system* (ANS) of an uncrewed maritime autonomous surface ship (MASS) operating in autonomous mode without human supervision as an example.

In line with other mobility robots (on autonomous vehicle architecture: Behere and Törngren 2016, p. 139; Liu et al. 2017; Wang et al. 2020, p. 5; Ahangar et al. 2021), an MASS ANS will likely contain two functionally interlocked subsystems. ANSs will first have a *situational awareness* (SA) or perceptive system that generates the ship's situational awareness. Second, the cognitions will include a *navigational planning and decision-making* (NPDM) system that plans and executes the ship's navigational actions (for MASS: Ringbom et al. 2020).

The SA system comprises numerous sensors, computing infrastructure, and software components that facilitate the processing of sensor data and the generation of a digital map of the ship's surroundings and internal state.

A typical MASS sensor setup consists of five blocks. Fixed visible light and thermal cameras and robot pan-tilt-zoom cameras capture imagery for detecting and labeling objects in the ship's vicinity. Lidars, radars, and sonars generate data for spatially mapping the vessel's external environment. The sensors help fix various objects' distances, movement vectors, and sizes. Inertial navigation systems, global navigation satellite systems, and local position reference sensors provide data on ego vessel location. Automatic identification system transceivers, VHF

radio, and sound sensors, in turn, provide data on the site of other vessels. Machinery, environmental, and computing hardware sensors produce data on the vessel's internal state and its external environment. In addition to the sensors, the SA system contains a significant computing infrastructure with general-purpose and tailored processors Ahangar et al. (2021).

In addition to hardware, SA systems also rely on software components. The software includes signal-processing algorithms for preprocessing sensor data feeds. Another layer of algorithms extracts spatial and semantic information from the preprocessed sensor data. Sensor fusion algorithms reconcile the previous layers' partial spatial and semantic maps into a single coherent digital map (Ringbom et al. 2020, chapter 1). While some SA algorithms may originate from standardized off-the-shelf libraries, many will likely be proprietary and have a machine learning provenance. For example, image recognition algorithms crucial to producing semantic information on the MASS external environment are typically neural networks trained on proprietary image sets. Some spatial analysis algorithms (Gold 2016) may also derive from machine learning processes (Guizilini et al. 2021).

The navigational planning system, in turn, consists of computing infrastructure and various software components. The computers are likely standard off-the-shelf equipment, while most software components are proprietary. Some of the algorithms may be machine learning-based, and various path finding and optimization algorithms with no machine learning components implement the brunt of navigational planning. However, object trajectory models used to forecast the actions of other traffic participants may have a machine learning provenance (Ringbom et al. 2020, chapter 1).

To summarize, an MASS navigational cognition, like equivalent cognitions in other mobility robots, consists of a massive array of physical equipment with varying pedigrees. Manufacturers will likely source most equipment from vendors but build some in-house. The same applies to software components. Manufacturers will source software both from commercial algorithm libraries and develop some internally. The algorithms will be varied, ranging from standard programming code to complex machine learning-based neural networks.

3 Robot cognitions as safety regulation targets

Identifying safety intervention targets and feasible measures is critical to designing safety regulations (van Kampen et al. 2023) and mobility robot cognitions (Leenes et al. 2017). Three intertwined technological features or metaphors in the literature discussing possible future regulatory interventions

(Leta Jones and Millar 2017) often feature prominently. Scholars think that *robot autonomy*, the *robot's capability to learn*, and the *unexplainability of robot code* affect how regulators could ensure mobility robot safety. In the following, I will argue that when deployed to orient regulatory measures, the two features misinform and misguide the discussion. At the end of the section, I propose a new typology for understanding robot technological features. I argue that the robot's inherent technical complexity, non-linear performance dynamics, and the complexity of the environments that robots will navigate give rise to emergent behavior that future safety rules for mobility robot cognitions should address.

3.1 Autonomy and learning

Detailed definitions of autonomous behavior in robotic contexts vary and are the subject of heated debates (Asada 2020). Robots and other artificial intelligence applications often attract the *_autonomy_label* in the legal literature (Chopra and White 2011; Karnow 2016; Wendehorst 2020). A considerable fraction of legal accounts hang robot autonomy not only on the robot's ability to adapt to its environment without direct human involvement (Leenes et al. 2017)—a typical refrain in robotics (Hellström 2013)—but extend it further toward *autonomy as agency* (Leenes et al. 2017, p. 5). The autonomy as agency account enacts a strong conception of autonomy typically reserved for humans in philosophical accounts (Haselager 2005). These *strong autonomy robots* (Gutman et al. 2012) eclipse their designers' intentions, react to their environment in a non-determinist fashion, or independently make real ethical or moral choices.

The crux of the learning narratives is straightforward. The narratives imagine that the robot cognitions constantly learn from the new data they encounter during use (Matthias 2004; Gless et al. 2016; Barfield 2018). Accounts that stress strong robot autonomy also often deploy the learning trope to account for robots' distinctive features. A ship, for example, might observe a novel maneuver during a voyage and incorporate it into its navigational algorithms (For MASS: Porathe 2019; Røsag 2020). If robots learn this way, their technological composition will inevitably become fluid and unpredictable during use, with humans losing control over how the now unstable robots behave.

For the current state-of-the-art robotics, both framings appear misguided. Strong autonomy remains science fiction (Richards and Smart 2016; Benjamin 2022; Sprenger 2022). Soon, robots will likely remain stable amalgams of hardware and software that simply execute and actuate code. They will lack the non-deterministic agency accounts adopting the strong autonomy framing ascribed to them. AI components may be sophisticated, and the

interactions between the code, hardware, and the environment complex, but the robots will remain machines (Leenes et al. 2017).

Similarly, continuous in-use or *online* learning is (and hopefully remains) a pipe dream. The narrative appears to build on a fundamental misconception of how, where, and at what time scales robot cognitions and their machine learning components learn. The account glosses over the fact that machine learning algorithms are code that emerges when data scientists deploy machine learning methodologies on data sets (Shalev-Shwartz and Ben-David 2014; Faul 2019). These processes are energy-intensive and often temporally lengthy. Typically, testing and validation caps then off. Learning, thus, is not something that happens during operational use.

For example, the image recognition algorithms that emerge as products of supervised machine learning are stable during use. They learn while in labs. Engineers and data scientists first design the basic network configuration specifying the number of layers in the network and activation functions. Then, the actual training work begins. The developers load the network nodes with multiple sets of different weights and use training datasets to explore which weight configurations produce the best results. While the work is automated and methodologies such as backpropagation speed up the process (Traore et al. 2018, pp. 262–263), creating a well-functioning algorithm may take hundreds of hours of processing time and significant energy. Once a candidate algorithm emerges, it must be validated and tested to ensure that it works adequately on real-life data. Only after validation tests convince the engineers that the algorithm works well under real-world conditions is it approved for use and deployment. Thus, learning by robot cognitions is a system-level process controlled by the system's developers.

While state-of-the-art robots do not learn at the instance level or online during operation, automating machine learning operations could transform algorithm development into a continuous instance-level lifelong learning process (Chen et al. 2018). Whether lifelong learning makes sense as a practical proposition is an open question. The answer hinges on whether the advantages of using online deployment data to train the algorithms during operations outweigh the computational and energy costs of training and validation. The genuine risks of system decay and bad learning outcomes also weigh into the decision. More importantly, one should question whether allowing lifelong learning in safety-critical contexts is acceptable in the first place. Deploying a lifelong machine learning approach in a safety-critical context is an extremely high-stakes bet on the automated learning and automated verification methodologies with no safeguards if and when the robot learns wrong things.

3.2 Unexplainability

Algorithmic unexplainability is a third trope that has attracted significant attention (Wortham and Theodorou 2017; Felzmann et al. 2019). Commentators have lamented that robot cognitions are opaque and contain unexplainable, and often contain intractable algorithms (Wachter et al. 2017; Umbrello and Yampolskiy 2022). The diagnosis is, per se, correct. Most deep learning components will remain largely uninterpretable or even inscrutable (Mittelstadt et al. 2016).

Technologies such as post hoc *interpretability tools* (Vale et al. 2022) increase algorithmic transparency for developers. They help developers understand what factors drive algorithmic performance and debug systems, and increase their understanding of system performance. Nevertheless, explainability is likely a misguided focus for safety regulation as it does little to promote safety objectives. The reason is simple. Explainability has significant immediate intrinsic and instrumental value in many use contexts but less so in robotics. To illustrate: a broad consensus exists for thinking that decisions affecting individuals' rights and obligations should be explainable, in particular, to justify the decision, facilitate contestation, and maintain and uphold the legitimacy of the decision-making processes (Vredenburg 2022). In robotic safety contexts, none of the factors are present. Robot-induced injuries often cannot be meaningfully contested as they are typically irreversible, and explanations do little to justify the damage or injuries the robots inflicted. Instead of focusing on explainability, safety regulation should seek to prevent injuries. Explainability is, however, a poor instrument for pursuing that end. Even if the cognitions were fully transparent and explainable, their internal and environmental complexity would make explainability-reliant efforts to predict their behavior and weed out undesirable outcomes largely futile, as argued below.

