A Path Dependent Approach for Characterizing the Legal Governance of Autonomous Systems

JOSEPH E. BORSON\textsuperscript{1} AND HUAN XU\textsuperscript{2}, (Member, IEEE)

\textsuperscript{1}Washington, DC, USA
\textsuperscript{2}Department of Aerospace Engineering, University of Maryland, College Park, MD 20742, USA

Corresponding author: Huan Xu (mumu@umd.edu)

This work was supported by the University of Maryland (UMD) Libraries’ Open Access Publishing Fund.

\textbf{ABSTRACT} Autonomous systems promise significant improvements in many fields. These systems will be subject to legal governance requirements. The literature has largely focused on “autonomous governance” as a framework that is broadly applicable to autonomous devices regardless of the type of system (e.g., aviation or motor vehicles) at issue. While there are regulatory principles applicable to autonomous systems generally, an “autonomy-focused” approach is an inadequate lens to consider the governance of these systems. Rather, because autonomous systems are improvements of currently regulated complex systems, the regulation of autonomous elements will occur within those systems’ preexisting regulatory framework. Accordingly, the nature of future autonomous regulation will likely depend on the preexisting features of that substantive system, rather than on an optimal approach divorced from that history, an attribute known in the social science literature as path dependency. In order to characterize diverse regulated systems with an eye toward assessing future autonomous developments, we develop a framework of regulatory approaches to identify specific features of the preexisting regulatory scheme for a given system. We then analyze that approach by examining three different regulatory regimes (aviation, motor vehicles, and medical devices), across two different continents, and consider how the same type of requirement, e.g., fail-safe systems, can lead to different types of regulations depending on the differing baseline framework.

\textbf{INDEX TERMS} Autonomous systems, legal governance, regulation.

\textbf{I. INTRODUCTION} By replacing or supplementing human judgment and intellect, autonomous systems promise improvements in areas ranging from vehicular transport to medical diagnoses to consumer entertainment devices. But because these systems may present risks to their users or to third-parties, they will be subject to some form of governmental regulation, to ensure their fair, safe, and efficacious operation.

This is not a new observation. Rather, scholars and policymakers have been considering how to regulate autonomous systems for many years. They have developed certain broadly applicable principles, such as requiring transparency or disclosure regarding autonomous algorithms [1]. In so doing, these researchers have been bound by an implicit premise: that “autonomy” or “artificially intelligent” devices constitute a discrete entity that can be studied, and therefore, regulated. While we agree that autonomous elements of complex systems require careful study and consideration as a defined category—particularly when these are the new additions to an existing system—in this paper, we argue that a focus solely on the autonomous elements of a regulated system is incomplete.

Rather, such an approach ignores the fact that systems incorporating autonomy are already subject to extensive and distinct regulatory approaches. Autonomous systems will exist within these existing governance frameworks, and thus the regulation of autonomous systems will be largely dictated by those approaches. Self-driving cars, therefore, will be regulated like cars, and artificial intelligence medical diagnostic systems will be regulated like medical devices; and these approaches will likely predominate over any sort of common “autonomy” governance. While there are common autonomous regulation principles that can extend beyond a specific regulatory application, those principles must be translated and incorporated into whatever specific
governance scheme applies to the system—and these principles will manifest differently depending on the complex system at play.

By focusing on general, often abstract regulatory principles, the existing literature lacks a clear method to map the requirements of an autonomous system or component to actual regulations applicable to actual products. To remedy this limitation, we develop an approach to identify the unique aspects of the governance structure that are applicable to a specific complex system. This structure will serve as the baseline for future autonomous applications within that governance system. The premise of this model is based on the social science concept of path dependence, i.e., whereby actors and institutions and constrained by previous choices into a specific modality in a way that makes altering such paths moving forward costly and difficult, even if alteration would create an objectively “better” outcome [2], [3], [4]. An example of path dependency is when a system becomes dominant, either by random chance or an initial condition, and then stays dominant simply because it is dominant, due to network effects or the difficulty of change. Operating systems or keyboard designs, like QWERTY, are classic path dependent systems [5]. Our framework is thus most relevant for practitioners and designers who must make predictions about the type of regulators and regulatory constraints that will affect their products.

This paper proceeds in five parts. First, we review the existing literature, which largely, though not exclusively, focuses on autonomy governance as a distinct concept regardless of the specific type of system at issue. Second, we develop a “regulatory approach” that identifies the various structural approaches to regulation that specific governance regimes employ. This novel regulatory approach can help engineers and policymakers identify the salient features of an existing regulatory scheme (for example, aviation regulation in the United States), which will be the starting point for how autonomous elements within that scheme will be regulated (for instance, autonomous drones). Third, we identify multiple real-world examples of how the nascent governance of autonomous systems closely resembles the governance of non-autonomous systems within that broader regulatory area. Fourth, we consider how future autonomy standards might be written in different subject areas, using fail-safe requirements as a case-study. This tangibly illustrates how pre-existing differences in regulatory structure can lead to materially different types of substantive regulations. Finally, we discuss the implications of a “regulatory application framework first” approach for future research and policy-making.

II. RELATED WORK
We define autonomous systems as entities that make adaptive decisions in response to input, independent of human interaction [6], [7]. Autonomous systems are often conflated with artificial intelligence, although we use “artificial intelligence” to refer primarily to machine learning techniques (while recognizing that “artificial intelligence” includes a broader set of concepts). These techniques “teach” autonomous systems to make predictive judgments using large data sets, from which the systems “learn” [8].

Previous work in the field of autonomous system governance and regulation has made a number of important findings.

