**Regulating Safety-Critical Autonomous Systems: Past, Present, and Future Perspectives**

M.L. Cummings
David Britton

**Abstract**

From unmanned aircraft to self-driving cars to closed-loop medical devices, autonomous systems offer great benefits but also pose new risks. Regulators must grapple with how to manage these risks, all while attempting to keep pace with technological developments and exhibit appropriate precaution without stifling innovation. Examination of how three agencies, the FAA, FDA, and NHTSA, regulate safety-critical systems, with a specific emphasis on their regulation of existing advanced automated systems, suggests important lessons learned that should be applied to the regulation of autonomous systems. These lessons include 1) Regulation that begins early in a systems development lifecycle may appear to be costly and time consuming, but for nascent technologies, systems engineering-based practices may provide critical checks and balances and ultimately such an approach may be the least expensive, 2) Current regulation of automated safety-critical systems has not adequately addressed automation surprises in the form of mode confusion and software certification, which are likely to become more problematic with autonomous systems, and 3) Expediting regulatory approvals for new technology with embedded software based on weak claims of existing equivalent technology likely significantly underestimates the risk that such technologies pose to the public. Going forward, not only do the federal agencies need to work together to better understand the how regulation needs to fundamentally shift for autonomous systems, they also need to consider companies and independent evaluators as critical stakeholders in defining future regulation.



**Introduction**

Autonomous systems in the world today include self-driving vehicles, which use sensors to estimate nearby obstacles and stored mapping data in order to safely navigate to a desired destination; artificial intelligence-based financial trading systems, which track market conditions and individual stocks and make independent decisions on when to buy or sell (Maney 2017); and even new medical devices which monitor a patient's physiological condition and alter the rate of drug delivery or direct other medical intervention without caregiver input (Schwartz 2017).

Differentiated from *automated* systems that operate by clear repeatable rules based on unambiguous sensed data, *autonomous systems* take in information about the unstructured world around them, process that information to analyze possible outcomes, and use that analysis to generate alternatives and make decisions in the face of uncertainty.

While autonomous systems hold great promise including increased access to education, public health, mobility, and transportation, there are also potential negative consequences. For example, consequences may include privacy invasions by camera vision and related tracking systems, significant opportunities for abuse and manipulation of autonomous systems such as that exhibited in the 2017 US election manipulation of social media algorithms (Woolley and Howard 2017), and threats to personal safety as seen in the recent death of a pedestrian due to

self-driving car sensor blind spots (Griggs and Wakabayashi 2018). As a result, calls for increased government regulation of autonomous systems are growing (Lietzen 2017, Laris 2018).

Technology regulation typically focuses on lowering risks and reducing potential negative consequences associated with an industry, activity, or product. Technology regulation could be seen as limiting the use of a technology, which could result in a decrease in innovation and incentives to invest in newer technologies (Jaffe, Peterson et al. 1995). However, competing research demonstrates that regulation can actually drive innovation and technological progress towards societal goals (Ashford and Hall 2012). Thus, the overarching challenge of regulating emerging technologies is to design regulations that both encourage fulfillment of a technology's potential but also manage associated risks.

There are many risks associated with autonomous systems that regulators will likely not have encountered with previous technologies, or risks will be manifested in new ways. Autonomous systems require new forms of computer-based sensing, information interpretation, and action generation in ways that are not always understood even by their own programmers (Knight 2017). The newness and unpredictability of autonomous systems means that many failure modes will be unforeseen, and therefore untested and unmanaged. Reducing the risk of human error is often cited as a main benefit of autonomous systems (Villasenor 2014), but that is only possible if autonomous systems become more reliable than humans.

Determining whether autonomous systems meet or exceed the reliability of humans is not straightforward due to the complexities of the software that drive these systems as well as what kind of testing is needed to make such assertions. For example, one study has asserted that in order to demonstrate a driverless car is as safe as humans, at least 275 million miles must be driven, which would take possibly up to a decade under current testing protocols (Kalra and Paddock. 2016). Thus, potentially new and different reliability assessment methods are needed if technology innovations are to be realized in more expeditious timeframes. Unfortunately, testing and certification of autonomous systems is still an immature field of inquiry.