3.3 The real safety concerns?

What, then, should be the central safety concerns? Robot safety regulation should be able to ensure that robots do not create excessive safety risks. Achieving this end requires that the regulatory interventions enact robots as clusters of features that, in reality, affect the safety risks and that regulators can leverage to manipulate the risk levels. The often-discussed robot features do not allow this. Robots first are not autonomous other than in the sense that they can function without direct human intervention. They do second not (and should) learn during use. Third, explainability will not help in controlling the risks.

What remains is a relatively dull control surface. Robot performance and safety outcomes seem driven by how the robots perform as machinic assemblages composed

of hardware and software components. Here, trajectories diverge. Hardware regulation has a track record that spans decades. We know how to make mechanically reliable ships, cars, and aircraft. The new challenge lies in managing the complex interplay between the sensory system, decision-making software, and the open world that mobility robots navigate.

As recounted above, mobility robot cognitions will likely contain hundreds, if not thousands, of hardware and software components. The components weave a complex, entangled web with countless interdependencies. Sensor performance affects labeling processes which in turn affects decision-making. At the same time, environmental variability is near-infinite in the open world's robots operate. A staggering number of possible unique sensor input states exist. Action path dependency contributes to variability by temporally expanding the universe of relevant inputs. Surprising, unanticipated confluences of expected and unexpected environmental factors are bound to emerge. Teslas, for example, have mistaken the moon for a yellow traffic light (Ramey 2021). Under these conditions, the fundamental problem of robot safety work and regulation is simple. Regulators must ensure that technologically complex cyber-physical systems operating in complex, endlessly variable environments will consistently and reliably do the right thing, year in and year out, in all conceivable conditions and settings.

This complexity alone rules out *formal verification*. Even if the systems contain no unexplainable machine learning code is used (Törngren and Grogan 2018), there is currently no feasible verification method available. Scientists cannot adequately specify test cases and analyze and verify the robot cognition responses within every unique set of possible inputs (Michael et al. 2021). Machine learning algorithms exacerbate the problem exponentially by adding a source of uncertainty. Machine learning algorithm outputs are typically non-linear: an infinitesimally small change to the input can trigger a drastic change in the output (Lipton 2018). The non-linear performance dynamics change the game. The interdependencies between various system components may alone render system outputs hard to anticipate. However, when the non-linearity of machine learning components kicks in, it will push mobility robots toward a de facto indeterminacy.

Even if de facto indeterminate, robots will remain machines and a far cry from the intractable, opaque, multilayered, and unstable constitutions that humans (Morin 2008, pp. 78–81) as thinking autonomous entities exhibit. Nevertheless, because of their inherent technological complexity, non-linear performance dynamics, and the complexity of the environments they will navigate, the robots will exhibit surprising, emergent behavior (Cf. Calo 2015) even if stable objects during use. Safety will be a function of the

designers' capability to weed out undesired behavioral emergence and turn it into dull predictability.

4 The safety regulation playbook

As regulators respond to mobility robot behavioral emergence, the critical question is will the robots destabilize the existing safety regulation approaches? To set the stage for charting future robot safety regulation trajectories and their consequences, I will first review what the traditional *physical system safety regulation* playbook looks like.

The traditional safety regulation playbook has two primary layers. Regulators often engage in *artifact regulation* and seek to affect the target artifacts' technological composition. Second, *liability rules* provide a buttressing layer that allegedly incentivizes manufacturers to implement safe designs by imposing civil or criminal liability for injuries and losses caused by unsafe designs (Polinsky and Shavell 2010). In the following, I will focus on the artifact regulation layer.

4.1 Rule-based standards

Regulators and their standardization organization proxies have multiple approaches when designing the artifact regulation layer.

The oldest trick in the book is to deploy a rule-based regulatory approach. In rule-based regulation, regulators set detailed binding technical specification standards (Decker 2018). The technical specification standards intervene directly in the artifacts' technological makeup and force manufacturers to adopt particular designs. These instruments are often further fleshed out in countless product standards (on standardization: Yates and Murphy 2019) that constitute the traditional backbone for, for example, machinery, vehicle, aircraft, and marine safety regulation.

4.2 Goal- and output-based standards

For complex artifacts, the rules-based approaches often reach the limits of their utility (Coglianese et al. 2003; May 2003; Decker 2018). On similar developments in regulating organizational behavior: Black and Baldwin 2010). There are several reasons for this. First, the space of feasible designs may be too vast to allow the regulators to specify the best approach. There might also be an information asymmetry between the industry actors and regulators, as the regulators often lack the necessary technical knowledge and expertise to choose alternative designs. Third, the technologies might move too fast for detailed rigid rules to make sense. Under such conditions, regulators often deploy a goal or performance-based regulatory

approach. The approach circumvents the problems of the more rigid methods by not laying out detailed specifications for the expected artifact features. Instead, it establishes performance targets that the artifacts must meet. The approach has several benefits. It gives leeway to the regulated to determine which designs to implement and coopts the industry's expertise to choose the best means for securing the desired outcomes (Coglianese et al. 2003).

4.3 Design process and management-based regulation

If both rule- and goal- and outcome-based approaches lose traction, regulators often resort to the design process- or management-based strategies. In these strategies, the regulatory focus shifts to artifact design processes. The regulators set standards for organizational structures, outline workflows, articulate foci for attention, create checklists, and frame mindsets. Instead of relying on direct behavioral prohibitions and commands as is usual in legal settings, the strategies leverage the normative power of processes (Coglianese and Mendelson 2010; Chiu 2015).

The hope is that an existence of an organizational process brings about the desired outcomes, even if rules do not articulate the results. For example, Christine Parker (2009) has argued that regulators design sustainability-oriented corporate governance interventions to induce the birth of "corporate consciences" in target firms.

The approach remains relatively rare in legislative instruments. Still, it has been rapidly proliferating from standardization body practice (Yates and Murphy 2019, pt. III) to legislation (for an early example: European Economic Communities 1993). The EU AI Act (European Commission 2021a) will incorporate the approach into the AI regulatory landscape, although the Act, due to Article 2(2), will likely not apply to mobility robots.

While the design process and management-based regulation may be an attractive template for robot regulators, its utility in safety-critical contexts seems dubious, given the failure dynamics in software development. As software flaws affect the entire installed base, they can trigger significant temporally concentrated accident cascades. With the threat of such cascades on the horizon, building and retaining public trust in mobility robots will likely require pre-deployment safety assurance methods and a set of performance-based standards all designs must meet. Even intensive in-use monitoring and reactive supervisory measures such as recalls (a robot equivalent to pharmacovigilance) have problems. Possible recalls would temporarily halt the use of critical high-value assets, making robot owners likely unhappy to see their robots shelved.

5 Would existing tools work?

In the following sections, I will review the challenges robots impose on rule-based and performance-based regulatory approaches. The discussion will focus on the rule-based and performance-based approaches as the design process and liability approaches seem inadequate in safety-critical contexts.

5.1 Rule-based robot safety rules?

Rule-based robot safety regulation will probably fall into trouble when encountering mobility robots. As recounted above, mobility robot cognitions will be complex assemblages of software and hardware with multiple alternative configurations likely offering equivalent performance. For example, various manufacturers' ANS sensor configurations and algorithms may all be capable of navigating a ship. Under these conditions, mandating that manufacturers use a restricted set of designs will likely result in sub-optimal outcomes and stifle innovation in an immature and fast-moving field.

Two system software traits, in addition, will, in particular, contribute to making rules-based approaches infeasible. First, the cyber-physical system design philosophy differs from traditional safety engineering methodologies. Many established safety engineering tricks that work well in physical systems, such as redundancy (Clarke and Hollister 2010), safety and design margins (Eckert and Isaksson 2017), and derating, are primarily useless for the software components. Engineers must design software flawlessly for it to function adequately. Second, mandating design standards for the machine learning components vital to ANS performance is also senseless. As the algorithms are creatures of training and validation data in important respects, the utility of rule-based approaches will be severely restricted.

Nevertheless, regulators will likely deploy rule-based standards for ANSs and their hardware components. Many of the rules will probably be high-level declaratory statements in the vein of Article 15 in the proposed European Union AI Act. The provision calls for all high-risk AI systems to "be designed and developed in such a way that they achieve [...] an appropriate level of accuracy, robustness and cybersecurity" (European Commission 2021).