First, the literature has identified common problems posed by autonomous systems that require a regulatory solution [9], [10], [11]. Recent surveys by Zuiderwijk et al. [12] and Taiebigh [13] have identified a number of these issues. One set of problems involve opacity concerns when policymakers do not fully understand how algorithms make autonomous decisions (the “black box” challenge). To mediate these challenges, researchers have emphasized the value of explainability, transparency, and auditing mechanisms that can attempt to make clear the basis for how autonomous technologies make their decisions [1], [14], [15], [16], [17], [18], [19]. A second set of problems involves the underlying usefulness of the data that is used by the autonomous system, including data acquisition challenges, data skew, data misuse, privacy concerns inherent in using or storing such data [20], [21]. A third issue is how artificial intelligence can perpetuate existing discriminatory structures, and how its potential for bias must be explicitly addressed [22], [23], [24]. A fourth problem involves questions of democratic accountability when autonomous systems are making decisions, particularly when those decisions involve public rights or adjudication of public benefits [25], [26]. A fifth problem involves the role of humans in such autonomous systems—in some contexts, like military systems, human oversight may be legally or ethically mandated, while in other contexts, like transportation, human involvement may be less required, particularly if human error is a major cause of operational accidents [27], [28], [29].

Second, the literature has identified regulatory solutions that can be applied to the problems identified above. Identifying the proper solution, however, can be problematic because autonomous technologies develop faster than the law can easily adapt [30]. Many of these solutions focus on transparency, i.e., requiring manufacturers to disclose information about their system, so that third-parties can help the product designer and regulators to identify vulnerabilities, or that those third-parties can certify the products as “safe” via audits or independent validation [17], [18], [31], [32], [33], [34]. Transparency is not a panacea, however, and researchers have noted that disclosure can have its own set of problems (such as disclosing trade secrets or potentially threatening the individual privacy of those whose data makes up the data-sets used in machine learning). Moreover, transparency alone does not necessarily verify the effectiveness of software in responding to unknown criteria [35], [36].

To address verification and validation concerns, researchers have suggested specific regulatory requirements, such as requiring that certain fault-tolerance techniques be incorporated into software designs (e.g., requiring redundancies, adjudications in the event of discrepancies, or specifying failure modes) [37], [38], [39]. Some, like Fisher et al., have
suggested conceptual frameworks to isolate out the high-level and low-level control elements in an autonomous system, and to regulate each separately, with special attention paid to more safety-critical elements [40]. Other work, particularly in the medical context, has suggested basic prerequisite steps to facilitate regulation of autonomous systems, such as common data definitions, standards, and security protocols [41], [42]. And still other researchers have focused on generating evidence that can be used by a certification authority to judge the validity of the autonomous system [43], [44], [45].

Third, the literature has discussed the institutions best suited to regulate autonomous systems. These include private companies, non-governmental organizations, international organizations, and governments [13], [46], [47]. Most researchers have concluded that government entities should play the primary role in regulating autonomous systems, particularly given the potential for negative consequences for third-parties and the user community [48], [49], [50], [51]. Researchers have also considered the efficacy of non-governmental regulation, such as codes of conduct or mandatory insurance regimes, and have some concluded that such techniques could be effective under some circumstances. [52], [53]. Other approaches involve a combination of public and private regulatory institutions, including private certifying and auditing entities that judge compliance with public regulations [54].

Assuming that the government is regulating autonomous systems, the literature has examined what type of government institution is best suited to perform such regulation. Researchers have considered the relative merits of legislatures, which are democratically responsive but have limited technical competency; government agencies, where the reverse is true; the common law tort system, which is reactive and fact-specific, and international organizations, which could create a harmonized, global regulatory structure, but would have the limited dexterity that attempting to standardize global rules implies [55], [56]. Moreover, the underlying regulatory institutions may differ geographically, particularly if those regions are motivated by different geopolitical values (such as international competition or human-rights values) [57], [58], [59]. In short, there is no universally applicable governing institution for autonomous systems.

Fourth, the literature has reviewed specific types of autonomous technologies. There is a deep literature on autonomous vehicles, which has looked at adapting existing tort liability principles to such vehicles, as well as prospects for affirmative regulation or legislation specific to such vehicles [60], [61], [62], [63], [64], [65], [66]. Similarly, medical devices that use artificial intelligence (such as diagnostic software) have also been the topic of considerable research, much of which has looked at ways of making the “black box” of these systems more visible, as well as the importance of preserving patient privacy and ensuring adequate safety review of the relevant systems [36], [67], [68], [69], [70].

Previous work has (1) identified common problems and potential solutions for regulating autonomous systems (i.e., transparency, fail-safe modes, etc.), (2) identified institutions capable of performing such regulation, and (3) considered subject-specific applications of autonomy. The literature has two related gaps. First, it does not significantly engage with the fact that governance structures vary, and different types of systems are regulated in very different ways. Second, most of the literature on autonomous governance is largely system agnostic. By failing to recognize that individual types of systems are and will be regulated in different ways, the literature overlooks key differences in how specific autonomous systems will be regulated in practice.

Engineers have assumed that autonomous elements within complex systems are regulated in similar ways—or, said differently, that autonomous planes, trains, and automobiles will be subject to the same types of autonomy-focused regulations, which will be enforced in the same way [71]. That isn’t so. Rather, these elements will be regulated in different ways in different types of systems, based on the pre-existing and path dependent regulatory frameworks that are already in place. This matters, because regulations that differ in structure, author, and origin will impose dramatically different functional legal requirements on practitioners charged with actually creating these systems; even if they are trying to accomplish the same goal, such as increasing data quality.

Our work develops a framework that can help to turn general principles of autonomy governance into concrete regulations. This framework contributes to the broader engineering community in two ways. First, it allows engineers to make reasonable assumptions about the form and content of future autonomous regulations, which will differ significantly based on the type of regulated system. These differences matter: Broadly drafted rules that allow for significant discretion impose different constraints on designers than those that specify specific technical requirements. Our framework allows practitioners to better assess—and thus consider—future legal requirements that will ultimately inform their creations. Second, this framework provides a summary of the legal framework applicable to autonomous systems, which can help the engineering community more intelligently participate in discussions surrounding future requirements.

