Lawyers and Engineers Should Speak the Same Robot Language

Bryant Walker Smith

Introduction

Engineering and law have much in common. Both require careful assessment of system boundaries to compare costs with benefits and to identify causal relationships. Both engage similar concepts and similar terms, although some of these are the monoglot equivalent of a false friend. Both are ultimately concerned with the actual use of the products that they create or regulate. And both recognize that the use depends in large part on the human user.

This chapter emphasizes the importance of these four concepts—systems, language, use, and users—to the development and regulation of robots. It argues for thoughtfully coordinating the technical and legal domains without thoughtlessly conflating them. Both developers and regulators must understand their interconnecting roles with respect to an emerging system of robotics that, when properly defined, includes humans as well as machines.

Although the chapter applies broadly to robotics, motor vehicle automation provides the primary example for this system. I write as a legal scholar with a background in engineering, and I recognize that efforts to reconcile the domains might in some ways distort them. To ground the discussion, the chapter frequently references four technical documents:

- **SAE J3016**: *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*, released by SAE International (formerly the Society of Automotive and Aerospace Engineers).\(^1\) I serve on the committee and subcommittee that drafted this report, which defines SAE’s levels of vehicle automation.

- **Preliminary Statement of Policy Concerning Automated Vehicles**, released by the National Highway Traffic Safety Administration (NHTSA).\(^2\) This document defines NHTSA’s levels of vehicle automation, which differ somewhat from SAE’s levels.

---


• ISO 26262: Road vehicles – Functional safety, released by the International Organization for Standardization (ISO). This automotive standard is based on a generic functional safety standard (IEC 61508).

• ISO/IEC 15288: Systems and software engineering – System life cycle processes, released by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) and sponsored by the Institute of Electrical and Electronics Engineers (IEEE).

These documents are only a tiny subset of relevant literature. As the next section discusses, however, every enterprise requires some bounds.

Systems
The concept of the system—a “group, set, or aggregate of things, natural or artificial, forming a connected or complex whole”—is foundational to nearly every discipline. Everything is a system, and every system (with the possible exception of the universe) forms part of another system: Molecules contain atoms; communities contain individuals; languages contain words; processes contain actions.

---

4 ISO 26262-1 at v; ISO 26262-10 at 4.1.
6 System (definition 4), OED.com; see also, e.g., DOMINIQUE LUZEAUX AND JEAN-RENÉ RUault, SYSTEM OF SYSTEMS 150-52 (Wiley 2013); DEREK K. HITCHINS, SYSTEMS ENGINEERING, A 21ST CENTURY METHODOLOGY 11 (Wiley 2007); JAMES RON LEIGH, CONTROL THEORY, A GUIDED TOUR 1-2 (IET 3rd ed. 2012).
7 That is, everything that exists, ever has existed, and ever will exist. But even some meanings of “universe” may not be this comprehensive. See, e.g., Multiverse (definition 2), OED.com.
These system elements typically interact with each other. When a building’s thermostat detects that the indoor air temperature has fallen below a certain threshold, for example, it turns on the heater, which then warms the air. Once the temperature reaches the threshold, the thermostat turns off the heater. This system, so modeled, has a feedback loop. The temperature affects the heating, which in turn affects the temperature—even though, at the same time, external heat sources like the sun may also raise this temperature.

In this way, system elements can also interact with things outside the system. Only open systems have inputs or outputs, but systems with trivial flows may nonetheless be modeled as closed. A sealed water bottle, for example, is an open system with respect to energy (which can penetrate the bottle to heat the contents) but effectively a closed system with respect to matter (which mostly stays inside or out).

Defining a system’s boundaries is a key conceptual challenge. “System boundaries delineate what we consider important from what we deem unimportant. We model the important part.” Excluding some elements could obscure feedback loops: An overactive heater will not meaningfully change the outside air temperature, but it could cause building occupants to open their windows, which could keep the indoor air temperature low enough that the thermostat never turns off the heater. Moreover, if the system is incorrectly treated as closed, important factors may be ignored altogether: In a system encompassing its human actors, subsidies are inputs and externalities are outputs.

System boundaries can be real or imaginary, but physical and temporal immediacy often have unwarranted appeal. In tort law, which struggles to divide the proximate from the distant, the foreseeable from the unforeseeable, and the direct from the indirect, this appeal helps explain why courts are still

---

8 MATTHIAS RUTH & BRUCE HANNON, MODELING DYNAMIC ECONOMIC SYSTEMS 35 (Springer 2012).
9 A subsidy is a benefit that is provided to rather than generated by a given activity (i.e., an input), and an externality is a cost or benefit that is created but not borne by that activity (i.e., an output). A closed system, however, has neither inputs nor outputs.
reluctant to permit recovery for emotional injuries that “occur far removed in time and space from the negligent conduct that triggered them.”