In addition to the high-level performative proclamations, sensible pockets for rule-based standards exist, such as sensor array composition and performance and operational design domain (ODD) ontologies. ODD ontologies emerge as crucial components of future mobility robot regulation. As humans currently perform the brunt of

sensory work and human sensory capabilities are primarily biologically determined, regulators have largely been able to eschew defining what objects mobility actors should be able to detect and act on. With the new technological sensory systems, a need for object detection requirements seems likely to emerge as biological constraints disappear.

5.2 Performance-based robot rules?

Performance-based regulation (PBR) will likely fare better than rigid rule-based approaches in robot contexts. PBR allows regulators to articulate safety standards despite the technological heterogeneity, informational asymmetries, and an unsettled, fast-moving technological field, all features of the robotics landscape. In addition, the problems arising from machine learning components' opacity, inscrutability, and unexplainability will lose at least some relevance as performance-based regulation treats its target artifacts as black boxes (Colaner 2022).

However, trouble may, nevertheless, be brewing. Mobility robots seem likely to push the limits of established PBR schemes. Note that performance-based standards typically contain three structural components. PBR standards first establish a *metric* for desired outcomes. After regulators have articulated a metric, they will outline a *verification methodology*, typically a pre-use test or an in-service monitoring scheme, for gauging artifact performance (Coglianese and Nash 2017). Finally, regulators must specify a target level for acceptable performance. The *performance targets* serve as proxies for the objectives underlying regulatory designs.

The three aspects often make up an interwoven knot for complex artifacts embedded in complex environments. Take the EU regulations for vehicle emissions and its 103 EU Official Journal pages of dense technical prose (European Commission 2017) on "real driving emissions" as an example of a highly complicated performance-based regulation scheme. The metric arose naturally in the Real Driving Emissions (RDE) protocols. The regulatory objective is to control pollutant emissions. Thus, tailpipe pollutant quantities emerge as the self-evident measurement target. However, measuring a car's "real" emissions is tricky. The emissions are not a smooth object with only one causal engine-related factor but a function of many complicated and interdependent causalities. The RDE regime seeks to navigate the causal mess by focusing on a "realistic" driving scenario to capture and average all the different causal drivers for emissions. These include engine operating mode, speed, speed variations; vehicle size and weight; fuel delivery and emissions control systems; fuel type; and environmental conditions (Sinha 2007, p. 255). As a result of this boundary work, the Euro6 metrics and performance targets remain tidy objects (Williams and Minjares 2016, p. 4), but

the verification methodology balloons into an exceedingly complex contraption that ultimately proved easy to game (Schiermeier 2015; Jung and Sharon 2019).

If the RDE design is already byzantine, the protocols for testing mobility robots will likely be several orders of magnitude more complex. The primary challenges arise from the characteristics of the processes ANS safety verification arrangements should control and cause problems in setting the regulatory objectives and ensuring adequate safety.

While a tricky and contentious subject, mainstream safety science understands safety as the absence of adverse events. Things are safe when nothing goes wrong, rendering safety a succession of non-events (Hollnagel 2014). Demonstrating the future absence of undesired events constitutes the fundamental challenge for performance-based safety regulation schemes. To do that, regulators must have access to the artifact's future and inspect them.

Should we transport the RDE into the safety frame, the safe future with no excess emissions becomes tangible as the "real driving" scenario enacts the car's present state. Here, the crucial design choice emerges into view. The regulators settled on a single stylized future to represent all possible futures. This approach is not feasible for robot safety schemes. Remember first that emergent behavior is the distinctive feature of mobility robot operations. The systems are complex, operate in high-variability environments, and exhibit non-linear traits. Enacting and collating a stylized future and building an isolated test to simulate the future would fail to account for all three key mobility robot features. Complex, non-linear responses to highly varying inputs require more than one stylized future to explore. The same dynamics will make worst-case-oriented testing infeasible, as anticipating what conditions challenge system performance is nearly impossible. Safety event dynamics further add to the futility of limited localized tests. High-impact accidents are rare tail events that actualize in a small fraction of all operational scenarios (Kröger and Ayoub 2022, pp. 142, 142). To explore the accident patterns, regulators must explore the high-risk low-probability tails of the likely environmental condition distributions.

Consequently, ensuring that undesirable events do not happen will not occur requires a massive amount of tests to be run. The tests must explore the complex internal ANS dynamics, account for the virtually limitless variation in environmental conditions, and produce evidence of adequate performance in the tail ends of environmental distributions.

Designing and building a testing regime capable of generating the necessary scenario volumes will not be trivial. A 2016 RAND study on securing autonomous ground vehicle safety illustrates the difficulties. The authors calculated what incident-free driving volume would demonstrate that self-driving cars were safer than human-operated vehicles. The analysis suggested that enough evidence for a 95%

confidence level would emerge first when an autonomous vehicle had run for driven billions of miles. Even with hundred vehicles in constant operation, the tests would take approximately 500 years (Kalra and Paddock 2016, p. 10).

Traditional testbed schemes and even extensive real-world testing are, thus, unlikely to generate sufficient evidence to demonstrate robot safety. The scenarios would accumulate too slowly to facilitate feasible commercial robot deployments. The only way forward is to move to the virtual world and simulation-based verification methodologies (Figueiredo et al. 2009; Ebert and Weyrich 2019; Kröger and Ayoub 2022).

5.3 Simulation-based robot regulation?

As limited real-world tests will likely fail to generate sufficient evidence of operational safety, simulations seem destined to become the inevitable cornerstones of future safety regulation regimes. Regulators will have to supplement the real world with its virtual twins.

The UNECE Inland Transport Committee World Forum for Harmonization of Vehicle Regulations, a primary venue for international regulatory cooperation, reached this conclusion in its new assessment framework for automated vehicles published in June 2021. The Forum argued that simulations would provide "an indispensable tool to verify the capability of the automated system to deal with a wide variety of possible traffic scenarios" (World Forum for Harmonization of Vehicle Regulations 2021, p. 9).

As cyber-physical entities, robots are opportune targets for simulation-based performance verification. Several complications may still emerge.

As recounted above, the first challenge lies in the testing volume required for demonstrating safety. Gaining confidence in a mobility robot's capability to navigate everyday and challenging situations will take at least hundreds of thousands of simulation runs. The simulations should cover a sufficient selection of common simple traffic scenarios in various environmental conditions. However, they should also include the more freakish outlier scenarios, such as erratic pedestrians traversing roads at unanticipated locations (National Transportation Safety Board 2019) and wild animals. A good example is that such commonplace objects as parked emergency vehicles forced Tesla to recall over 300 000 cars in February 2023 (Nyce 2023). Designing and generating the scenarios will take significant expertise and investments in personnel and the simulation platforms (Ringbom et al. 2020).

Building the scenario inventories and simulation platforms will likely prove a complicated process. Informational asymmetries and the lack of expertise and resources will likely push the regulators toward self-regulation-based approaches. The constraints may force the regulators to

rely on the industry and its proprietary platforms. The fragmented and uneven jurisdictional geography of mobility robot production may strengthen the self-regulation trend. While the UNECE Forum and IMO, for example, are essential regulatory cooperation venues, no global regulators with the budget to build the simulation environments currently exist. Consequently, the simulation-based performance verification landscape may be patchy, with regional and national actors adopting varying approaches and facing severe resource constraints.

Running tests on SA generation may offer further significant technical challenges. The navigational planning and decision-making systems could build on standard dynamic map inputs. Even if input templates varied across systems, transposing the traffic scenarios into maps used in NPDM systems will probably be relatively easy. SA systems, in contrast, may significantly differ across ANS manufacturers and ANS versions as to their configurations and data feeds. As ANS performance is a function of both the SA and NPDM systems, simulating NPDM systems alone would likely lead to an insufficient understanding of ANS safety. To explore ANS performance, regulators must understand how the SA system performs under various conditions and interacts with the NPDM components. However, most simulation environments still focus on mapping navigational planning and decision-making (Yang et al. 2021), while modeling sensors remains fraught with technological difficulties (Schlager et al. 2020).

6 Theorizing simulation-based regulation

Above I argued that simulation-based regulatory approaches are critical parts of any viable regulatory approach to securing mobility robot safety. Notably, the techniques are relative novelties in the regulatory landscape and remain undertheorized. In the following, I will argue that the emergence of simulation-based safety regulation approaches will trigger significant transformations in regulatory knowledge production patterns, temporalities, and types of knowledge. I will start by discussing regulatory knowledge production and its temporalities.