Autonomous systems rely on probabilistic reasoning and significant estimation through a mathematical estimate approach called machine learning, aka deep learning. Such pattern recognition algorithms are a data-intensive approach to developing an autonomous system world model, which serves as the core set of assumptions about who, what and where agents in the system are and what their likely next set of behaviors and actions will be (Hutchins, Cummings et al. 2015). To date, there exists no industry consensus on how to test such systems, particularly in safety-critical environments, and such approaches to computer-based reasoning have been criticized as deeply flawed (Marcus 2018).

Given that there are new and emerging risks that must be mitigated with the introduction of autonomous systems in safety-critical environments, it is not clear how regulatory agencies could and should respond. Regulatory agencies typically struggle to keep pace with technological change, often referred to as the pacing problem (Krisher and Billeaud 2018). The inertia created by the procedural requirements of administrative law causes agencies and regulations to lag behind technological innovation, which is especially problematic in the current climate of rapid autonomous technology development. Institutional expertise also lags as, for example, robots and artificial intelligence are introduced into industries whose traditional regulators are

unfamiliar with advanced computing and need to acquire the technical knowledge needed to understand such systems (Calo 2014).

In order to better understand how regulatory agencies of safety-critical systems could and should adapt as autonomous systems become more commonplace, we first discuss how such technologies come into existence from a systems engineering perspective. We then discuss how three different federal regulatory agencies, the Federal Aviation Administration (FAA), the Food and Drug Administration (FDA), and the National Highway Transportation and Safety Administration (NHTSA) approach regulation of new technologies in general, and more specifically their progress with automated and autonomous systems. We conclude with a comparison of the three different approaches to regulation of new technologies and discuss possible paths forward.

**The Systems Engineering V Model**

In order to understand how autonomous systems could or should be regulated, especially those in safety-critical applications, it is first important to understand how such systems come into existence, so that critical regulatory decisions can be mapped to major gateways and milestones of system development. There are many prescriptive frameworks that describe in detail how such complex systems should be engineered, including the traditional waterfall system engineering process (Blanchard and Fabrycky 1998), the spiral model (Boehm 1988), and the agile software model (Crowder and Friess 2013). The systems engineering model that will be used in this discussion is the V-model since it represents a concise visual representation of the major steps in a system's development lifecycle, as well as the iterative nature of the underlying processes (Mitre 2014).

The V-model (Figure 1) is a visual representation of the main steps of systems engineering, sometimes referred to as the life-cycle building blocks (Mitre 2014). On the horizontal axis is time, indicating that the activities in each block occur in a certain order, but also overlap in time. This implies that some iteration between steps may be necessary, and that some elements of a project may proceed to the next block while others stay behind. The vertical distribution indicates the level of system decomposition, such that the lower blocks address components of a system, while higher blocks look at system definition, integration, and implementation.

As depicted in Figure 1, in an ideal setting when a new system is conceived like a self-driving car, a management plan is developed that accounts for initial budgets, project timelines, and necessary personnel, which then leads to the definition of a concept of operations. A concept of operations (CONOPs) is then conducted, which generally describes how operations with a new technology should occur in order to meet some common goal or identified need. CONOPs frequently embody plans for resource allocation and scheduling (both objects and people), particularly for complex operations that involve many different entities. In the example of a self-driving car, a CONOPs would lay out when, where, and under what conditions self-driving cars would operate, including contingency operations and identification of relevant stakeholders like pedestrians, bicyclists, etc.

Once the CONOPs is agreed upon, then system and component requirements can be generated, which are critical in defining the design space. Good requirements are unambiguous, measurable, testable, and traceable statements that form the intermediary step between the

concept of operations and detailed system design. They define the criteria that must be met by designs and implementation and then theoretically form the backbone of testing on the right side of the V-Model.

System requirements typically revolve around anticipated functionalities, i.e., a self-driving car must be able to operate in rain, snow and fog and on city and highway roads, while component requirements typically focus on lower level detail such as a LIDAR (Light Detection and Ranging) that should be able to detect a static or dynamic obstacle within 200m of a moving vehicle.