In fourth amendment jurisprudence, the fact of physical intrusion underpinned the Supreme Court’s conclusion that physically attaching a GPS tracker to a vehicle constitutes a search, while some justices would have looked to the duration of the monitoring.

Rising automation and connectivity will demand increasingly thoughtful systems analysis. For at least three reasons, increasingly complex systems could involve or implicate a multitude of others that, at least on first thought, seem unrelated.

First, developments like machine-to-machine networks, over-the-air updates, and three-dimensional printing may lead to products that are neither discrete nor static, complicating traditional distinctions between products and services, among individual products, and among particular versions. Robots may interact not just in the physical realm but also in the cloud, as they draw from and contribute to an assortment of overlapping databases. If so, demarcating these robots on the basis of their physical form may make little sense. A 2013 story about Tesla’s factory, for example, counted 160 total robots. But what does this number signify? Why is it not one—or one thousand?

Second, human enhancement and augmentation will blur the boundary between human and machine. Interaction will become integration, and today’s space- and time-based lines will no longer be easy to draw. Particularly in time-critical situations, decisions may be wholly delegated to or effectively dictated by automated systems that necessarily present only a subset of the total information available to them.

Third, sophisticated tracking and analysis of economic, environmental, and social impacts over entire product lifecycles may lead to revelations, real or presumed, approaching that of Lorenz’s theoretical butterfly effect. Previously unrecognized causal relationships may emerge, and designers and regulators may need to begin from the assumption that everything is connected or connectable.

Developers and regulators of robots may struggle to define these systems, but they should at least do so explicitly. A system, carefully defined, is a powerful concept for describing scope and relation. But when used without this content, the word itself can mislead as much as it can inform. The system that the regulator regulates may have boundaries wholly incongruous with the system that the developer has

---

10 Conrail v. Gottshall, 512 U.S. 532, 545 (1994). Recovery for negligent infliction of emotional distress was first limited to victims who had “contemporaneously sustained a physical impact (no matter how slight) or injury due to the defendant’s conduct, id. at 547, or who were at least “placed in immediate risk of physical harm by that conduct,” id. at 548 (describing the “zone of danger”). Many states have since adopted a broader “relative bystander” test, which nonetheless considers physical, temporal, and relational proximity. Id. at 548.


12 Id. at 964 (Alito, J. concurring).

13 In this context, “the cloud” refers collectively to networks like the Internet that enable infrastructure, information, and applications to be shared.


15 This is not to say that these revelations will necessarily bring greater clarity. See Peter Dizikes, The Meaning of the Butterfly, Boston Globe, June 8, 2008 (“Instead of a vision of science in which any prediction is possible, as long as we have enough information, Lorenz’s work suggested that our ability to analyze and predict the workings of the world is inherently limited.”).
developed. A “surveillance system,” for example, might be a single pair of binoculars or an entire police state.

The language of systems dominates the remainder of this chapter. Language is, more broadly, another crucial meeting point for law and technology—and the focus of the next section.

**Language**

The inconsistent use of several key terms within and across the legal, technical, and popular domains contributes potential and ultimately unnecessary confusion. This section highlights several terms susceptible to this confusion: control, risk, safety, reasonableness, efficiency, and responsibility. The concepts that they represent dominate the remainder of this chapter.

**Control**

Engineered systems, which can be products or services,¹⁶ are generally designed to achieve particular goals.¹⁷ These two qualifications—the existence of an external designer imposing external goals—are essential to conventional control theory, which contemplates “a goal-oriented action” by a subject upon an object.¹⁸ Successful control requires:

(i) A purpose or objective...

(ii) A set of possible actions that offers an element of choice... [and]

(iii) ... some means of choosing the correct action (ii) that will result in the desired behaviour (i) being produced.¹⁹

Given this expansive definition, control theory is useful when the pertinent control system is clearly defined and potentially confusing when it is not. Because they obscure rather than clarify that system, casual references to humans who are “in control,” “in the loop,” “out of control,” or “out of the loop”¹²⁰ and to automated systems that are “under control,” “under human control,” “under computer control,” or “out of control” are particularly unhelpful.