6.1 Producing knowledge of robot safety

During recent decades, science and technology studies have blossomed to trace how normative concerns shape (or co-produce) scientific knowledge and technological artifacts (Jasanoff 1995a). Similarly, all lawyers are at least instinctively aware of the peculiar but often undertheorized way courts create and shape facts in legal flows (Latour 2009). In addition, significant research has gone into understanding the construction of expertise and

"regulatory science" in legal and regulatory processes (Jasanoff 2012, 2018) and mapping the production of knowledge on risks, in particular for chemicals, drugs, and food (Demortain 2011). These advances, however, seem to struggle when transplanted into robotics contexts. In the following, I will outline the material regulatory formations that generate knowledge about artifact safety and the regulatory temporality embedded in rules. To do that, I will draw inspiration from Michel Foucault's early work on truth production in judicial and quasijudicial practices (Foucault 1977, 2002, pp. 1–91).

In his 1973–1974 Brazilian lectures on truth and law, Foucault traced the diachronous emergence of *inquiry* and *examination* as legal truth production regimes. The inquiry was a particular technology of truth production that emerged during the Middle Ages from administrative and pastoral practice to constitute the primary technique of judicial truth production. According to Foucault, the inquiry as a practice template put a "form of knowledge" in place. In the template, the sovereign's representatives gathered notable people together. The representative then asked the people what they knew about past events, listened to their testimonies, and parsed together the truth. (Foucault 2002, pp. 49–51).

Three directions are crucial to understanding the inquiry as a knowledge production template. First, the inquiry is part and parcel of a particular (pre-)governmentality (Foucault 2008, 2014). The inquiry is a crucial technology of power in facilitating the exercise of sovereign power. The inquiry allows the sovereign to react to and avenge rule infractions and reestablish the primacy of its will. Second, the inquiry as a technique of knowing has a particular temporality. As Foucault argues, the inquiry "extends the actuality" of the past by "transferring events from one time period to another" and making them subject to the sovereign's gaze even if already in the past (Foucault 2002, pp. 49–51; Hunt 2012). As a material formation, inquiry third rests on human cognitive capabilities for remembering, recalling, and judging and the specific rules that govern assessing testimony.

Examination, the second of Foucault's forms of knowledge, emerged as a technique of disciplinary governmentality seeking to manage and increase the strength of a population. The examination subjects its targets to a gaze that seeks to describe their various traits to "maintain [them] in [their] individual features, in [their] particular evolution, in [their] own aptitudes or abilities, under the gaze of a permanent corpus of knowledge. Conducting the subject's conduct then becomes possible through the "constitution of a comparative system that made possible the measurement of overall phenomena, the description of groups, the characterization of collective facts, the calculation of the gaps between individuals, their distribution in a given 'population'" (Foucault 1977). Instead of making the past visible to the sovereign's gaze and its reactive interventions, examination as a

technique allowed the disciplinary and biopolitical actors to peer into the futures of its subjects and prime interventions into the subjects' constitutions.

While Foucault's legacy in legal studies mostly arises from his analyses of disciplinary power (Hunt 1994), governmentalities (Rose 1999; Rose et al. 2006), and biopolitics (Lemke 2019), the legal forms of knowledge Foucault outlined still constitute the basic formations of knowledge production in legal processes. Courts gather evidence and testimony to recreate pasts. Quasijudicial processes, such as parole assessments, examine the subjects' traits to peer into their possible futures. Consequently, while Foucault's work remains relevant in many corners of legal knowledge production, the two forms of knowledge appear strained as heuristics when used to explain how safety regulation and other regulatory (e.g., Anwar 2020) practices co-produce knowledge about technological artifacts.

Due to its temporality, the inquiry as a template fits ill with safety verification procedures. The technique looks backward, not forwards. It only facilitates reactive interventions that can govern futures indirectly. Examination, however, fares better. In rule-based technology regulation, the knowledge production about the artifact travels between the general regulatory science "corpus of knowledge" on safe designs that the technical rules and standards represent and the particular observable traits of a technological artifact. The corpus provides the anchoring point for safety. It is an object co-produced in negotiations between the industry, scientists, lawyers, and politicians representing a frail and temporally discrete state-of-the-art (Jasanoff 1995b; Irwin et al. 1997). This knowledge underlay allows the legal production of knowledge about particular artifacts. The examination transforms observed artifact traits into writing (Cf. on clinical drug trials: Will 2007). The regulators compare the signs to those that stand in for safe designs. If there is a match, the future is secure.

As already discussed, performance-based regulation is a response to changes in regulatory knowledge. If the regulators do not have access to a corpus of knowledge on safe artifact constitutions or the corpus is evolving rapidly, there are articulable safe futures. To respond, the regulators shift regulatory efforts to developing meta-level tests to manage the knowledge void. The tests serve as the proxies for knowing what artifacts are safe, allowing the regulators to retain control over what things emerge while remaining agnostic over their compositions.

In the transition to performance-based regulation, the material formations of knowing change, but the temporality of knowledge remains. The techniques still allow regulators to peer into the future by inspecting the present state of an artifact. However, the artifact is no more the direct subject of the inspection. Instead, inscriptions reporting the test results represent the artifact, while the tests stand in

for the knowledge of safe artifact constitutions. The tests enact small-scale futures to verify safety. To facilitate this transformation, the material formations of legal knowledge production expand outwards, concretely embedding material assemblages of equipment and facilities (Latour 1994) into the formerly language-dominated legal flows. The regulators will certify that the future will be safe if the artifact successfully passes through purpose-built material formations (On clinical trials: Will and Moreira 2010). For example, a lane-keeping system in a car must perform adequately on a purpose-built test track with particular intricately detailed physical properties to generate signs indicating the presence of a safe design (UNECE 2020).

The transition to a simulation-based regulatory paradigm, however, entails another reconfiguration. Whereas the performance-based regulatory schemes with their tests bring the future into the present as test outcomes, simulation-based strategies seem to enact a more decisive temporal shift. Simulations enact massive arrays of virtual futures. Instead of building environments where the artifacts can go through localized trials, simulations as a form of regulatory knowledge require a capability to enact the futures en masse. To do this, the regulators need virtual models of the artifact and the world the artifact inhabits. Knowing the future is possible if one has access to a complex industrial-scale information technology assemblage, computing power, energy, digital twins of the laws of physics and countless objects, and an endless succession of possible futures.

6.2 A new type of robot knowledge

How to triangulate the characteristics of this new type of regulatory knowledge? Insurance emerges as an apt point of comparison. Actuaries gather years of data on how many insurance events have happened and how much the insurer had to pay in indemnities. As insurers use actuarial techniques, they gain discrete visibility into the "laws of the physics" of the insurance pool. The laws are probabilistic: data on how indemnity payments total vary yearly indicates the variability of the insured process. Armed with data on the pool's past, actuarial mathematical tools, and a conviction that the future will be like the past, actuaries can create—often less than scientifically founded and rigorous (Doyle and Ericson 2010, p. 237)—futures. The futures are particular: the probability distribution of indemnity payment totals for the next policy year (E.g., Smith and Kane 1994) stands in for the life of the pool. With this information at hand, the future becomes actuarially manageable. The insurer can quantify the risks and ultimately share or move them.

Simulations produce a state of knowledge comparable to that of an insurer. There are, however, marked differences. While both technologies are future-oriented and enact

statistical futures, simulations do not track how a mass of individual actors behaves as a collective in simultaneous enactments of multiple possible future world instances as insurance technologies do. Instead, simulations reverse the dynamics of future generation by focusing on a particular actor, a single robot, and its behavior across multiple simultaneous alternative locations of a world. Two transformations take place.

First, in robot simulations, data scientists enact a mass of alternative local worlds, each with distinctive peculiarities, whereas insurance simulations enact multiple versions of a global world for the pool. Second, data scientists collate the robot simulation outcomes to make up the distribution of the future performance of the robot across a selection of locations. The contrast to insurance simulations is clear. In insurance contexts, simulations make alternative future worlds, but robot simulations create a mass of local world instances that, as a collective, stand in for *a* world. Third, the future in robot simulations is a function of the traits of a particular individual actor, not the outcomes of a collectivized process with tens of thousands of pool members whose specific features can remain opaque.

This difference is crucial. The simulated robot worlds are synthetic snapshots of possible episodes in the life of the simulated entity. The emerging future is singular, particularized, and episodic, not average, totalizing, or undifferentiated, as in insurance. Robot simulation platforms need plugins to create a totalized view of overall robot performance and costs to escape their inherent particularity. However, with plugins, the extended simulation environment can create robot and robot fleet life paths out of the episodes and collate them to arrive at a totalizing knowledge of robot futures. For MASSs, for example, simulation scientists would, first, compose multiple realistic voyages and, second, aggregate the trips to fleet actions over expanded periods. Once the plugins perform these manipulations, simulated knowledge approaches the real but imagined collective worlds actuaries deal with.