Once the system and component level requirements are generated, then the design space

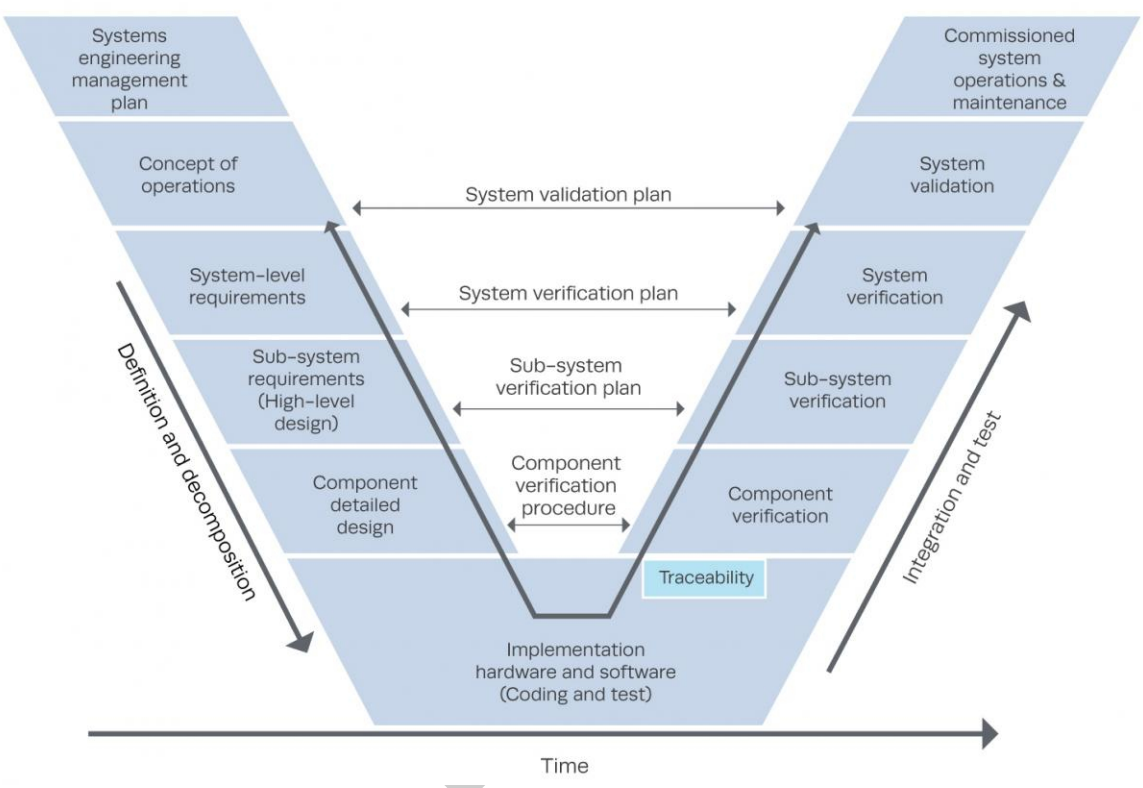


**Figure 1: The Systems Engineering V-Model (Mitre 2014)**

can be characterized, which then results in the development of both hardware and software. The CONOPs, requirements generation, and design phases are iterative and fluid, especially in the early days of a system's development. Often design issues, such as the inability of a sensor to work in all weather conditions like LIDAR, will cause engineers to revisit the CONOPs and requirements to fill in a gap that was only revealed in the design and implementation stages. Engineers, in theory, should be able to trace design decisions back to the relevant requirements, in an effort to assure that the end product matches stakeholder needs. Traceability to requirements is a key systems engineering concept at play in the design and implementation phase.