---

¹⁶ ISO/IEC 15288 at 4.31 note 1 (“A system may be considered as a product or as the services it provides.”).
¹⁷ E.g., *id.* at 4.31 (defining “system” as a “combination of interacting elements organized to achieve one or more stated purposes”); see also L'UZEAUX & RUAULT 152 (describing a similar definition as appropriate for artificial but not natural systems).
¹⁸ ZDZISLAW BUBNICKI, MODERN CONTROL THEORY 3 (Springer 2005). The subject is the controller, and the object is the control plant. *Id.*
¹⁹ LEIGH 1-9.
²⁰ Given their early association with aviation automation, the phrases “in the loop” and “out of the loop” likely referred originally to a control loop. See in the loop, OED.com (“1970 Sunday Tel. 22 Mar. 7/7 Fully automatic landing has now been perfected, though it will still be necessary to keep the pilot ‘in the loop.’”); out of the loop, OED.com (“1976 Aviation Week 12 Apr. 63/2 Automation technology can lead to complacency when it takes the controller ‘out of the loop’ by reducing the need for his interaction with a flightcrew and deemphasizing the cooperative aspects of the air traffic system.”); see also “in the loop” and “control loop,” Google Books Ngram View, https://books.google.com/ngrams/graph?content=+in+the+loop%2C%20control+loop%2C+&year_start=1900&year_end=2000&corpus=15&smoothing=1 (showing rise in usage from the 1960s).
Phrases like these are susceptible to numerous technical, legal, and popular meanings. Consider, for example, the “control” of an automated vehicle. An engineer might picture a real-time control loop with sensors and actuators, a lawyer might envision a broad grant of authority from human to machine analogous to a principal-agent relationship, and the public might imagine runaway cars and killer robots. These connotations could produce different answers to the legally significant question of whether a human driver who delegates certain elements of the driving task to her motor vehicle nonetheless “controls” that vehicle.\textsuperscript{21}

Moreover, most automatic control systems can be defined broadly enough that they involve a human and narrowly enough that they do not.\textsuperscript{22} Even a fully automated vehicle might still rely on a human to select its destination and to specify certain parameters of operation.\textsuperscript{23} Indeed, ISO/IEC 15288 expressly notes that “humans can be viewed as both users external to a system and as system elements (i.e., operators) within a system.”\textsuperscript{24} Here too, the system definition could drive the legal conclusion.

Rather than attempt to define control (as well as particular control systems), SAE J3016 deliberately eschews most standalone uses of the term.\textsuperscript{25} Instead, for each level of vehicle automation, it specifies whether the “human driver” or the “automated driving system” performs the “dynamic driving task.”\textsuperscript{26} It also defines each of these terms; “dynamic driving task” refers to “all of the real-time functions required to operate a vehicle in on-road traffic, excluding the selection of destinations and waypoints ... and including without limitation object and event detection, recognition, and classification; object and event response; maneuver planning; steering, turning, lane keeping, and lane changing; acceleration and deceleration; and enhancing conspicuity...”\textsuperscript{27}

In contrast, NHTSA’s Preliminary Statement uses “control” twenty-eight times in the five paragraphs describing its levels of automation.\textsuperscript{28} Combined function automation, for example, “involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. Vehicles at this level of automation can utilize shared authority when the driver cedes active primary control in certain limited driving situations. The driver is still responsible for monitoring the roadway and safe operation and is expected to be available for control at all times and on short notice. The system can relinquish control with no advance warning and the driver must be ready to control the vehicle safely.”\textsuperscript{29}

\textsuperscript{22} See infra note 77.
\textsuperscript{23} Cf. SAE J3016 at 4.4 (excluding “the selection of destinations and waypoints (i.e., navigation or route planning” from the “dynamic driving task”); Smith, supra note 21, at 4.4.4.
\textsuperscript{24} ISO/IEC 15288 at 5.1.2.
\textsuperscript{25} It does refer once to “adaptive cruise control” as an established term of art and, unfortunately, once to the “release [of] steering control.” SAE J3016 at 10.
\textsuperscript{26} Id. at 2.
\textsuperscript{27} Id. at 6 (capitalization and spacing modified).
\textsuperscript{28} Preliminary Statement at 4-5.
\textsuperscript{29} Id. at 4 (emphasis added).
Ultimately, control is more useful as a structure than as a standalone term. Those who would deploy it should first describe the control system they actually intend: the goals, inputs, processes, and outputs to the extent they are determined by a human designer and the authority of the human or computer agents to the extent they are not.

Risk
The risk of a particular harm is the product of the probability of that harm and the severity of that harm; the risk of an act or omission is the sum of the risks of the particular associated harms. In the same way that a single lottery ticket might be extremely unlikely to produce a high payout and merely unlikely to produce a low payout, a single activity might have a low risk of death but a higher risk of nonfatal injury. This aggregate conception of risk is assumed much more than it is actually expressed. ISO 26262, for example, defines risk merely as the “combination of the occurrence of harm ... and the severity of that harm.”