7 Consequences of robot knowledge

I argued that simulations could produce insurance-like probabilistic and totalizing knowledge of the future. Importantly, this new knowledge production formation affords its users an unparalleled information resource that may have significant downstream implications. It may upend mobility system safety politics and cause other disruptions, particularly in liability allocation.

7.1 Disrupting safety politics

To understand the implications for safety politics, consider the safety politics of human-dominated mobility systems. In these systems, people are typically hard to govern objects and impervious to regulatory interventions. While studies suggest that training and informing people have clear impacts (Turner et al. 2021; Fisa et al. 2022), such interventions often fall short. As a 1960's New York traffic commissioner wrote, "as time goes on, the technical problems become more automatic, while the people problems become more surrealistic." (Vanderbilt 2008, p. 8). Many human risk factors remain beyond the reach of feasible interventions for practical or political reasons (Daniels et al. 2019). Consequently, the problems force policymakers to tolerate a certain background level of accident costs.

Second, when regulators design human-targeting interventions, their work is often like grappling in the dark with an unknown opponent. Regulators can make assumptions about the humans they engage, build theories about what drives their behavior, gather evidence, and devise regulatory interventions they think may be effective. However, even evidence-based policymaking is no objective science (Pawson 2006; Head 2008). While feedback loops may allow the regulators to assess and retool interventions, no Planet B's exist, at least within the time constraints of policy cycles. Typically, regulators will remain uncertain whether interventions can imprint permanent human changes. Sometimes, interventions also prove counterproductive (Friedman 2016). Consequently, regulators can rarely rely on anything other than vague guesses of the likely consequences of their actions when they make trade-offs between alternative regulatory options. As a result, the human-facing safety politics in human-dominated systems emerges as an eminently uncertain practice.

The policy-setting uncertainties seem prone to dissipate when mobility robots gain ground. Robots eliminate and suppress the tensions emanating from human agency's capriciousness and unruliness but create new uncertainties. Robots' machinic constitutions, however, make them easier to control than humans. Compared to humans, robots are dependable, compliant, and docile. They do not take stupid risks when they are young and male, get distracted by mobile phones, commit suicide, drink, use drugs, or deteriorate physically and mentally when reaching old age, at least not like humans do. They are, further, not resistant to reading instructions. Instead, robots, if hardware tear, wear and dirt, and instance-level lifelong learning excluded, perform day in, day out, as their machinic constitutions determine. Flaws in robot code, once identified, can be fixed. The patches also reliably stick. As this change occurs, a radical transformation in the degree of control regulators can have on mobility may be in the offing. Regulators may gain new policy horizons.

Given robot plasticity, simulation platforms could allow regulators to explore various alternative robot designs and analyze the outcomes of each design. If this happens, the informational basis for safety politics will transform. Instead of grappling in the dark with uncertain information (Alcaraz Carrillo de Albornoz et al. 2022), the new technologies could give regulators unprecedented visibility into the likely outcomes of their choices. Cost-benefit analyses could gain in accuracy, and policy justifications grow more detailed. Policymakers could set granular accident level targets for, for example, different road user groups and, importantly, have the leverage to enforce them.

The increased control, however, has several drawbacks. The transition toward simulation-based regulation could contribute to ushering in a code-driven law that "folds enactment, interpretation, and application into one stroke, collapsing the distance between legislator, executive and court" (Hildebrandt 2020). The transition to code-driven law "freezes the future," suppressing the flexibility and adaptability of the traditional textual law. Quantitative wretched choices also seem poised to multiply. While regulators already make explicit choices between lives, money, and comfort in their cost-benefit analyses (Florio and Pancotti 2022), the turn to code-driven law will multiply the decision points and make the choices increasingly explicit.

7.2 Disrupting liability allocation

However, with robots, the wretched choices would flow downstream to manufacturers. Speed selection decisions are an illustrative example. As speed is a crucial, if not the most critical, risk factor in road traffic (Carsten and Tate 2005; Elvik et al. 2019), selecting an appropriate speed is an ethical flash point in autonomous vehicle design. Adopting a conservative vehicle speed policy could ensure that the vehicles constantly adapt their speed to match road surface, visibility, and traffic conditions, and significantly reduce accident numbers compared to a more aggressive approach.

When and if injured parties contest AV speed selection decisions in liability trials, simulation data have the potential to destabilize the existing liability allocation patterns. The destabilizing potential emerges out of three dynamics.

First, simulations will likely create an unprecedented number of detailed article trails documenting risk imposition decisions, possibly making standard mobility robot product liability or negligence cases resemble the infamous Ford Pinto case (Grimshaw v. Ford Motor Company (119 Cal. App.3d 757, 174 Cal.Rptr. 348)). Claimants will likely have a heyday in court, drawing attention to the manufacturers' callous choices. Whatever the state of legal doctrines will be once they emerge, the dynamic may encourage manufacturers to adopt conservative policies.

The second pathway emerges in the EU Product Liability Directive (PLD, (European Economic Communities 1985)) rules. The Directive imposes strict liability for products that do "not provide the safety which a person is entitled to expect" (PLD Art. 6). The new type of knowledge disrupts the important "development risk defense" to which most EU member states allow manufacturers to appeal. According to the defense, manufacturers escape liability if "the state of scientific and technical knowledge at the time [the product was designed] was such as not to allow the unsafe feature to be discovered" (PLD Art. 7).

Simulations will shrink the space of inadvertent safety defects considerably. All scenario failures indicate a potential unsafety not covered by the development risk defense. For these scenarios, the manufacturers will lose the chance to claim that they did not know of system inadequacy. The shrinking development defense will likely contribute to pushing manufacturers toward safer designs. Still, whatever courts consider legitimate robot safety expectations will draw the final line between acceptable and unacceptable system safety.

A third pathway seems prone to shake the foundations of negligence-oriented liability rules, which typically govern robot accidents if the courts cannot apply the product liability framework. The EU legislative landscape is fragmented as the field is not harmonized, and the member state rulesets differ. Most jurisdictions have, however, settled on standard frameworks for assessing whether the defendant's conduct should attract liability. The defendant is first liable if they caused the loss wilfully. Wilful conduct typically requires that the defendant either intended to cause a particular consequence, was certain that it would arise or understood it to be a probable consequence of their action. The second option, negligence, is harder to frame. For example, under Finnish law, three options exist. First, a defendant's conduct was negligent if they intended to cause harm, knew that such harm was a reasonably certain consequence, or understood that the harm was a probable consequence of the conduct. Second, a court would similarly find the defendant liable if they breached a relevant safety rule. Third, under a variation of the Hand formula, the defendant's conduct is negligent if they failed to take reasonable precautions considering the costs of possible safeguards, the gravity of the injury or the losses, and the likelihood that they were to occur (Hemmo 2006).

In setting speed policies, each alternative has a weakness. First, as usual, relevant safety rules are unlikely to exist in sufficient numbers to facilitate norm-based assessments. Second, simulation-induced knowledge could throw a wrench into legal intentionality analyses. Design decisions collate and compress causality into a single point in time. When the designers settle on a policy design, the consequences of that decision expose a large group of people and

other entities to the risk the decision created. If successfully conducted, simulations allow the designers to know the aggregate outcomes of the risk processes they create. When choosing what design option to adopt, they make conscious system-level risk choices where the risk of injury is low for each exposed party but high for the collective. In this sense, mobility robots resemble traps. They have a known potential for causing injuries, but when the accident happens and who it affects is mainly stochastic. A fault line emerges into view. When opting for an aggressive speed policy over a more conservative one, a robot designer does not intend to cause any particular accidents but chooses to cause accidents on the systemic level. In the first case, intentionality builds on individualized knowledge. In the latter, knowledge is statistical (Simons 2012). Intentionality analyses are ill-equipped to deal with these situations (Simons 2012; Geistfeld 2017; Simons and Cardi 2017). The existing approaches have evolved to govern particularized processes, focusing on the immediate decisions available to defendants and struggling with disentangling systemic consequences.

The disruptive nature of robot choices could also constrain Hand formula defenses. How the costs for precautions in such defenses, for example, should be framed appears ambiguous. Significant outcome differences likely arise depending on which frame the courts choose. Costs to the manufacturer for specifying, for example, a safer AV speed policy may be negligible, but the overall social costs in terms of lost time may be considerable. Escaping liability could become difficult if the courts choose the manufacturers' cost frame. All three tendencies push liability allocation toward no-fault approaches (Schellekens 2018).

8 Conclusion

In the article, I argued that the regulatory challenge mobility robots pose to regulation arises out of robot emergent behavior that, contrary to many accounts, is a function of three features. The robots are complex technological constitutions, traverse complex open worlds, and machine learning code displays non-linear performance dynamics. I also argued that focusing on strong robot autonomy, capability to learn, and unexplainability of machine learning code misguides the discussion on robot safety regulation.