III. INTRODUCING A REGULATORY APPROACH

“Regulation,” and governance more generally, is the process by which society constrains or authorizes private behavior in a way that promotes the social good [50], [72], [73]. In this section, we develop an approach that identifies key features of regulatory systems. While related work has identified individual components within this framework, this paper’s novel contribution is three-fold. First, it organizes these components into three levels: (i) the primary regulatory, (ii) the means by which the regulatory regulates, and (ii) the mechanisms by which those regulations are enforced. Our proposed framework is depicted in Figure 1. Second, the paper validates this three-level framework by using it to show how different type of governance structures regulate autonomy differently. Third, the paper uses a case-study approach
to illustrate how pre-existing differences in regulatory approach can lead to dramatically different types of regulation, even when applied to the same substantive engineering challenge. Overall, regulators and the regulated public can use this paper to better assess how future autonomy regulations in their specific spheres may emerge.

We pause for one note about methodology. This paper reviews existing literature to develop a novel framework, which it then applies to the emerging field of autonomous system regulation. This work necessarily involves descriptive and predictive judgments, supported by evidence, as opposed to formal scientific proof. This approach accords with general practices in these types of works, including those published in this journal [74], [75], [76], [77]. We recognize that there have been attempts to more formally describe the interactions between different agents and institutions. The most notable example of this is the “electronic institution” model, which precisely defines what agents are allowed to do or not to do under specific “rules of the game” [78], [79]. While this framework is intriguing, we believe that a more qualitative description is appropriate in this circumstance, given the volume of institutional actors and the continued evolution of the relevant legal principles, which remain somewhat indeterminate. Future work could include the application of the electronic institution model when the “rules of the game” become more defined.

A. IDENTIFYING THE PRIMARY REGULATOR(S)

Regulators with jurisdiction over different geographies regulate complex systems, sometimes exclusively, and sometimes in tandem with other entities [80], [81]. These include national bodies, such as the U.S. Federal Aviation Administration (FAA) and U.S. Food and Drug Administration (FDA), sub-national bodies, such as state governments (in the U.S. context), or international or supra-national entities, such as the European Union (EU). As a general rule of thumb, the smaller the unit of government, the more responsive it can be to local needs, though often at the expense of creating a patchwork of different rules and standards across different jurisdictions. As the relevant regulating polity grows larger, there are fewer opportunities for conflicting regulatory requirements, even as the substantive rules may become harder to change given their broader scope [82], [83], [84].

B. ARTICULATING THE PRIMARY REGULATORY MECHANISMS

Regulatory standards are effectuated in different ways.

1) NON-BINDING STANDARDS

The most basic set of regulatory standards are those that are persuasive or influential, but are not legally binding. Sometimes termed “soft law,” these include private voluntary standards, such as ISO or ASTM standards that are issued by non-governmental expert committees, as well as non-binding guidance and advisory documents issued by government agencies. These standards are usually detailed and technical [85], [86], [87], [88], [89]. Moreover, they are often more flexible than other types of regulatory standards (since they involve fewer procedural hurdles to promulgate) and can be more directly influenced by technical experts. Because they do not directly create legal mandates, however, they may be less effective at directly governing behavior.

2) INSURANCE STANDARDS

Regulatory requirements can be imposed or enforced by insurance companies. Insurance is a social mechanism to manage and distribute risks, and parties can either voluntarily or be legally required to procure such a product [90]. Insurers, by determining what types of actions they will reimburse, can impose practical requirements on the insured population. For example, insurers may require sprinkler systems as a condition of fire insurance, or insurers may reimburse only those types of medical procedures that meet certain affirmative standards [91], [92].

3) TORT STANDARDS

Regulatory requirements can be imposed by tort law, which governs wrongs committed by one party against another [93]. Of these, the most relevant is negligence, where a party has an obligation to act reasonably toward those she owes a duty of care [94]. In practice, these standards regulate private behavior according to a standard of reasonableness. However, tort standards do so through a case-by-case, backward-looking process (with manufacturers not knowing if their designs comport with a reasonableness standard unless and until there is a lawsuit) [95], [96], [97].

4) SPECIFIC STANDARDS

Regulators may impose affirmative, specific requirements on a regulated entity. There are three basic types [98]. The first is prescriptive standards, sometimes also termed technology-based standards or design standards. These standards specify the particular means by which a regulated entity must achieve a specific result. For example, automobile safety standards may require a defroster button to have a particular graphical icon, or that an aircraft must use a specific type of de-icing system. These regulations are relatively easy to define, execute, and enforce [99]. They are limited, however,
by the fact that because they define the requirements in terms of a particularly technological state, they cannot easily adapt to developments in the underlying technological landscape. The second type of regulation is performance-based standards, which define the outcome that a system must achieve, but do not require a particular means of achieving that outcome [100]. For example, an environmental regulation can specify a level of permissible effluent discharge, but allow regulated parties to use any means they judge appropriate to comply with that requirement. These types of regulations are more adaptive to changes in technology, but can be relatively difficult to validate or assess. The third type of regulation, disclosure-based standards, do not impose substantive requirements, but instead require the regulated entity to release information, either to the public or to the regulator, in hopes that such transparency will allow third-parties to make better decisions about whether and how to interact with the regulated entity [101], [102].

C. SPECIFYING THE METHOD OF ENFORCEMENT
Regulations must be enforceable in some fashion. There are four basic techniques. First is the “name-and-shame” method, which does not impose legal consequences for non-compliance, but publicizes that non-compliance to encourage the regulated entity to change its behavior [103], [104]. Second is self-certification, where the regulated party— but not the regulator—will certify that the product meets the relevant standards [105]. Third is affirmative governmental certification, where the regulator itself will certify that the product complies with the relevant criteria [106]. Finally, certain types of regulations, particularly tort standards, are enforced by private parties who are injured in some way by the allegedly improper behavior, and who bring a lawsuit to seek recompense.

D. INTERRELATED REGULATORY ELEMENTS
These regulatory elements are interdependent, and various elements will either work in harmony or take priority relative to each other [107]. For example, non-binding standards can establish the relevant standard of care used to determine whether a manufacturer was negligent, and can be incorporated by reference into specific, government-issued binding standards [108], [109]. Insurance can cover the costs of fines and penalties imposed by a regulator for non-compliance with government regulations, or for a court judgment in the event of negligence liability [110]. Tort standards serve as a form of “background rule” that imposes legal obligations unless other, more specific forms of government regulation state to the contrary [111], [112]. And, by the same token, specific government rules can overrule tort provisions, while, under certain circumstances, rules from one geographic jurisdiction can overrule those of another [113].