Once initial designs are completed, verification and validation tests ensue, typically first at the component levels, i.e., testing a LIDAR first in a lab before it is integrated into a moving

vehicle. Eventually tests address full scale system operation in realistic environments, like having a self-driving car drive on public roads with a safety driver behind the wheel. It is not typically feasible for engineers to test all possible scenarios faced by the system, including environmental influences, varieties of user input, and all the ways that system components might fail. Thus, engineer must develop a cost-benefit tradeoff plan to determine how much testing is needed to demonstrate a system is safe-enough. Of course, these test plans are heavily influenced by what, if any, testing regulations exist.

For deterministic systems, those that react in the same way to the same situation every time, comprehensive testing is much easier to achieve, and such testing is well established in current safety critical systems such as aviation. But for autonomous systems that incorporate machine learning and other probabilistic reasoning techniques, proving that these systems will act as needed across a comprehensive set of tests, including corner cases at the limits of a system's capability, is not well established, and is still an open area of basic research (Andrews, Abdelgawad et al. 2016, Cummings in press).

Once a system passes its final validation tests, then in theory the system is ready for large scale operation and deployment. However, testing often reveals latent issues either in design or even perhaps a requirements gap, which then requires revisiting an earlier block as represented by the horizontal lines in Figure 1 that reach back to a previous set of decisions. Thus, the design of such systems is iterative and not linear. Revisiting earlier blocks, especially those that require movement from right to left across the V, represent very costly decisions, both in time and money. Thus, companies should strive to address as many problems as possible before moving up the right leg of the V with final designs and associated testing. Uncovering design flaws at the top right of the V can spell disaster for a system, especially one with significant embedded software, which is hallmark of autonomous systems.

Without principled and extensive testing, often software issues do not reveal themselves until the late stages of testing or even operation, sometimes resulting in fatal outcomes. The airline industry is replete with examples of latent software issues not revealing themselves until years into operations (Favarò, Jackson et al. 2013). Medical communities (Leveson and Turner 1995, Grissinger 2012) and air traffic control systems (Bar-Yam 2003) have also experienced such latent problems due to a lack of comprehensive software testing, a lack of understanding of the critical human-automation interface, and a lack of a principled systems engineering process. Most recently, a latent problem with a self-driving car's computer vision system and its inability to make a decision about the existence of a pedestrian led to the death of this pedestrian in March of 2018 while Uber was conducting late stage validation tests (Griggs and Wakabayashi 2018).

It is important to note that the V-model systems engineering process, or really *any* system engineering process, is an ideal, best-case process. Technical risks are theoretically identified early in the process and mitigated through careful design, development, and testing, with a focus on assuring safety, reliability, or redundancy where needed. In reality many companies do not have an adequate understanding of the risks embedded in their systems as they take short cuts through their engineering process, for example often not defining a clear CONOPs, not generating requirements, or skipping potentially critical testing gates. Often pressure to get a product to

market motivates such short cuts, or a lack of monetary resources (extensive testing can be very expensive) or a combination of both.

For safety critical systems like those in transportation and medicine, it is because of the propensity of some companies to take such short cuts that regulatory bodies are needed so that an independent assessment can be made as to whether such technologies are truly safe and ready for large scale deployment. In the next sections, we will contrast and compare how three different regulatory bodies view their role in regulating aviation, medicine, and surface transportation with an emphasis on which stage of the V-Model these agencies get involved in technology regulation.

**The FAA Approach to Regulation of New Technologies**

Commercial passenger airline safety is at an all-time high, with an accident rate of one fatal accident for every 16 million flights as of 2017 (Shepardson 2018). The aerospace industry has led the broader transportation industry in terms of incorporating significant automation into vehicles, with planes routinely landing themselves at hundreds of airports worldwide. Because of the routine use of advanced automation in commercial aviation, reviewing FAA regulation of aircraft safety provides a window into how one regulatory agency deals with safety-critical automation and related complex systems.

The FAA asserts that safety is its primary value in overseeing air transportation (FAA 2018). In support of this mission that emphasizes safety first and efficiency second, the FAA insists that any aerospace company wishing access to the US national airspace first must engage in a *Partnership for Safety* Plan. This program defines the working relationship between a company and the FAA independent of any specific project, in order "to build mutual trust, leadership, teamwork, and efficient business practices (AIA, AEA et al. 2017)." Trust is a key element for working with the FAA and working with them as early as possible can, in theory, help to reduce regulatory delays by identifying issues and requirements for certification of new technologies and/or procedures.