However, the difficulties in controlling robots should not lead to robots running wild. Robot regulation is a feasible project. Robot technological features will, however, likely force regulators to abandon both rule- and simple performance-based regulatory approaches and develop new simulation-based regulatory instruments. Simulations, then, will allow both developers and regulators to gain sufficient anticipatory capability over robot performance to ensure that they will not cause excessive accidents.

Robots seem bound to usher in a new era of large-scale simulation environments that produce knowledge on safety. Once simulations enter regulatory flows, they transform how regulatory practices produce knowledge. The emergent simulation formations are interesting in their materiality and temporality. They also open new possibilities for future research. They stress the materiality of regulation (Kang 2018) and the care regulators should take when designing complex regulatory assemblages.

Simulation environments may also have important implications for future safety politics and doctrinal developments. Simulation environments will likely allow regulators and industry actors' unprecedented visibility into the future. For regulators, this visibility may open up new policy horizons. For industry actors, the same tendencies could speed up a de facto strict liability for mobility robots as available defenses shrink in their scope. However, significant practical research gaps are visible. Little work has been done to canvass the breadth and depth of regulation needed to create and sustain the environments. Similarly, research on the legal implications of increasing knowledge of process outcomes is warranted.

Acknowledgements The author would like to thank the participants of the Kåseberga seminar and the anonymous reviewers for their valuable comments.

Funding Open Access funding provided by University of Turku (UTU) including Turku University Central Hospital.

Data availability No data is available.

Declarations

Conflict of interest The research was financed in part by the Academy of Finland under Grant Nos. 315007 (AALAW) and 327357 (ETAI-ROS). The author has no direct financial or non-financial interests in robots.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahangar MN, Ahmed QZ, Khan FA, Hafeez M (2021) A survey of autonomous vehicles: enabling communication technologies and challenges. *Sensors* 21(3):706. <https://doi.org/10.3390/s21030706>
- Anwar T (2020) Unfolding the past, proving the present: social media evidence in terrorism finance court cases. *Int Political Sociol* 14:382–398. <https://doi.org/10.1093/ips/olaa006>
- Asada M (2020) Rethinking autonomy of humans and robots. *J AI Consci* 7:141–153. <https://doi.org/10.1142/S2705078520500083>
- Bäck V (2021) Unlocking EU roads: the importance of policy for automated driving. *EU Policy Law Rev* 1(2021):238–258
- Barfield W (2018) Liability for autonomous and artificially intelligent robots Paladyn. *J Behav Robot* 9:193–203. <https://doi.org/10.1515/pjbr-2018-0018>
- Behere S, Törngren M (2016) A functional reference architecture for autonomous driving. *Inf Softw Technol* 73:136–150. <https://doi.org/10.1016/j.infsof.2015.12.008>
- Benjamin G (2022) Drone culture: perspectives on autonomy and anonymity. *AI Soc* 37:635–645. <https://doi.org/10.1007/s00146-020-01042-7>
- Bennett Moses L (2017) Regulating in the Face of Sociotechnical Change. In: Brownsword R, Scotford E, Yeung K (eds) *The Oxford handbook of law, regulation, and technology*. Oxford University Press, Oxford, pp 573–596
- Bennett Moses L, Gollan N (2015) The Illusion of newness: the importance of history in understanding the Law-Technology Interface. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.2697311>
- Black J, Baldwin R (2010) Really responsive risk-based regulation. *Law Policy* 32:181–213. <https://doi.org/10.1111/j.1467-9930.2010.00318.x>
- Calo R (2015) Robotics and the Lessons of Cyberlaw. *Calif Law Rev* 103:513–564
- Carrillo A, de Albornoz V, Molina Millán J, Lara Galera A, Muñoz Medina B (2022) Road speed limit matters—are politicians doing the right thing? *Socio-Econ Plan Sci* 79:101106. <https://doi.org/10.1016/j.seps.2021.101106>
- Carsten OMJ, Tate FN (2005) Intelligent speed adaptation: accident savings and cost–benefit analysis. *Accid Anal Prev* 37:407–416. <https://doi.org/10.1016/j.aap.2004.02.007>
- Chen Y, Chen S, Zhang T et al (2018) autonomous vehicle testing and validation platform: integrated simulation system with hardware in the loop. *IEEE Intell Veh Symp*. <https://doi.org/10.1109/IVS.2018.8500461>
- Cheng G, Zelinsky A (2001) Supervised autonomy: a framework for human-robot systems development. *Auton Robot* 10:251–266. <https://doi.org/10.1023/A:1011231725361>
- Chiu IH-Y (2015) Regulating (from) the inside: the legal framework for internal control in banks and financial institutions. Hart Publishing, Oxford
- Chopra S, White LF (2011) *A legal theory for autonomous artificial agents*. University of Michigan Press, Ann Arbor
- Clarke DM, Hollister I (2010) Introduction to redundancy. *Saf Reliab* 30:4–15. <https://doi.org/10.1080/09617353.2010.11690919>
- Coglianesi C, Mendelson E (2010) Meta-regulation and self-regulation. In: Baldwin R, Cave M, Lodge M (eds) *The Oxford handbook of regulation*. Oxford University Press, Oxford, pp 145–168
- Coglianesi C, Nash J (2017) The law of the test: performance-based regulation and diesel emissions control. *Yale Journal on Regulation* 34:33–90
- Coglianesi C, Nash J, Olmstead T (2003) Performance-based regulation: prospects and limitations in health safety and environmental protection. *SSRN J*. <https://doi.org/10.2139/ssrn.392400>
- Colaner N (2022) Is explainable artificial intelligence intrinsically valuable? *AI Soc* 37:231–238. <https://doi.org/10.1007/s00146-021-01184-2>
- Daniels S, Martensen H, Schoeters A et al (2019) A systematic cost-benefit analysis of 29 road safety measures. *Accid Anal Prev* 133:105292. <https://doi.org/10.1016/j.aap.2019.105292>