The above framework is summarized in Figure 1.

IV. REGULATORY APPLICATIONS
Next, we apply the above regulatory framework to aircraft, motor vehicles, and medical devices. This confirms our assessment that this framework is an accurate way of describing the regulation of autonomous systems. As we illustrate, existing differences in overall regulatory approaches translate into different ways of regulating autonomous elements of these systems. While we focus on regulation in the United States, we also touch upon European regulation, and observe similar path-dependent effects, which boosts our confidence in our approach.

A. AVIATION REGULATION
1) IDENTIFYING THE PRIMARY REGULATOR
In the United States, the Federal Aviation Administration (FAA) is responsible for comprehensive regulation of aircraft, their crew and operators, and their operation.

2) IDENTIFYING THE PRIMARY REGULATORY MECHANISMS
The FAA governs aircraft design by issuing airworthiness standards, which are a mix of design-standards and performance-based standards, and which prescribe specific technical requirements for regulated vehicles [114]. Some standards impose specific, objective requirements, such as defining the size of cable diameter used in the aileron system. Others are more general, such as imposing validation processes for engine control systems, including requiring that the “applicant must design, implement, and verify all associated software to minimize the existence of errors by using a method, approved by the FAA, consistent with the criticality of the performed functions” [115]. These general standards often go in tandem with more specific, non-binding guidance, usually in the form of an FAA-issued Advisory Circular or a privately issued standard, that articulates a specific method of complying with these obligations [116]. The FAA also certifies that aircraft conform with these standards, via type certifications (the overall design of the aircraft), production certificates (the manufacturing process) and airworthiness certificates (that a specific aircraft satisfies the above requirements) [117], [118]. While some aspects of aircraft design and manufacture are subject to tort liability standards under certain circumstances, in general the FAA has expansive and near exclusive regulatory scope over modern aircraft design and manufacture within the United States [119].

Regulation of unpiloted aircraft has generally comported with this framework. The FAA has complete control over the operation of all uncrewed drones in U.S. airspace, as with crewed aircraft, to the exclusion of state or local governments [120]. As of now, there is no provision for true autonomous operations. Instead, the FAA has maintained the requirement that there be a remote pilot in ultimate command of the uncrewed vehicle, who is responsible for its operations and who is able to manually take over the vehicle if necessary. The FAA has also begun developing airworthiness criteria for uncrewed aircraft that are larger than 55 pounds (25 kg) [121]. As proposed, these are explicitly based on the airworthiness criteria for piloted aircraft, with modifications as appropriate for uncrewed operation. Those include the requirement that control stations provide adequate data to allow a remote pilot
to take control, as well as contingency procedures, where the vehicle would be designed to “automatically execute a predetermined action” in the event of loss in communications. Other requirements include the capacity to detect and avoid other aircraft [122], [123]. The FAA’s regulation of uncrewed aircraft thus uses as its baseline the existing regulatory approach for crewed aircraft.

Pilots operating within the United States are also licensed by the FAA. Pilots must demonstrate aeronautical knowledge (i.e., knowledge of airspace rules and procedures, and the characteristics and limitations for the type of vehicle they are going to fly), having received adequate training hours, and having demonstrated proficiency and competency in the make and model of the relevant aircraft, as judged by an authorized instructor [124]. As with crewed aircraft, there is also a certification process and knowledge test requirement in order to certify an uncrewed operator [125].

The FAA is responsible for the conduct of aircraft in the U.S. airspace, which is subject to exclusive federal (as opposed to state) control. The specific requirements for operation in different airspace regions depend on criteria such as altitude and distance from airports and population areas [126], [127]. Drones are also subject to flight and operational restrictions, such as limitations in operation over persons or under certain weather conditions [125].

3) IDENTIFYING REGULATORY ENFORCEMENT MECHANISMS
The FAA directly approves aircraft certifications (sometimes in partnership with manufacturers), licenses pilots, and levies fines and sanctions when necessary to enforce those rules.

4) EUROPEAN REGULATORY COUNTERPOINT
The European aircraft governance framework is similar to the American model in terms of its structure and substantive obligations, however, responsibilities are divided between the EU and its component member states. The European Union, through the European Union Aviation Safety Agency (EASA), is responsible for issuing airworthiness standards and conducting type and airworthiness certification of proposed aircraft designs that operate within the EU [128].

As in the FAA context, EASA issues binding regulations, which are followed up via non-binding technical guidance that specifies appropriate means of complying with the binding regulations. Aircrews are certified by their home member state aviation authorities, subject to standards set by the European Union [129]. Control over European airspace is more complicated, with ultimate authority at the country level, but subject to coordination and technical standards issued by EASA and other pan-European entities [130], [131].

EASA has already issued regulations governing uncrewed aircraft [132]. These regulations divide drone operations into three domains: open, specific, and certified, and these categories are distinguished by the mass of the aircraft and the risk posed by operations (including operations near people or other populated areas). Specific and certified drones require operational authorization from a national regulatory body, open drone operations generally do not. EASA has pronounced rules and requirements for remote pilot operations (to be enforced by national authorities), as well as provisions for airworthiness standards [133], [134].

These rules can be relatively broad. For example, they require that the aircraft be designed so that any catastrophic failure condition is extremely improbable and does not result from a single failure, or that information concerning unsafe system operating conditions be presented to the crew so they can take appropriate actions. While autonomous operations are not specifically provisioned, however, EASA has established a risk-based approach to uncrewed aircraft certification, particularly those in “specific” and “certified” domains, and has specified that “the risk assessment of autonomous operations should ensure, as for any other operations, that the risk is mitigated to an acceptable level” [134]. EASA appears to have a more explicit risk-based approach relative to the FAA, with more relaxed rules for lower risk, non-autonomous operations.

European autonomous aviation will likely continue to follow this framework, where competency is divided between EU and member state institutions, with requirements for aircraft and (remote) aircrew certification specified by the EU and enforced by both. Indeed, EASA has stated that its existing certification standards and regulations will provide the model for future autonomous operations [135].