Different domains are concerned about different harms. ISO 26262 confines harm to “physical injury or damage to the health of persons.” Tort law recognizes a broader class of harms, albeit inconsistently, and administrative rulemaking may identify even more harms. Although corporate risk management may also encompass a wide range of harms, it focuses on financial and reputational risk to the company from the occurrence or allegation of injury.

This secondary risk to actors who may have created the primary risk must not be conflated with the primary risk itself. The two do belong to the same system of risk: Secondary risk can increase primary risk by discouraging desirable behavior or decrease it by discouraging undesirable behavior. Nonetheless, a given risk to life or limb cannot be “transferred” to another party or “reduced” by a grant of tort immunity.

Safety, Reasonableness, and Efficiency
Safety can be defined as protection from the risk of physical harm. Without more, however, the claim that a product is safe (or secure or privacy-protecting) is meaningless and, whether rhetorical, empirical, or absolute, tends to be refuted by the very occurrence of injury. Reasonableness typically provides the required qualification: ISO 26262 defines safety as the “absence of unreasonable risk,” and the

---

30 ISO 26262-1 at 1.99.
31 Restatement (Third) of Torts: Liab. for Physical and Emotional Harm § 3 (2010). This corresponds to the PL (or probability of harm multiplied by severity of harm) in Judge Learned Hand’s famous equation, see, e.g., United States v. Carroll Towing, 159 F.2d 169, 173 (2d Cir. 1947).
32 ISO 26262-1 at 1.56; see also ISO 26262-3 at B.2.2 (explaining the abbreviated injury scale (AIS), which considers only physical injuries to humans).
33 These can include, for example, violations of property and dignitary interests as well as bodily harm.
34 See, e.g., supra note 10 (discussing negligent infliction of emotional distress).
35 Curtis W. Copeland, Cost-Benefit and Other Analysis Requirements in the Rulemaking Process, Congressional Research Service (Aug. 30, 2011) at 2-15 (surveying executive and Congressional requirements for cost-benefit and cost-effectiveness analysis); Executive Order 12866, Regulatory Planning and Review, 58 F.R. 51735 (October 4, 1993) (“Costs and benefits shall be understood to include both quantifiable measures (to the fullest extent that these can be usefully estimated) and qualitative measures of costs and benefits that are difficult to quantify, but nevertheless essential to consider.”).
36 ISO 26262-1 at 1.102.
Restatement (Third) of Torts: Products Liability states that “products that are defectively designed or [are] sold without adequate warnings or instructions ... are ... not reasonably safe.”

There is a subtle difference between these two linguistic approaches: Although ISO 26262 attempts to infuse reasonableness into the term “safety,” the Restatement implicitly invites readers to consider what is reasonable with every instance of the phrase “reasonably safe.”

Engineering and law both struggle, however, to define reasonableness. ISO 26262 states that “unreasonable risk” is “risk judged to be unacceptable in a certain context according to valid societal moral concepts” but declines to specify what these concepts are or who should apply them. Legal scholars have forced these questions but disagree on the answers. Product liability, for example, has at least two arguably competing tests for determining the existence of a design defect: one that explicitly uses cost-benefit analysis and another that looks to the safety expectations of consumers.

Meaningful cost-benefit analysis requires an appropriate system. Excluding certain decisions could affect whether a behavior is deemed reasonable. For example, a car crash that seems unavoidable at the given speed might be avoidable at a lower speed. Or a product like lawn darts might seem to have the safest possible design when there may actually be no design that is reasonably safe. Moreover, choosing a particular goal could impact those outside the system. For example, trying to protect the occupants of a vehicle by increasing its mass could endanger pedestrians who are struck by it. If the objectively desirable outcome is not the “efficient” outcome, then the model is wrong: Goals have been improperly defined, externalities have been incorrectly excluded, or elements have been inaccurately valued.

**Responsibility**

Responsibility can be used in a legal, technical, or moral sense. The suggestion in NHTSA’s Preliminary Statement that the human driver is “solely responsible for safe operation” of vehicles in its first two levels of automation, for example, is probably meant to describe the driver’s technical role rather than to conclusively assign legal responsibility. SAE J3016, in contrast, avoids the potential conflation of these domains by speaking of performance rather than responsibility.