Once this partnership is established, when an aerospace company wants to develop a new aircraft, or even part of a new aircraft like an aircraft engine, the FAA's aircraft certification process begins. The most onerous regulatory certification is called a "type certification" and refers to the approval process for the design of an entire aircraft, including subcomponents like jet engines and avionics. Much like the V-Model in Figure 1, type certification consists of five phases: conceptual design, requirements definition, compliance planning, implementation, and post certification (AIA, AEA et al. 2017). However, this systems engineering approach to regulation is designed to structure the information flow between the company and the FAA, in effect giving both sides a script that guides the regulatory process.

The FAA's Conceptual Design phase consists of "familiarization meetings" between the FAA and company to discuss the new product idea and its concept of operations. The goal is to identify critical areas and difficult certification issues at the outset and to develop shared awareness of what the company hopes to achieve. A "Project Specific Certification Plan" is started during this phase, which includes a project timeline, checklists for moving on to the next phases, means of compliance, testing plans, and other project management information. In

effect, the FAA uses this plan to determine how aligned a company is with a systems engineering plan.

The FAA Requirements Definition stage focuses on how a company plans to address specific regulations and methods of compliance. Specific regulations known as airworthiness standards address all safety-related aspects of an aircraft including such items as weight limits, takeoff speeds, fabrication methods, and even passenger information signs. If a new product feature is not a covered by existing airworthiness regulations, a "special condition" assessment may be needed. A special condition is a new regulation particular to that aircraft, typically developed through notice-and-comment rulemaking, which fills the gap in existing airworthiness standards. Rulemaking can be a time and labor-intensive process, so communication between a company and the FAA at the Requirements Definition stage allows for earlier processing of a perceived need for a special condition.

The Compliance Planning phase consists of completing the Project Specific Compliance Plan, including finalizing airworthiness standards and detailing test plans to show compliance with those standards. It is at this stage where the critical role of the Designee is formalized. Under statutory authority dating to at least 1958, the FAA delegates some of its certification responsibilities to employees of those companies with an established Partnership for Safety Plans. These employees, known as Designees, theoretically provide efficient collaboration between companies and the FAA, and thus reduce the regulatory footprint. The legitimacy of the designee system has not always been viewed in a positive light (Henderson, Scott et al. 2013, Koenig 2018)

In the Implementation phase, the company works closely with the FAA/designee to ensure that certification requirements are met. Testing is an important component of this phase and tests only count towards certification if the FAA agrees prior to testing as to the purpose and scope of the tests. Thus, testing for new technologies is a mutually agreed upon plan. When all compliance activities are complete, the type certification approval can be issued. By the time a new aircraft is fully approved for commercial operation, as much as 8 years may have passed since the beginning of the type certification process (Boeing Commercial Airplanes 2013).

The final phase, Post Certification, focuses on continued airworthiness through maintenance and operational awareness, as well as conformance to approved manufacturing processes. Any subsequent required inspections are typically carried out by company's Designees and overseen by the FAA. If a safety risk is discovered post operational deployment, the FAA can implement an "Airworthiness Directive" through notice-and-comment rulemaking (14 CFR §39), which makes it unlawful to operate an aircraft unless an inspection or some other mitigating action occurs. Airworthiness directives, which happen relatively frequently, allow the FAA to manage risk and address potential threats to safety in a relatively expeditious manner.

The FAA clearly approaches regulation of aviation systems through the lens of system engineering. The FAA builds trust and confidence in those companies that wish to fly in the national airspace through an intensive relationship building process. While such a regulatory approach is comprehensive, it also can be extremely costly and add significant time to the deployment of new technologies, which is often at odds with rapid design and implementation cycle of many autonomous vehicle technologies, particularly those coming from Silicon Valley.