B. MOTOR VEHICLE REGULATIONS
1) IDENTIFYING THE PRIMARY REGULATORS
In the United States, motor vehicles and their drivers are subject to both state and federal regulation. The federal government is primarily responsible for regulating the safety-related features of the vehicles themselves, while the state government is responsible for regulating the drivers, and also playing an important back-stop role in regulating vehicles via state-enforced tort standards.

2) IDENTIFYING THE PRIMARY REGULATORY MECHANISMS
At the national level, the U.S. Department of Transportation’s (DOT) Federal Motor Vehicle Safety Standards (FMVSS) sets objective standards, capable of measurement by repeatable tests, for specific safety devices and configurations that motor vehicles are required to maintain. These standards, which are usually more design-based than performance-based, range from specifying the control and display requirements for onboard indicators, to rear-visibility camera requirements, to impact protection and crash survivability requirements [136], [137], [138], [139]. State tort law complements and expands upon these requirements, and may impose additional requirements for vehicle design or construction, by imposing liability on manufacturers under certain circumstances [140]. Drivers are regulated primarily by state governments, who license operators and are responsible for setting traffic and safety rules [138]. State tort law also serves as a form of driver regulation, as drivers can be liable
to each other in the event of traffic accidents (with insurance playing an important mediating role by providing financial compensation so long as its requirements are satisfied).

Autonomous vehicle regulation will continue within this mix of state and federal prescriptive and tort standards. To date, DOT has issued voluntary guidance discussing autonomous safety standards (including the need for a robust validation process, crash avoidance, object detection and fallback provisions), but has not yet updated the FMVSS to require any specific standards [141], [142]. Presumably, the FMVSS will be the baseline for federal safety regulations of autonomous vehicles, and researchers have considered how to adapt these standards for driverless vehicles [136], [137]. Indeed, DOT has sought comment about how it should adapt the FMVSS for autonomous vehicles, and whether more flexible standards than those currently implemented are best as technology evolves rapidly [143].

States are also able to set their own laws, and some, like California, have already done so. Those laws emphasize insurance requirements, disclosure to regulators in the event that a driver has to manually re-take control of the vehicle, and the requirement that the manufacturer disclose how the system is intended to respond to poor conditions (e.g., bad weather [144], [145]). Privately issued standards, such as ISO 26262, which apply to certain safety-related automobile electrical systems, may also play a role in institutional governance, although these standards would be enforced only through formal adoption or by government bodies or incorporation into tort-law standards of care [146]. To date, these standards are relatively permissive, rather than prescriptive, focusing on ensuring that there is a driver available to take over in the event of emergencies, as well as a public reporting process in the event that the vehicle is disengaged from autonomous mode. Absent a fundamental change in existing regulatory structures, tort law will play a role in governing the design and manufacture of these vehicles, although there remain open questions about how “operator” liability will work when the operation is autonomous [63], [64], [147]. Indeed, lawsuits have already been filed against car companies with autonomous or semi-autonomous software, challenging the systems’ ability to detect and respond to other vehicles, or to properly account for driver unresponsiveness [148]. U.S. National Transportation Safety Board (NTSB) crash investigations have also identified real-world failures of such systems, which may form the basis for future lawsuits [149].

3) IDENTIFYING REGULATORY ENFORCEMENT MECHANISMS

Federal FMVSS standards, while mandatory, are certified by private auditor organizations. State driver regulations are enforced by the state directly, through its civil regulatory and criminal enforcement mechanisms.

C. EUROPEAN REGULATORY COUNTERPOINT

European vehicles are subject to a similar regulatory framework as in the United States, with authorities divided between different regulators and between specific, prescriptive standards and more general tort standards. As with the United States, European vehicles must comport with specific safety standards, which are usually issued by the United Nations Economic Commission for Europe (UNECE) [150]. Unlike the American context, which relies on self-certification, European vehicles are subject to a “Whole Vehicle Type Certification” process, where the specific model of vehicle is certified by a national authority for compliance with those safety standards [151], [152]. Tort law remains a primary mechanism for ensuring the safety of individual vehicles [153], [154]. Vehicle operation is generally handled by individual countries, subject to their own local laws [153].

Autonomous vehicle regulations have already emerged within this context. Individual countries are beginning to regulate autonomous vehicles within their borders. Germany, for example, has recently passed a draft bill to amend its Road Traffic Act and Compulsory Insurance Act to permit autonomous vehicles in certain spatial areas. These permissions are subject to certain requirements, including mandates that the vehicle be able to independently cope with driving responsibilities and comply with traffic laws, as well as to place itself into a lower-risk fail-safe condition if necessary. The car owner must carry adequate insurance to cover the vehicle and the person who is using its autonomous functions, and the vehicles themselves must maintain information about its operation, which is to be made available to the government under certain circumstances [155]. The UNECE is also developing common, high-level principles toward autonomous vehicles (such as requiring validation, object detection, fail-safe criteria, etc.), although these remain voluntarily principles, as opposed to legally mandated requirements [156].

D. ARTIFICIALLY INTELLIGENT MEDICAL DEVICES

1) IDENTIFYING THE PRIMARY REGULATOR

Medical devices are regulated in the United States by the Food and Drug Administration (FDA). Some of these devices have begun incorporating autonomous elements via machine learning diagnosis algorithms.

2) IDENTIFYING THE PRIMARY REGULATORY MECHANISMS

Medical device regulation in the United States operates pursuant to a risk-based approach, with the simplest devices (such as surgical gloves) being subject to the least regulation, while the most complex and safety-critical devices (such as implantable medical devices) require premarket approval, where the applicant must demonstrate the safety and effectiveness of the device with clinical data [157], [158], [159]. Medical devices that were not in distribution before 1976 are automatically subject to the most stringent requirements, unless the applicant can demonstrate that (1) the device is the “substantial equivalent” to a device that was in distribution before 1976, or (2) pass the “de novo process,” where the product is a new type of device, but
special controls (i.e., the submission of premarket data or postmarket surveillance) can provide “reasonable assurances of safety and effectiveness” [160], [161].