Even within the legal domain, responsibility requires careful definition. The concept encompasses both obligations and liabilities: A human driver, for example, is legally required to exercise due care, may be criminally liable if she drives recklessly (in some cases even if she causes no injury), and may be civilly liable if she causes injury (in some cases even if her conduct was reasonable). Similarly, an automaker is

---

38 ISO 26262-1 at 1.129; see also id. at 1.93 (defining “reasonably foreseeable event” as an “event that is technically possible and has a credible or measurable rate of occurrence”).
43 Preliminary Statement at 4.
44 See SAE J3016 at 2.
legally required to certify that the vehicles it markets meet certain safety standards, 45 may be subject to civil fines if it does not comply with the certification requirements, 46 and may be subject to civil liability even if it does. 47

The legal question I am asked most frequently about vehicle automation concerns “who is responsible” in the event of a crash and is often accompanied by a binary choice: either the driver or the manufacturer. Liability, however, is not an either/or proposition: Depending on the facts and the jurisdiction, the vehicle owner, operator, seller, manufacturer as well as upstream providers of parts and services may all be civilly and in some cases even criminally liable. The operator may have failed to properly monitor the vehicle (for which the owner may be also be liable), the manufacturer may have failed to properly guard against this misuse (for which the seller may also be liable), and other firms may have supplied defective components or incorrect data; any or all of these acts may have contributed to the injuries alleged.

This section began with control and ended with responsibility. The growing autonomy of robots will increase the need to articulate a principled relationship between these two concepts both within and across the technical, legal, and moral domains. One key to this relationship is the notion of use, 48 to which this chapter now turns.

**Use**

Engineered systems are used. This actual use (including misuse and abuse) may or may not be an intended use, a legal use, or even a reasonable use. An intended use is one for which the product is marketed (rather than just designed), a legal use is one that is not proscribed by law (even if it gives rise to civil liability), and a reasonable use is one with societal costs that do not exceed societal benefits (consistent with the discussion above). The Venn diagram below illustrates these four elements.

Tensions among the actual, intended, legal, and reasonable uses suggest particular structural failures. A mismatch between legal and reasonable use implies that law as written is either too permissive or too restrictive. A mismatch between reasonable and intended use suggests that the product is, for that use, categorically unsafe. A mismatch between intended and actual use suggests the possibility of a design or warning defect. A mismatch between intended and actual use suggests that users are either uninformed or irrational. A mismatch between actual and legal use suggests that law is either underenforced or obsolete. And a mismatch between legal and intended use suggests a “criminal product.”

Texting-while-driving restrictions illustrate several of these failures: Drivers who text while behind the wheel are behaving illegally and dangerously but not unusually, and some bans on the practice may actually decrease road safety.

An open question is the extent to which product design should attempt to confine actual uses to those that are legal, reasonable, or intended. Speed regulators and ignition interlocks on vehicles, trigger

---


53 See, e.g., Joseph Marutollo, No Second Chances: Leandra’s Law and Mandatory Alcohol Ignition Interlocks for First-Time Drunk Driving Offenders, 30 Pace L. Rev. 1090 (2010); The Use of Alcohol Ignition Interlocks for Reducing Impaired Driving Recidivism, NHTSA,
locks on firearms,\textsuperscript{54} and much of the Internet\textsuperscript{55} already raise this question. The expanding ability of manufacturers to monitor and update their products in the field may increase their obligations to also facilitate reasonable use of those products.\textsuperscript{56} Indeed, the notion of X-by-design—where X is safety, security, privacy, sustainability, or some other value—suggests that products must be designed for their entire lifecycle. While technical and contractual limitations on use might improve product safety, they may also calcify existing law, invite tampering, and preclude some uses that are reasonable and legal.

Traffic speed provides a more extended illustration of the complex interactions among and within these four elements. Imagine that you are driving your car at 70 miles per hour; are you speeding? Answering this question is easy—so easy that it has numerous answers.\textsuperscript{57} Consider a few, divided crudely among legal speed, actual speed, reasonable speed, and design speed:

**Legal Speed**

Speed restrictions have multiple sources within law:\textsuperscript{58}

- **Statutory speeds.** Many state vehicle codes specify maximum speeds for certain road classes, such as Interstate highways, rural two-lane roads, and urban streets.

- **Posted regulatory speed.** A road’s posted speed may be lower than its statutory speed. The state or local maintaining authority typically sets such a limit by reference to the road’s operating speed.

- **Posted advisory speed.** On some road segments, the maintaining authority may advise a speed that is lower than that which is posted. Enforcement of such an advisory speed is at most indirect, through the basic speed law.

- **Basic speed.** The “basic speed law” common to most states requires drivers to maintain a reasonable and prudent speed. If visibility is poor, for example, driving slower than the posted speed might nonetheless violate this requirement.\textsuperscript{59}


\textsuperscript{55} See, e.g., Lydia Pallas Loren, Deterring Abuse of the Copyright Takedown Regime by Taking Misrepresentation Claims Seriously, 46 Wake Forest L. Rev. 745 (2011).


\textsuperscript{57} The most common of which may be: “If you are behind me, yes; if you are in front of me, you are not going fast enough.”