Indeed, the FAA's traditional and process-laden approach to regulation has caused issues for the emerging unmanned aerial vehicle, aka drone, industry. After years of delay, the FAA was pushed through industry lobbying to address the growing demand for commercial drone use and accommodate them in the national airspace (Morgan and Seetharaman 2015). In addition to commercial applications for small drones weighing less than 55lbs, there is currently a push to begin passenger-carrying drone operations, which would involve significantly more reliance on autonomous systems than what is in current drone systems. However, the regulatory obstacles presented by the FAA for doing so are substantial (Nneji, Stimpson et al. 2017), and so it is not clear how these regulatory barriers could be overcome to take advantage of what could be transformational transportation technologies.

**The FDA Approach to Regulation of New Medical Devices**

Medical devices are another potential outlet for autonomous technology application, including surgical robots, drug delivery systems, and artificial intelligence-based diagnostic programs. When such technology falls within the statutory definition of a medical device as an instrument, machine, or a contrivance "intended for use in the diagnosis . . . cure, mitigation, treatment, or prevention of disease", it falls into the regulatory purview of the FDA (21 USC §321(h)).

Tasked to both protect the public health and advance it through innovation, the FDA's medical device evaluation program attempts to ensure the safety and effectiveness of proposed systems in treating patients without placing overwhelming regulatory obstacles in the way of device developers (Maisel 2015). In general, the FDA requires all medical device companies to follow a set of manufacturing best practices, which also includes controls on how a device is developed. These "design controls" essentially mandate that a medical device company follow a version of a systems engineering framework that includes design planning, input, output, review, verification, validation, and transfer (21 CFR §820). Although the FDA does not directly monitor a developer's adherence to a system engineering process like the FAA does, the FDA's stance is that "Design controls increase the likelihood that the design transferred to production will translate into a device that is appropriate for its intended use (21 CFR §820.30)."

Communication between medical technology developers and the FDA begins when a medical device developer wants to start testing a new device on humans, which occurs just before the System Validation stage of the V-Model in Figure 1. Human trials cannot be conducted without prior FDA clearance under an Investigational Device Exemption (IDE, 21 CFR §812.1). Because a device might attack a disease or condition in a novel way as compared to a currently approved device, no standard metrics of success can necessarily apply to new medical devices, which is why an applicant must work with the FDA to determine the criteria, end-points, and objectives of clinical trials. Compulsory meetings before clinical trials thus serve to help determine what the goals of the testing will be and also mitigate risk.

Acquiring an IDE requires meetings with the FDA and submission of significant information about the device. This information includes a device description, drawings, components, specifications, materials, principles of operation, analysis of potential failure modes, proposed uses, patient populations, instructions, warnings, training requirements, clinical evaluation criteria and testing endpoints, and summaries of bench or animal test data or prior clinical experience (Center for Devices and Radiological Health 2001).

The FDA currently has two main regulatory pathways for placing a new medical device on the market: premarket approval (PMA) and a 510(k) clearance. PMA is the more stringent of the two and applies to devices intended to support or sustain human life or when the device presents a potentially unreasonable risk of illness or injury (21 USC §360c(a)(1)(C)). For PMA, the FDA must determine that sufficient, valid scientific evidence exists to indicate the proposed device is safe and effective for its intended use (21 CFR 814). Thus, a PMA applicant generally must provide results from clinical investigations involving human subjects showing safety and effectiveness data, adverse reactions and complications, patient complaints, device failures, and other relevant scientific information (21 CFR 814.20(6)(ii)).

In 2005, the FDA estimated that reviewing one PMA application costs the agency an average of $870,000 (Crosse 2009). One survey of medical device companies found that it took an average of 54 months to reach approval from first communication with the FDA about a potential innovation. The same survey found that the average total costs for a medical device company from the time of product conception to approval was $94 million, although these costs cannot all be attributed to compliance activities (Makower, Meer et al. 2010).

The 510(k) pathway for approval is more straightforward as it applies to moderately risky devices. Generally, developers only need to show that the new device is "substantially equivalent" to a "predicate" device already on the market. A predicate device is one that was available on the market before 1976, or any device cleared since then via 510(k). In contrast to PMA, human-subject clinical trials for safety and effectiveness are typically not required (Johnson 2016). However, the FDA can respond to a 510(k) application by requesting additional information it deems relevant (CFR §807.100(a)(3)), and data from device testing is typically provided for agency review.