The FDA has begun to approve several machine learning algorithms under the existing de novo process. These include the IDx-DR system, which uses a deep learning algorithm to evaluate ophthalmic images to identify mild diabetic retinopathy, and the QuantX system, which is a computer-assisted diagnostic tool that identifies and characterizes breast abnormalities [162], [163], [164]. In providing de novo approval, the FDA has required that these manufacturers include software verification and validation as well as performance testing sufficient to demonstrate efficacy [165], [166]. This focus on safety and effectiveness is driven by the FDA’s statutory mandates. It is also consistent with the FDA’s guidance on machine learning regulatory principles generally, which are evolutionary offshoots of its existing guidance on regulating medical software, and focus on principles of transparency and validation, particularly around large, evolving machine learning training sets [167].

3) IDENTIFYING THE REGULATORY ENFORCEMENT MECHANISM
The FDA’s medical device regulatory scheme involves pre-approval (or pre-notification, depending on the type of device) by the agency, with penalties for noncompliance.

4) EUROPEAN REGULATORY COUNTERPARTS
By contrast, medical devices in Europe are not regulated by a centralized authority. “Low risk” European devices - the equivalent of surgical gloves in the United States - are self-certified by the manufacturers as meeting the applicable standards [157]. Riskier devices are subject to external certification, whereby national states can designate a “Notified Body,” which is an organization that has been accredited to perform a conformity assessment, i.e., to certify that the device meets the requisite EU standards for safety [157], [168], [169]. To the extent there are European Union regulations or directives that focus on Artificial Intelligence specifically, those requirements would be considered during the conformity assessment stage [170]. Unlike in the United States, there are not public databases of European medical device approval documents, and so it is more difficult to assess how EU regulations are applied, or whether any special conditions have been associated with approvals [157], [170].

V. CASE STUDIES FOR IMPLEMENTING FAIL SAFE REQUIREMENTS
There are multiple ways to design a regulatory system to effect the same general mandates. In this section, we take one example of a regulatory requirement often applied in autonomous systems, contingency planning/fail-safe requirements. We reject the assumption that these requirements will be applied the same in different types of systems [37]. Rather, based on the assumption that pre-existing regulatory approaches will continue to be applied to autonomous elements, we consider how this same principle may translate into different operative requirements in autonomous aircraft and cars in light of those different regulatory contexts. This insight is significant for practitioners because it illustrates that different regulatory frameworks, applicable to different types of systems, will lead to significantly different types of regulations. Those regulations work in different ways, at different levels of generality, are written by different types of people with more less technical expertise, and have greater or lesser degrees of enforcement “teeth.” By understanding how these different regulatory approaches actually equate to different types of concrete legal requirements, practitioners can better assess the specific types of requirements that will be imposed on their systems, as well as the types of people who will write the relevant rules.

A. AUTONOMOUS AIRCRAFT
The U.S. FAA is likely to require performance-based regulation of fail-safe conditions as part of the type certification process. For example, it may require that UAS be designed to automatically execute a predetermined action in the event of loss of communication with the operator, as it has proposed for larger drones [122]. It may also impose methodological requirements to confirm that the systems can actually meet those standards. For example, it may require that fail-safe criteria be “substantiated by tests, analysis, or a combination thereof, that the . . . control system performs the intended functions in a manner that allows for safe operation,” as it does currently for engine certification requirements [115].

In any event, it is likely that the FAA will issue (or recognize) non-binding guidance that articulate specific means of complying with the formally binding performance standards.

B. AUTONOMOUS CARS
1) FEDERAL GOVERNMENT REGULATION
Based on its current approach, the U.S. DOT could impose design-based standards for fail-safe requirements, for example, updating the FMVSS to require multiple sensor types (e.g., LIDAR and a camera). It could also implement performance-based standards, such as requiring that autonomous functions can only operate in specific operational design domains under specified operating conditions (e.g., highways, certain weather conditions). DOT has not yet decided which of these mechanisms it will adopt. These would likely be enforced via self-certification [143].

2) STATE GOVERNMENT REGULATION
State regulation will, as it is now, be divided between operator and vehicle regulation. Autonomous vehicles could be regulated under performance-based standards. For example, California requires manufacturers to describe “how the vehicle is designed to react when it is outside of its operational design domain or encounters the commonly-occurring or restricted conditions.” These include transitioning control to the driver, transitioning to a minimal risk condition, or moving the vehicle away from travel lanes [145]. Other states may follow this approach, which is enforced by mandating such
standards in the permits that provide for legal operation of the vehicle within the state. State tort law may also impose fail-safe requirements. For example, courts may determine that manufacturers are negligent if autonomous software does not detect and respond to driver inattentiveness (an NTSB-identified cause of crashes in the past) [149]. In both cases, federal law could override these requirements.

VI. IMPLICATIONS AND FUTURE WORK

Future efforts to regulate autonomous applications of systems will largely build on the existing regulatory apparatus for those systems, which uses different types of regulatory tools, issued by different types of regulators. The reason is straightforward. Fundamentally new approaches require greater investments in information, resources, and training, because the policymakers must learn how to do something new (and convince others to do so as well), and it is less costly to simply adapt the existing approach, even if that leads to non-optimal structures and institutional inertia [2], [3]. Our conclusion has multiple implications:

A. GOVERNANCE STRUCTURES FOR AUTONOMOUS SYSTEMS CANNOT BE UNDERSTOOD BY CONSIDERING ONLY THEIR AUTONOMOUS ELEMENTS

There are undoubtedly elements of autonomous systems that will require special, autonomy-focused rules, such as requiring fail-safe procedures, data transparency and disclosure regarding machine learning training sets, and validation requirements. But these features by themselves are not complete autonomy governance systems, and researchers should not limit their focus to those elements alone. Rather, these autonomous elements should be mandated or recommended by different types of institutions, should be framed with different degrees of specificity, and should be enforced in different ways. These choices matter: autonomous systems will develop in different ways if they are subject to specific, prospective standards (such as requiring LIDAR as a collision avoidance system) versus outcome-oriented standards (such as requiring redundant sensors in the same system) versus simply disclosing accidents. And they will develop in different ways if they are enforced by different types of regulators, with different enforcement philosophies.