\textsuperscript{59} Smith, supra note 21, at 6.3.3.
Even these limits are not absolute. Exceeding the statutory speed may be expressly authorized for emergency vehicles and, in a few states, for passing another vehicle. More ambiguously, certain emergencies may occasionally justify noncompliance. Furthermore, several states effectively preclude a speeding conviction based on a recorded speed that is near the posted speed.

**Actual Speed**

Historically, traffic engineers have used actual traffic speed as an indirect measure of a road’s appropriate speed. The dominant metric is the 85th percentile speed, which is the speed that 85 percent of unimpeded vehicles are traveling at or below. A posted limit is typically set near this point (provided that it does not exceed the statutory limit), which in one sense implies that exactly 15 percent of drivers are speeding. This approach, however, has at least two significant flaws. First, the collective judgment of drivers may marginalize others who use or are impacted by the road, particularly pedestrians, cyclists, and neighbors. Second, that judgment persistently underestimates certain risks.

Notably, the speed of other vehicles may be as relevant to other drivers as to traffic engineers. Smoother traffic tends to be safer traffic—although this effect is not consistently observed in the context of differential speed limits for trucks and cars.

**Reasonable Speed**

An individual’s optimal travel speed, which would reflect her marginal travel costs related to time, fuel, potential civil and criminal liability, and potential personal injury, likely differs from the societally optimal speed, which would also account for impacts on natural and human environments, health and safety, roadway capacity, and public resources. Congress, for example, indirectly imposed a national maximum

---

64 Eric. T. Donnell et al., Speed Concepts: Informational Guide, FHWA-SA-10-001 (Sept. 2009), available at http://safety.fhwa.dot.gov/speedmgmt/ref_mats/fhwasa10001 (“The 85th percentile speed is used extensively in the field of traffic engineering and safety. Since the majority of drivers are considered reasonable and should be accommodated, some numerical definition for this segment of the driver population is needed. Over time, the 85th percentile driver (or speed) has been used to characterize reasonable and prudent behavior.”).
speed limit in 1974 primarily to reduce energy consumption and retained it in some form for over 20 years in large part because of the perceived (crash) safety benefits.\(^{68}\)

Safety itself can be defined expansively or narrowly. Emissions from motor vehicles directly harm human health: One study estimated that these emissions annually kill roughly 58,000 people in the United States alone.\(^{69}\) Typically, however, motor vehicle safety contemplates “only” crashes, which each year kill more than 30,000 and injure more than 2,000,000 in the United States.\(^{70}\)

Vehicle speed can increase both the probability and severity of a crash. A pedestrian, for example, is highly unlikely to die if struck by a car traveling at 20 mph and highly likely to die if struck by a car traveling at twice that speed; every 1 percent change in speed is associated with an increase of about 4 percent in fatal crashes.\(^{71}\)

Crash probability and severity, however, can also depend on vehicle design. Ralph Nader’s seminal book on the automotive industry of the mid-20th Century sought in part to rebut the broad perception that high-speed vehicle crashes were simply unsurvivable.\(^{72}\) Since then, improvements to active and passive safety systems in vehicles have contributed to a significant decline in crash-related fatalities.\(^{73}\)

Risk tolerance has also decreased over that period. NHTSA estimated the value of a life at $1 million in 1973, $4 million in 2005, and $6 million in 2010 (adjusted for inflation).\(^{74}\) This persistent rise suggests that people were undervalued either by the federal government or by the people themselves—and that this might continue to be the case today.

**Design Speed**

Transportation engineers frequently use a particular “design speed” to determine minimum geometric requirements for a road, including how sharply it can turn and how far along it a driver must be able to see. Only recently have state transportation authorities begun to separate design speed from the concept

---


of safe speed.\textsuperscript{75} Inherent in the design speed, which often exceeds the posted speed (whether regulatory or advisory), are myriad assumptions about environmental conditions and vehicle capabilities: A truck on wet pavement will handle differently than a sports car on dry pavement, for example. Furthermore, design speed has, at least prior to modern traffic calming approaches, functioned as a floor but not a ceiling: A horizontal curve might be widened or banked to accommodate travel at the design speed, but extraneous curves would not be introduced solely to discourage higher speeds. As one state department of transportation noted, this could yield “an infinite Design Speed” on a straight and flat road.\textsuperscript{76} Such a speed would be difficult to exceed.

***

In short, legal speed affects design speed, which affects reasonable speed, which affects actual speed, which affects legal speed, which affects actual speed, which affects reasonable speed. These three concepts are distinct and dynamic, and designing a road that satisfies all of them without distorting any of them may well be impossible. Automated systems face a similar dilemma, in part because they still depend in part on human judgment or performance. The next section develops this point.