- Decker C (2018) Goals-based and rules-based approaches to regulation. BEIS Research Paper Number 8. Department for business, energy and industrial strategy, London
- Demortain D (2011) Scientists and the regulation of risk Standardising control. In: Glos C (ed) Northampton Edward Elgar. UK
- Doyle A, Ericson R (2010) Five Ironies of Insurance. In: Clark GW, Anderson G, Thomann C, von der Schulenburg JM-G (eds) The appeal of insurance. University of Toronto Press, Toronto
- Ebert C, Weyrich M (2019) Validation of Automated and Autonomous Vehicles. *ATZ Electronics Worldwide* 14:26–31. <https://doi.org/10.1007/s38314-019-0090-9>
- Eckert C, Isaksson O (2017) Safety margins and design margins: a differentiation between interconnected concepts. *Procedia CIRP* 60:267–272. <https://doi.org/10.1016/j.procir.2017.03.140>
- Elish MC (2019) Moral crumple zones: cautionary tales in human-robot interaction. *Engag Sci Technol Soc* 5:40–60. <https://doi.org/10.17351/ests2019.260>
- Elvik R, Vadeby A, Hels T, van Schagen I (2019) Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accid Anal Prev* 123:114–122. <https://doi.org/10.1016/j.aap.2018.11.014>
- European Commission (2017) Commission Regulation (EU) 2017/1151 of 1 June 2017 supplementing Regulation (EC) No 715/2007 of the European Parliament and of the Council on type-approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information, amending Directive 2007/46/EC of the European Parliament and of the Council, Commission Regulation (EC) No 692/2008 and Commission Regulation (EU) No 1230/2012 and repealing Commission Regulation (EC) No 692/2008 (Text with EEA relevance)
- European Commission (2021a) Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) COM(2021a) 206 final
- European Economic Communities (1993) Council Directive 93/42/EEC of 14 June 1993 concerning medical devices
- European Economic Communities (1985) 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
- Faul AC (2019) A concise introduction to machine learning. CRC Press, Boca Raton, Florida
- Felzmann H, Fosch-Villaronga E, Lutz C, Tamo-Larriex A (2019) Robots and transparency: the multiple dimensions of transparency in the context of robot technologies. *IEEE Robot Autom Mag* 26:71–78. <https://doi.org/10.1109/MRA.2019.2904644>
- Figueiredo MC, Rossetti RJF, Braga RAM, Reis LP (2009) An approach to simulate autonomous vehicles in urban traffic scenarios. In: 2009 12th International IEEE Conference on Intelligent Transportation Systems, pp 1–6
- Fisa R, Musukuma M, Sampa M et al (2022) Effects of interventions for preventing road traffic crashes: an overview of systematic reviews. *BMC Public Health* 22:513. <https://doi.org/10.1186/s12889-021-12253-y>
- Florio M, Pancotti C (2022) Applied welfare economics: cost-benefit analysis of projects and policies, 2nd edn. Routledge, London
- Foucault M (1977) Discipline and punish. Vintage Books, New York, The birth of the prison
- Foucault M (2008) The birth of biopolitics: lectures at the Collège de France, 1978–1979. Palgrave Macmillan, New York
- Foucault M (2014) Security, territory, population: lectures at the Collège de France, 1977–78. Palgrave Macmillan, Basingstoke
- Foucault M (2002) Power. The essential works of Michel Foucault, 1954–1984. Vol. 3. Penguin Books, London; New York; Victoria, Friedman LM (2016) Impact. Harvard University Press, Cambridge, How Law Affects Behavior
- Geistfeld MA (2017) Conceptualizing the Intentional Torts. *J Tort Law* 10:159–196. <https://doi.org/10.1515/jtl-2017-0024>
- Gless S, Silverman E, Weigend T (2016) If robots cause harm, who is to blame? self-driving cars and criminal liability. *New Crim Law Rev* 19:412–436. <https://doi.org/10.1525/nclr.2016.19.3.412>
- Gold CM (2016) Spatial context: an introduction to fundamental computer algorithms for spatial analysis: an introduction to fundamental computer algorithms for spatial analysis. CRC Press, London
- Guizilini V, Vasiljevic I, Ambrus R et al (2021) Full surround monodepth from multiple cameras. *IEEE Robot Autom Lett*. <https://doi.org/10.48550/arXiv.2104.00152>
- Gutman M, Rathgeber B, Syed T (2012) Action and autonomy: a hidden dilemma in artificial autonomous systems. In: Decker M, Gutman M (eds) Robo-and Informationethics. Lit Verlag, Wien, Some Fundamentals, pp 231–256
- Hansson L (2020) Regulatory governance in emerging technologies: the case of autonomous vehicles in Sweden and Norway. *Res Transp Econ* 83:100967. <https://doi.org/10.1016/j.retrec.2020.100967>
- Haselager WFG (2005) Robotics, philosophy and the problems of autonomy. *Pragmat Cogn* 13:515–532. <https://doi.org/10.1075/pc.13.3.07has>
- Hayles NK (2017) Unthought: the power of the cognitive nonconscious. University of Chicago Press, Chicago, IL
- Head BW (2008) Three lenses of evidence-based policy. *Aust J Public Adm* 67:1–11. <https://doi.org/10.1111/j.1467-8500.2007.00564.x>
- Hellström T (2013) On the moral responsibility of military robots. *Ethics Inf Technol* 15:99–107. <https://doi.org/10.1007/s10676-012-9301-2>
- Hemmo M (2006) Vahingonkorvausoikeus. Talentum, Helsinki
- Hildebrandt M (2020) Code-driven law: Freezing the future and scaling the past. In: Deakin S, Markou C (eds) Is law computable?: Critical perspectives on law and artificial intelligence. Hart Publishing, Oxford, pp 67–83
- Hollnagel E (2014) Is safety a subject for science? *Saf Sci* 67:21–24. <https://doi.org/10.1016/j.ssci.2013.07.025>
- Hunt A (1994) Foucault and law: towards a sociology of law as governance. Pluto Press, London
- Hunt A (2012) Encounters with juridical assemblages: Reflections on Foucault, law and the juridical. In: Golder B (ed) Re-reading Foucault: On Law, Power and Rights, Routledge, London, pp 64–84
- Irwin A, Rothstein H, Yearley S, McCarthy E (1997) Regulatory science—towards a sociological framework. *Futures* 29:17–31. [https://doi.org/10.1016/S0016-3287\(96\)00063-8](https://doi.org/10.1016/S0016-3287(96)00063-8)
- Ivanov AM, Shadrin SS, Makarova DA (2022) The Analysis of International Standards in the field of safety regulation of highly automated and autonomous vehicles. *Syst Signals Gener Process Field Board Commun* 2022:1–6. <https://doi.org/10.1109/IEEECONF53456.2022.9744341>
- Jasanoff S (1995a) Science at the bar: law, science, and technology in America. Harvard University Press, Cambridge, Mass
- Jasanoff S (1995b) Procedural choices in regulatory science. *Technol Soc* 17:279–293. [https://doi.org/10.1016/0160-791X\(95\)00011-F](https://doi.org/10.1016/0160-791X(95)00011-F)
- Jasanoff S (2012) The practices of objectivity in regulatory science. In: Camic C, Gross N, Lamont M (eds) Social knowledge in the making. University of Chicago Press, pp 307–338
- Jasanoff S (2018) Science, common sense & judicial power in US courts. *Daedalus* 147:15–27. https://doi.org/10.1162/daed_a_00517

- Jung JC, Sharon E (2019) The Volkswagen emissions scandal and its aftermath. *Glob Bus Organ Excell* 38:6–15. <https://doi.org/10.1002/joe.21930>
- Kalra N, Paddock SM (2016) Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability? RR-1478-RC. RAND Corporation, USA
- Kang HY (2018) Law's materiality. In: Philippopoulos-Mihalopoulos A (ed) *Routledge handbook of law and theory*. Routledge, London, pp 453–474
- Karnow CEA (2016) The application of traditional tort theory to embodied machine intelligence. In: Froomkin M, Kerr I (eds) Ryan Calo, A. *Robot Law*. Edward Elgar, Cheltenham, pp 51–77
- Koopman P (2018) Practical experience report: automotive safety practices vs. accepted principles. In: Gallina B, Skavhaug A, Bitsch F (eds) *Computer safety, reliability, and security*. Springer International Publishing, Cham, pp 3–11
- Koopman P, Wagner M (2018) Toward a framework for highly automated vehicle safety validation. *SAE Int US*. <https://doi.org/10.4271/2018-01-1071>
- Koopman P, Ferrell U, Fratrick F, Wagner M (2019) A safety standard approach for fully autonomous vehicles. In: Romanovsky A, Troubitsyna E, Gashi I et al (eds) *Computer safety, reliability, and security*. Springer International Publishing, Cham, pp 326–332
- Kouroutakis AE (2020) Autonomous vehicles: regulatory challenges and the response from Germany and U.K. *Mitchell Hamline Law Review* 46:1104–1121
- Kröger W, Ayoub A (2022) Towards «type approval» of automated vehicles: Means of safety validation and simulation-based methods, in particular. In: 14. Workshop Fahrerassistenz und automatisiertes Fahren. pp 141–150
- Latour B (1994) *Science in action: How to follow scientists and engineers through society*. Harvard University Press, Cambridge (Mass.)
- Latour B (2009) *The making of law: an ethnography of the Conseil d'Etat*. Polity, Oxford
- Lee EA (2008) Cyber Physical Systems: Design Challenges. In: 2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC). pp 363–369
- Lee D, Hess DJ, Heldeweg MA (2022) Safety and privacy regulations for unmanned aerial vehicles: a multiple comparative analysis. *Technol Soc* 71:102079. <https://doi.org/10.1016/j.techsoc.2022.102079>
- Leenes R, Palmerini E, Koops B-J et al (2017) Regulatory challenges of robotics: some guidelines for addressing legal and ethical issues. *Law Innov Technol* 9:1–44. <https://doi.org/10.1080/17579961.2017.1304921>
- Lemke T (2019) *A critique of political reason: Foucault's analysis of modern governmentality*. Verso, London
- Leta Jones M, Millar J (2017) Hacking metaphors in the anticipatory governance of emerging technology: the case of regulating robots. In: Brownsword R, Scotford E, Yeung K (eds) *The Oxford Handbook of Law, regulation and technology*. Oxford University Press, Oxford, pp 597–619
- Lipton ZC (2018) The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16:31–57. <https://doi.org/10.1145/3236386.3241340>
- Liu S, Tang J, Zhang Z, Gaudiot J-L (2017) Computer architectures for autonomous driving. *Computer* 50:18–25. <https://doi.org/10.1109/MC.2017.3001256>
- Matthias A (2004) The responsibility gap: ascribing responsibility for the actions of learning automata. *Ethics Inf Technol* 6:175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- May PJ (2003) Performance-based regulation and regulatory regimes: the saga of leaky buildings. *Law & Policy* 25:381–401. <https://doi.org/10.1111/j.0265-8240.2003.00155.x>
- McDermid J, Koopman P, Hierons R, et al (2019) *Certification of Highly Automated Vehicles for Use on UK Roads: creating an Industry-Wide Framework for Safety*. Five AI Ltd
- Michael JB, Drusinsky D, Wijesekera D (2021) Formal Verification of Cyberphysical Systems. *Computer* 54:15–24. <https://doi.org/10.1109/MC.2021.3055883>
- Mittelstadt BD, Allo P, Taddeo M et al (2016) The ethics of algorithms: mapping the debate. *Big Data Soc* 3:2053951716679679. <https://doi.org/10.1177/2053951716679679>
- Morin E (2008) *On complexity*. Hampton Press, Cresskill, NJ
- Nakamura H, Kajikawa Y (2018) Regulation and innovation: how should small unmanned aerial vehicles be regulated? *Technol Forecast Soc Chang* 128:262–274. <https://doi.org/10.1016/j.techfore.2017.06.015>
- National Transportation Safety Board L (2019) collision between vehicle controlled by developmental automated driving system and pedestrian. Tempe, Arizona. March 18, 2018. Accident Report. NTSB/HAR-19/03. PB2019–101402. Washington, DC
- Nyce CM (2023) Why the Tesla Recall Matters. In: *The Atlantic*. <https://www.theatlantic.com/ideas/archive/2023/02/tesla-recall-elon-musk-missy-cummings/673124/>. Accessed 26 Feb 2023
- Parker C (2009) Meta-regulation: legal accountability for corporate social responsibility. In: Kinley D (ed) *Human rights and corporations*. Routledge, London, pp 335–368
- Pawson R (2006) *Evidence-based policy: a realist perspective*. SAGE, London
- Petit N, Cooman JD (2021) Models of law and regulation for AI. In: Elliott A (ed) *The routledge social science handbook of AI*. Routledge, Abingdon, pp 199–221
- Polinsky AM, Shavell S (2010) The uneasy case for product liability. *Harv Law Rev* 123:1437–1492
- Porathe T (2019) Safety of Autonomous Shipping: COLREGS and Interaction between Manned and Unmanned Ships. In: *Proceedings of the 29th European Safety and Reliability Conference (ESREL)*. Research Publishing Services, pp 4146–4153
- Ramey J (2021) Tesla's Full Self Driving System Mistakes the Moon(!) for Yellow Traffic Light. In: *Autoweek*. <https://www.autoweek.com/news/green-cars/a37114603/tesla-fsd-mistakes-moon-for-traffic-light/>. Accessed 27 Feb 2023
- Richards NM, Smart WD (2016) How should the law think about robots? In: Kerr I (ed) Calo R, Froomkin AMi. *Robot Law*. Edgar Elgar, Abingdon, pp 3–22
- Ringbom H, Viljanen M, Poikonen J, Ilvessalo S (2020) *Charting Regulatory Frameworks for Maritime Autonomous Surface Ship Testing, Pilots, and Commercial Deployments*. Ministry of Transport and Communications. Helsinki. <https://julkaisut.valtiouevosto.fi/handle/10024/162560>
- Røsag E (2020) Diabolus ex machina: When an autonomous ship does the unexpected. In: Ringbom H (ed) *Autonomous Ships and the Law*. Routledge, Abingdon
- Rose NS (1999) *Powers of freedom: reframing political thought*. University Press, Cambridge
- Rose N, O'Malley P, Valverde M (2006) Governmentality. *Annu Rev Law Soc Sci* 2:83–104. <https://doi.org/10.1146/annurev.lawso.csci.2.081805.105900>
- Schiermeier Q (2015) The science behind the Volkswagen emissions scandal. In: *Nature*. <http://www.nature.com/articles/nature.2015.18426>. Accessed 18 Aug 2022
- Schellekens M (2018) No-fault compensation schemes for self-driving vehicles. *Law Innov Technol* 10:314–333. <https://doi.org/10.1080/17579961.2018.1527477>