B. DIFFERENCES IN REGULATORY STRUCTURE MATTER TO PRACTITIONERS

Practitioners are ultimately required to comply with government regulations. The form of those regulations matter: some regulations may precisely define mandatory technical requirements that must be included in a system (such as having multiple types of sensors), while others are much more open ended. Some regulations are functionally written by experts, in the form of non-binding standards that are incorporated by reference into government regulations. Others are written by generalist judges or civil juries composed of regular people. Some regulations require the government to pre-approve regulated systems, while others rely on self-certification. These regulatory choices have significant implications for the design, costs, and potential liability of future autonomous systems. And the best way to predict future regulations is to understand the way in which the current regulations are structured.

C. EVEN WHERE THERE ARE ATTEMPTS TO REGULATE AUTONOMY OR ARTIFICIAL INTELLIGENCE HOLISTICALLY, THOSE ATTEMPTS WILL BORROW AND INCORPORATE EXISTING REGULATORY ELEMENTS FROM EXISTING STRUCTURES

There will certainly be attempts to regulate autonomy and artificial intelligence as a specific category. In April 2021, for example, the European Commission proposed a regulation harmonizing certain rules related to artificial intelligence, including mandating certain requirements for “high-risk” systems (i.e., that pose a high risk to health and safety or fundamental rights) [171]. But even these substantive requirements, if adopted, would still be implemented by a network of private standard-setting and certifying entities, as well as nation-states authorities. Moreover, for certain heavily-regulated fields that will be using autonomous systems, such as aviation, these efforts largely graft these requirements onto the existing regulatory scheme, and do not attempt to comprehensively govern the design and operation of autonomous systems. In these circumstances – and in other circumstances where broad autonomy requirements are proposed – they will likely work within the existing regulatory approaches, or as a compliment to those approaches, rather than completely displacing them.

D. PATH DEPENDENT REGULATORY APPROACHES HAVE CHALLENGES AND LIMITATIONS SPECIFIC TO AUTONOMOUS SYSTEMS

By its nature, a path-dependent regulatory approach uses existing legal frameworks and employs personnel who are used to existing processes. This works reasonably well for progressive or incremental adaptations of existing technologies, since those processes can usually be “bent” to adapt to new technologies. For example, U.S. FAA regulators can likely adapt to increasingly autonomous components within crewed aircraft by adopting or modifying the existing approaches currently used for autopilots and flight guidance systems.

A path dependent approach is less adaptable to truly revolutionary technologies, which do not have existing regulatory analogues. Under these circumstances, a path dependent approach has limitations because it is premised on the idea that there are relevant analogous regulatory approaches, and if no such analogous structures exist that premise does not follow. In such circumstances, one of three situations are likely:

First, existing, general regulatory principles will have to adapt to new situations, which will be fleshed out on a case-by-case basis. Tort law is a good example; regardless of the specific technology, there will still be a general, background requirement that the system not be negligently designed. What “negligent” means, however, will be determined
by individual judges in the context of individual cases, so there will not be an overall, guiding approach at the onset. (Another example of such standards would be generalized performance-based standards, such as the overall requirement that medical devices be “safe and effective.”) This requirement would remain in place, but there would be significant ambiguity about how to deploy that requirement in the context of fundamentally new technologies. While these “common law” approaches would likely eventually collapse into a new set of standards, it would take time, and would occur in fits and starts, without the benefit of specific, expert-driven ex ante standards.

Second, the path dependent approach has limitations in the sense that regulators will likely attempt to graft existing regulations onto new systems where they do not align. In the context of autonomous systems, that could include requiring some form of operator involvement (because the existing regulators are premised on the concept that there is always an operator “in-the-loop” and ready to take over), even when new autonomous technologies would benefit from being truly autonomous. One way to respond to such limitations is through “regulatory sandboxes,” which allow for controlled testing of new technologies under a regulator’s supervision, but without the need to be bound by existing rules [172].

Finally, the path dependent regulatory approaches - because they use existing frameworks and personnel - tend to make it harder for new people to enter into a regulatory domain. These approaches may deploy formal (licensing) or informal (culture) barriers to entry. This can be challenging in the context of autonomous systems, particularly when new types of people need to enter - and be trusted - in a new system. For example, autonomous cars require automobile regulators to trust software engineers in new ways; and for those software engineers to understand different cultural mores. (The consequences of failure, for example, may be very different in highway safety versus software design.) These cultural approaches can be bridged, but it requires significant work and effort.

E. THERE IS A DISCONNECT BETWEEN ACADEMIC RESEARCH INTO AUTONOMOUS TECHNOLOGIES AND POLICY-FOCUSED CERTIFICATIONS OR STANDARDS

Academic research into autonomy and artificial intelligence is primarily focused on four topics: (1) verification, which addresses deterministic guarantees (i.e., searching the entire solution set), determining reachable sets, guaranteeing convergence, statistical guarantees, etc; (2) testing, which govern coverage criteria and test case generation; (3) adversarial attack and defense which focuses on how systems can “break” and compensating accordingly; and (4) interpretability is being able to explain the outputs or steps involved in the decisions a neural net or the autonomy software makes [173]. While important, this research is not likely to directly translate into policy certification or regulation, which will occur at a different level. Accordingly, researchers may benefit by thinking about how to “translate” their research into a regulatory context.

This can happen in several ways. Researchers could define key goals, such as verification or transparency, that are necessary to ensure system confidence and are at a sufficiently high level to be adopted into a legally-binding standard. They can also provide input on what types of requirements are so critical for safety that manufacturers should be required to adopt them (i.e., where design standards are necessary, and what those standards should be). Or researchers can think about how to make “enforceable” standards, i.e. those where compliance can be easily assured, even by a non-expert. Still other work could bridge the gap between technical research and regulatory standards, by developing guidance on specific analytical methods that would satisfy more general regulatory standards. Ultimately, by recognizing that the nature of the standards will vary in different fields, scholars can fine-tune their approaches to develop more useful and relevant technical guidance that leads to better policy outcomes.