**Users**

In the words of the Defense Science Board (DSB), “all autonomous systems are joint human-machine cognitive systems.”\textsuperscript{77} They “are supervised by humans to some degree, and the best capabilities result from the coordination and collaboration of humans and machines.”\textsuperscript{78} In a sense, the DSB implicitly argues that automated systems should be conceived broadly enough that humans who might otherwise be viewed as external designers or users are instead viewed as internal operators or supervisors.\textsuperscript{79} This is a wise approach.

ISO 26262 illustrates the interplay between human and machine in today’s safety-relevant electric and electronic systems. Some of the active safety technologies that fall within its scope, such as antilock brakes and electronic stability control, are designed to compensate for the shortcomings of the human driver. Conversely, the required integrity level of these technologies depends in part on the driver’s ability to detect failure and on her ability to avoid the specific harm caused by that failure.\textsuperscript{80} In this way, both the safety technology and the human driver are part of a larger safety system.

Increasing automation actually amplifies the complexities of this human-machine relationship. SAE J3016’s levels of driving automation, which are summarized in the chart below, reveal some of this complexity. Driving is a task that is shared, consecutively or concurrently, by a “human driver” and a “vehicle automation system.” The four key variables are whether steering, acceleration, and deceleration

\textsuperscript{76} State of New Jersey Department of Transportation Roadway Design Manual, http://www.state.nj.us/transportation/eng/documents/RDM.
\textsuperscript{78} Id.
\textsuperscript{79} See supra notes 22-24.
\textsuperscript{80} See ISO 26262-3 7.4.3.7 note 2; id. at annex B.
are executed by the human driver or by the automation system;\textsuperscript{81} whether the driving and traffic environment is monitored by the human driver or by the automation system; whether a human must be available while the automation system is engaged—which depends on whether that system reverts to a “minimal risk state” in the absence of human input; and whether the automation system is functional in all driving environments or only in certain driving environments.\textsuperscript{82}

\textsuperscript{81} This is the only nonbinary variable: At the “assisted” level of automation, the human and the automation system share these functions.

\textsuperscript{82} This variable imperfectly maps a multidimensional taxonomy onto a single axis.
### Summary of Levels of Driving Automation for On-Road Vehicles

This table summarizes SAE International's levels of driving automation for on-road vehicles. Information Report J3016 provides full definitions for these levels and for the italicized terms used therein. The levels are descriptive rather than normative and technical rather than legal. Elements indicate minimum rather than maximum capabilities for each level. "System" refers to the driver assistance system, combination of driver assistance systems, or automated driving system, as appropriate.

<table>
<thead>
<tr>
<th>Level</th>
<th>Definition</th>
<th>Execution of steering and acceleration/deceleration</th>
<th>Monitoring of driving environment</th>
<th>Fallback performance of dynamic driving task</th>
<th>System capability (driving modes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 No Automation</td>
<td>the full-time performance by the human driver of all aspects of the dynamic driving task, even when enhanced by warning or intervention systems</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Human driver</td>
<td>n/a</td>
</tr>
<tr>
<td>1 Driver Assistance</td>
<td>the driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task</td>
<td>Human driver and system</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>2 Partial Automation</td>
<td>the driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task</td>
<td>System</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>3 Conditional Automation</td>
<td>the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene</td>
<td>System</td>
<td>System</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>4 High Automation</td>
<td>the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>5 Full Automation</td>
<td>the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>All driving modes</td>
</tr>
</tbody>
</table>

To these four variables I would eventually add another: whether the human driver or the automation system exercises ultimate authority over steering, acceleration, and deceleration. This authority might be immediate (enabling instant override), delayed (requiring some transition), or otherwise mediated. It may matter in a variety of circumstances, including when an automation system can respond quicker than a

---

human, when the automation system acts in concert with others, and when deliberate human signals to steer, accelerate, or decelerate are inconsistent with the environmental constraints perceived by the automation system. The committee discussed authority at length but reasonably considered it beyond the scope of its initial taxonomy.

The domains introduced in the previous section help clarify these levels of automation. SAE J3016 attempts a purely descriptive taxonomy; accordingly, it specifies the intended use on the assumption that this use is reasonable, and it does not specify the actual or legal use. Considering these uses raises difficult questions.

At partial automation (level 2), the human operator monitors the driving and traffic environment and steers, accelerates, and decelerates only when necessary. Within this category, the reasonable and the legal human behaviors are largely coterminous; only New York expressly requires the human driver to keep at least one hand on the steering mechanism at all times. The more pressing question is whether this legal and definitionally reasonable behavior is actually attainable: Would a human who is not actively providing input to her vehicle be willing or even able to maintain the level of vigilance that is assumed?