- Schlager B, Muckenhuber S, Schmidt S et al (2020) State-of-the-art sensor models for virtual testing of advanced driver assistance systems/autonomous driving functions. *SAE Int J Connect Autom Veh* 3:233–261. <https://doi.org/10.4271/12-03-03-0018>
- Shalev-Shwartz S, Ben-David S (2014) *Understanding machine learning*. Cambridge University Press, New York
- Simons KW (2012) Statistical Knowledge Deconstructed BU. *L Rev* 92:1–88
- Simons KW, Cardi WJ (2017) Restating the intentional torts to persons: seeing the forest and the trees. *J Tort Law* 10:343–404. <https://doi.org/10.1515/jtl-2018-0004>
- Sinha KC (2007) *Transportation decision making principles of project evaluation and programming*. John Wiley, Hoboken, N.J.
- Smith ML, Kane SA (1994) The Law of Large Numbers and the Strength of Insurance. In: Gustafsson SG, Harrington SE (eds) *Insurance, Risk Management, and Public Policy*. Springer, New York, pp 1–27
- Sprenger F (2022) Microdecisions and autonomy in self-driving cars: virtual probabilities. *AI & Soc* 37:619–634. <https://doi.org/10.1007/s00146-020-01115-7>
- Törngren M, Grogan PT (2018) How to deal with the complexity of future cyber-physical systems? *Designs* 2:40. <https://doi.org/10.3390/designs2040040>
- Traore BB, Kamsu-Foguem B, Tangara F (2018) Deep convolution neural network for image recognition. *Eco Inform* 48:257–268. <https://doi.org/10.1016/j.ecoinf.2018.10.002>
- Turner B, Job S, Mitra S (2021) *Guide for Road Safety Interventions. Evidence of What Works and What Does Not Work*. World Bank, Washington, DC
- Umbrello S, Yampolskiy RV (2022) Designing AI for explainability and verifiability: a value sensitive design approach to avoid artificial stupidity in autonomous vehicles. *Int J of Soc Robotics* 14:313–322. <https://doi.org/10.1007/s12369-021-00790-w>
- UNECE (2020) UN Regulation No. 157. Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems. ECE/TRANS/WP.29/2020/81
- Vale D, El-Sharif A, Ali M (2022) Explainable artificial intelligence (XAI) post-hoc explainability methods: risks and limitations in non-discrimination law. *AI Ethics*. <https://doi.org/10.1007/s43681-022-00142-y>
- van Kampen J, Lammers M, Steijn W et al (2023) What works in safety. the use and perceived effectiveness of 48 safety interventions. *Saf Sci* 162:106072. <https://doi.org/10.1016/j.ssci.2023.106072>
- Vanderbilt T (2008) *Traffic: why we drive the way we do (and what it says about us)*. Alfred A, Knopf, New York
- Vöckler K, Eckart P (eds) (2022) *Mobility Design: Shaping Future Mobility Volume 1: Practice*. JOVIS Verlag GmbH
- Vredenburg K (2022) The right to explanation. *J Polit Philos* 30:209–229. <https://doi.org/10.1111/jopp.12262>
- Wachter S, Mittelstadt B, Floridi L (2017) Transparent, explainable, and accountable AI for robotics. *Sci Robot* 2:eaan6080. <https://doi.org/10.1126/scirobotics.aan6080>
- Wang J, Zhang L, Huang Y, Zhao J (2020) Safety of autonomous vehicles. *J Adv Transp* 2020:e8867757. <https://doi.org/10.1155/2020/8867757>
- Wendehorst C (2020) Strict liability for ai and other emerging technologies. *J Eur Tort Law* 11:150–180. <https://doi.org/10.1515/jetl-2020-0140>
- Weng Y-H, Barfield W, Pagallo U (2018) Robot Law 1.0: on social system design for artificial intelligence. In: *Research Handbook on the Law of Artificial Intelligence*. Edgar Elgar, Cheltenham, pp 609–623
- Will CM (2007) The alchemy of clinical trials. *BioSocieties* 2:85–99. <https://doi.org/10.1017/S1745855207005078>
- Will C, Moreira T (eds) (2010) *Medical proofs, social experiments: clinical trials in shifting contexts*. Ashgate, Farnham, Surrey, England ; Burlington, Vt
- Williams M, Minjares R (2016) A technical summary of Euro 6/VI vehicle emission standards. In: *International Council on Clean Transportation*. https://theicct.org/sites/default/files/publications/ICCT_Euro6-VI_briefing_jun2016.pdf. Accessed 24 Aug 2022
- World Forum for Harmonization of Vehicle Regulations (2021) *New Assessment/Test Method for Automated Driving (NATM) - Master Document*. UNECE, Geneva
- Wortham RH, Theodorou A (2017) Robot transparency, trust and utility. *Connect Sci* 29:242–248. <https://doi.org/10.1080/09540091.2017.1313816>
- Yang G, Xue Y, Meng L et al (2021b) Survey on autonomous vehicle simulation platforms. *Int Conf Depend Syst Appl (DSA)*. <https://doi.org/10.1109/DSA52907.2021.00100>
- Yates J, Murphy C (2019) *Engineering rules: global standard setting since 1880*. Johns Hopkins University Press, Baltimore

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.