ACKNOWLEDGMENT

This paper is written in his personal capacity and does not necessarily represent the views of any present or past employer.

REFERENCES

[1] M. C. Buiten, “Towards intelligent regulation of artificial intelligence,” Eur. J. Risk Regulation, vol. 10, no. 1, pp. 41–59, Mar. 2019.

[2] H. Trouvé, Y. Couturier, F. Etheridge, O. Saint-Jean, and D. Somme, “The path dependency theory: Analytical framework to study institutional integration. The case of France,” Int. J. Integr. Care, vol. 10, no. 2, p. e049, Jun. 2010.

[3] E. A. Kirk, A. D. Reeves, and K. L. Blackstock, “Path dependency and the implementation of environmental regulation,” Environ. Planning C, Government Policy, vol. 25, no. 2, pp. 250–268, Apr. 2007.

[4] E. A. Fulton, “Opportunities to improve ecosystem-based fisheries management by recognizing and overcoming path dependency and cognitive bias,” Fish Fisheries, vol. 22, no. 2, pp. 428–448, Mar. 2021.

[5] W. Barnes, M. Gartland, and M. Stack, “Old habits die hard: Path dependency and behavioral lock-in,” J. Econ. Issues, vol. 38, no. 2, pp. 371–377, Jun. 2004.

[6] M. Luckeuck, M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher, “A summary of formal specification and verification of autonomous robotic systems,” in Proc. Int. Conf. Integ. Formal Methods. Cham, Switzerland: Springer, 2019, pp. 538–541.

[7] D. Harel, A. Marron, and J. Sifakis, “Autonomics: In search of a foundation for next-generation autonomous systems,” Proc. Nat. Acad. Sci. USA, vol. 117, no. 30, pp. 17491–17498, Jul. 2020.

[8] T. Punch, P. Szolovits, and R. Atan, “Artificial intelligence, machine learning and health systems,” J. Global Health, vol. 8, no. 2, Dec. 2018, Art. no. 020303.

[9] P. Reinecke, O. Kokshagina, and S. Karanasios, “Framing the regulation of artificial intelligence-based technologies,” in Proc. 29th Eur. Conf. Inf. Syst., Marrakesh, Morocco, 2021, Paper 1556. [Online]. Available: https://aisel.aisnet.org/ecis2021_rip/35?utm_source=aisel.aisnet.org%2Fecis2021_rip%2F35&utm_medium=PDF&utm_campaign=PDFCoverPages

[10] H. Sun, “Legal examination and regulation of artificial intelligence algorithm,” in Proc. Int. Conf. Comput. Inf. Sci. Artif. Intell. (CISAi), Sep. 2021, pp. 588–591.

[11] O. J. Erdélyi and J. Goldsmith, “Regulating artificial intelligence: Proposal for a global solution,” Govern. Inf. Quart., vol. 39, no. 4, Oct. 2022, Art. no. 101748.
[165] J. Hogan, “DeNovo Classification Request for IDx-DR,” Food Drug Admin., Silver Spring, MD, USA, Tech. Rep. 1, 2018.

[166] Food and Drug Administration, “Evaluation of automatic class III designation for QuantX,” Food Drug Admin., Silver Spring, MD, USA, Tech. Rep. 7, 2017. [Online]. Available: https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170022.pdf

[167] Food and Drug Administration, “Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan,” Food Drug Admin., Silver Spring, MD, USA, Tech. Rep. 1, 2021. [Online]. Available: https://www.fda.gov/media/145022/download

[168] E. French-Mowat and J. Burnett, “How are medical devices regulated in the European union?” J. Roy. Soc. Med., vol. 105, no. 1, pp. 22–28, Apr. 2012.

[169] F. Pesapane, C. Volonté, M. Codari, and F. Sardanelli, “Artificial intelligence as a medical device in radiology: Ethical and regulatory issues in Europe and the United States,” Insights into Imag., vol. 9, no. 5, pp. 745–753, Oct. 2018.

[170] K. N. Vokinger and U. Gasser, “Regulating AI in medicine in the United States and Europe,” Nature Mach. Intell., vol. 3, no. 9, pp. 738–739, Sep. 2021.

[171] M. Veale and F. Z. Borgesius, “Demystifying the draft EU artificial intelligence act—Analysing the good, the bad, and the unclear elements of the proposed approach,” Comput. Law Rev. Int., vol. 22, no. 4, pp. 97–112, Aug. 2021.

[172] H. J. Allen, “Regulatory sandboxes,” Geo. Wash. L. Rev., vol. 87, no. 3, p. 579, 2019.

[173] X. Huang, D. Kroening, W. Ruan, J. Sharp, Y. Sun, E. Thamo, M. Wu, and X. Yi, “A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability,” Comput. Sci. Rev., vol. 37, Aug. 2020, Art. no. 100270.

JOSEPH E. BORSON received the A.B. degree (magna cum laude) from Brown University, and the J.D. degree (magna cum laude) from Harvard Law School. While at Harvard, he served as the Vice President and Treasurer for the Harvard Law Review. He was a Law Clerk for both Judge William G. Young of the United States District Court for the District of Massachusetts and Judge Patrick E. Higginbotham of the United States Court of Appeals for the Fifth Circuit. He has been a practicing attorney since 2015, and has appeared in a number of federal district courts across USA, where his practice focuses on regulatory policy and constitutional law. This paper is written in his personal capacity and does not necessarily represent the views of any present or past employer.

HUAN XU (Member, IEEE) received the B.S. degree in mechanical engineering from Harvard University, Cambridge, MA, USA, in 2007, and the M.S. and Ph.D. degrees in mechanical engineering from the California Institute of Technology, Pasadena, CA, USA, in 2008 and 2013, respectively. From 2013 to 2015, she was a Research Assistant Professor of aerospace engineering at the Institute for Systems Research, University of Maryland, College Park. Since 2015, she has been an Assistant Professor with the Aerospace Engineering Department, University of Maryland, where she is currently an Associate Professor of aerospace engineering at the Institute for Systems Research. Her research interests include controls and dynamical systems, safety certification for autonomous systems, and multi-agent control.

***