At conditional automation (level 3), the vehicle automation system monitors the driving and traffic environment, and the human operator steers, accelerates, and decelerates only when prompted in advance to do so by the vehicle automation system. Within this category, the reasonable and the legal human behaviors are, arguably, still largely coterminous; while many states require due care and prohibit distracted driving, most of these provisions specify or imply that a driver is only distracted when she is behaving unreasonably.

Again, the more pressing question here is whether the behavior is actually attainable: Would a human who is not even monitoring the roadway be willing to stay in an optimal driving position and able to stay awake? Would she be willing and able to acquire and then to retain the driving skills necessary to safely maneuver the vehicle when actually required? Moreover, when her vehicle encountered unusual road or traffic conditions, would she be willing to defer to the automation system’s response?

These questions implicate the very idea of safety as well as the values that underlie or conflict with it. As the discussion of travel speed indicated, there is no uniform understanding of safety from a technical much less legal perspective. Indeed, safety assessments necessarily involve assumptions about scale, timeline, and causation—classic issues related to system boundaries.

Consider, for example, how “safe” a human-vehicle pair must be. One answer might be that it must perform at least as well as an expert human driver for any conceivable individual maneuver or scenario.

84 An AEIS, for example, might brake or steer to avoid an imminent crash.
85 Vehicle platoons and automated intersections, for example, might require such active coordination.
86 If an AEIS brakes or steers, for example, a human might attempt to accelerate or countersteer.
87 Authority is particularly important to the operation of automated emergency intervention systems (AEISs) because of the critical situations in which they act. Although SAE J3016 pointedly excludes these systems from its scope, its taxonomy may nonetheless be informative. Equivalent AEIS levels, for example, might describe warning (level 0), braking or serving (level 1), momentary braking and swerving (level 2), longer braking and serving (level 3), stopping on the shoulder (level 4), and driving directly to a hospital (level 5).
Such a stringent standard, however, might mean that automation technologies reach the market more slowly and at greater cost, resulting in the loss of lives that they could have saved. The dramatization: Perpetuation of a tragic status quo.

An answer on the other side of the spectrum might be that the human-vehicle pair must, in a broad statistical sense, be safer than today’s cars and drivers. Based on rough 2009 estimates, a vehicle is involved in a crash about once every 160 thousand miles and in a fatal crash about once every 65 million miles. Automated vehicles that could simply beat these figures might, in broad terms, represent a safety improvement. Because crashes involving such vehicles might be different than those exclusively involving more conventional vehicles, however, the result of this logic might be fatalities that would be preventable under today’s assumption of unambiguously human-driven vehicles. The dramatization: Headlines proclaiming “machines kills child,” ruinous litigation, and long-term damage to the credibility of the technologies.

Related to this question of reasonable safety is the role of the human user. Conceivably, for example, an automated vehicle might not require real-time human monitoring to satisfy a particular safety threshold. Nonetheless, what if that human oversight further increased the system’s overall safety? Alternately, what if occasional human intervention were to prevent some crashes while causing others? Deliberately eclipsing human authority—the variable excluded from SAE J3016—may well alienate actual use from legal use even if it more closely aligns actual and reasonable use.

On the assumption that humans may still need to maneuver their vehicles in some situations, particularly emergencies and unusual conditions, degradation of or failure to acquire critical driving skills is a concern. Unclear, however, is how that concern should factor into an assessment of safety. A recent safety alert from the Federal Aviation Administration (FAA) raises this point: While acknowledging that autoflight systems “have improved safety,” the memorandum nonetheless urges airlines to ensure sufficient opportunity for “manual flight operations” to prevent “degradation of the pilot’s ability to quickly recover the aircraft from an undesired state” —a contributing factor to, for example, the 2009 crash of an Air France flight. But while airline passengers might understand the need for practice, they might not appreciate being on the flight used for that practice.

Human factors scholars have long recognized that circumscribing the human role in a human-machine system does not necessarily make that system less susceptible to human failure. This “iron[ɔ] of automation” is also present in tort law: Injuries associated with automated vehicles and other robots will result at least in part from the behavior of the humans who use or encounter these machines. To mix metaphors, the human in the loop may be the weakest link in the chain. For this reason, a developer of

92 Id.
an automated system must understand human behavior at least as well as it understands machine performance.

**Conclusion**

The systems analysis introduced in this chapter reveals the conceptual, linguistic, and practical difficulties that developers and regulators will confront on the path of increasing automation. The safety of these automated systems will be determined both by their design and by their use, and the humans who remain a key part of this design or use are best understood as part of the systems themselves. Sensibly defining these systems in turn requires thoughtful dialogue between the technical and legal domains: Lawyers and engineers can—and should—speak the same robot language.