A 510(k) application is significantly cheaper for the FDA to review, at an estimated average cost of $18,200 per application (Crosse 2009). A company's total costs from product concept to clearance is around $31 million on average with an average time of 10 months from first submission of an application to clearance (Makower, Meer et al. 2010). This faster timeline and the lower evidentiary requirements may make 510(k) appealing to device companies over PMA, but it also potentially allows them to circumvent the systems engineering process.

Similar to the FAA, the FDA is struggling to understand the impact of autonomy as it relates to medical devices in the future. While the FDA does not have much experience in regulating autonomous systems that incorporate probabilistic reasoning, it does have experience with automated medical devices that operate with within well-defined feedback control loops such as the Automated External Defibrillator (AED) and automated insulin monitoring and delivery devices (FDA 2018). Just recently the first autonomous medical device, i.e., one that leverages probabilistic reasoning, was approved by the FDA, which allows a system to autonomously detect diabetic retinopathy (US FDA 2018). This device, developed through a relatively new FDA Breakthrough Devices Program, another version of a 510(k) approval, is only a diagnostic device and takes no action based on the information it senses. Thus, it remains to be seen how the FDA will approach the regulation of truly closed-loop autonomous systems that leverage probabilistic reasoning to both detect and then take action based on input.

**The NHTSA Approach to Regulation of New Technologies**

Federal regulation of the automotive industry falls to NHTSA, whose mission is to "save lives, prevent injuries, and reduce economic costs due to road traffic crashes through education, research, safety standards, and enforcement activity (NHTSA 2018)." NHTSA has authority over motor vehicles and related equipment including all components, accessories, and software which impacts the safety of a vehicle. With respect to cars, trucks, motorcycles, and other motor vehicles on public roadways, NHTSA attempts to assure safety through two mechanisms: minimum safety standards and recall authority (NHTSA 2016).

NHTSA administers the Federal Motor Vehicle Safety Standards (FMVSS), which provide minimum safety requirements to be followed by vehicle manufacturers (49 CFR §571.) Established through notice-and-comment rulemaking, the FMVSS consist of 73 separate standards grouped generally into three categories: crash avoidance, crashworthiness, and post-crash survivability. These minimum safety standards address safety-related aspects of a vehicle, including headlights, braking systems, turn signals, electronic stability control, seat belts, motorcycle helmets, bus emergency exits, flammability, and many others (49 CFR §571). The FMVSS can be very specific, dictating sub-component requirements as well as the objective tests needed to show compliance (e.g., 49 CFR §571.104 S4.1.2). Thus, the FMVSS can essentially dictate design requirements and testing standards, which streamline the systems engineering process for automotive manufacturers at the component level.

While NHTSA sets the FMVSS compliance tests, it does not independently test vehicles for compliance before they reach the market. Instead, manufacturers must self-certify that their vehicles comply with all relevant FMVSS. Moreover, while the FMVSS address verification testing, NHTSA does not require manufacturers follow a specific certification process. Instead, manufacturers are allowed to take those actions they deem meet compliance standards (Office of Vehicle Safety Compliance 1998). They are also expected to monitor vehicles post marketing for any compliance concerns.

After vehicles are on the market, the Office of Vehicle Safety Compliance (OVSC) buys cars from real-world new-car dealerships and then tests 30 of the 44 testable FMVSS on these cars. Due to budget limitations, these tests are effectively quality assurance spot checks, with the majority of vehicle makes and models never tested, although OVSC prioritizes testing high risk vehicles (Office of Vehicle Safety Compliance 1998). According to NHTSA, instances of manufacturer non-compliance with significant safety compromises are rare (NHTSA 2016).

As long as manufacturers meet the FMVSS, they are not prevented or restricted in adding new features into vehicles. Thus, unlike the FAA and the FDA, when an automotive manufacturer wishes to insert a new technology in a vehicle, there is no expectation of discussions or formal notification to the regulatory body as long as the FMVSS are met. Indeed, NHTSA generally only initiates dialogue with manufacturers when a defect is suspected in a car already released to the public or when manufacturers seek an exemption to the FMVSS.

NHTSA considers a problem to be a "defect" if it "poses an unreasonable risk to motor vehicle safety (NHTSA 2016)." A defect determination can be made whether the cause of the defect is known or unknown. In theory, NHTSA will act on a suspected defect so long as there is a likelihood that the suspected defect could or has caused crashes, injuries, or deaths. Once a defect is identified, NHTSA notifies the manufacturer, who can then choose to remedy the defect through repair, replacement, or refund (49 USC §30120). NHTSA has the authority to carry out

civil enforcement actions, like mandating a recall, and can impose civil penalties if manufacturers do not comply with orders to remedy defects (NHTSA 2016). However, NHTSA's actual ability to investigate defects is limited, with only 20 investigators in 2017 and a budget of less than $23 million dollars to fund defect investigations for the 250 million cars on American roadways (NHTSA 2016).

In addition to conversations that happen between NHTSA and manufacturers over possible defects, the other event that triggers formal dialogues are requests for FMVSS exemptions. This exemption process was designed to allow automotive manufacturers the ability to conduct more streamlined development and field evaluation of new safety technologies, with an assumption that these proposed new technologies would provide equivalent or better levels of safety as compared to legacy technologies. Under this exemption process, manufacturers can sell up to 2500 cars per year for two years (49 USC 301). However, such requests are rare, with reportedly only 8 such exemptions requested since 1994 (Fraade-Blanar and Kalra 2017).

Of the three federal regulatory agencies facing revolutionary changes due to autonomous systems, NHTSA by far has been the agency most targeted by industry and lobbyists to relax its regulatory approach. With more than 80 billion dollars invested in the research and development of cars that are self-driving (a car that occasionally requires human intervention) or driverless (a car with no controls available to a human), the autonomous vehicle industry is anxious to realize returns on its investments and see the FMVSS as a major barrier (Kerry and Karsten 2017). Because manufacturers are currently limited to 2500 cars per exemption, the existing exemption process is generally seen as too restrictive and major efforts are underway to remove these regulatory restrictions so that these experimental cars can be sold on the market (Fraade-Blanar and Kalra 2017, Kulisch 2018).

A major problem with the removal of such regulatory restrictions designed to ensure safe technologies operate on public roads is establishing whether these new autonomous vehicles (either self-driving or driverless) truly have equivalent or better levels of safety than current cars. The 2018 death of a pedestrian caused by an Uber self-driving car highlighted the brittleness of an autonomous vehicle's perception system, as well as an overreliance on human vigilance meant to prevent such engineering flaws from surfacing (Laris 2018).

In addition, there have been numerous deaths attributed to the Tesla Autopilot system, which is a precursor technology to self-driving cars but currently sold as a driver-assist feature and not covered by the FMVSS. While NHTSA absolved Tesla of any defects in the Autopilot system after the death of a driver in Florida in 2016 (Habib 2017), the National Transportation Safety Board (NTSB), an independent government agency responsible for determining the probable cause of transportation accidents, found that Tesla did bear responsibility in the design and marketing of the Autopilot system (NTSB 2017). Additional NTSB and NHTSA investigations are currently underway for another Tesla Autopilot-related fatality in California in 2018.

As a result of these fatalities and other problems exhibited by autonomous cars, many have called for the development of new testing protocols such as computer vision tests to ensure these new and unproven technologies are ready for the myriad of conditions they could face in different environments (Claybrook and Kildare 2018, Cummings in press). However, it is not clear how and when such testing would be inserted into NHTSA's regulatory process, especially since it currently regulates cars *after* they have been sold on the market. Thus, regulation occurs in the

operations and maintenance block in Figure 1, and not in the testing blocks like what occurs for the FDA and FAA.

To better elucidate how the different agencies can both learn from one another but also recognize where true regulatory innovation is needed, the next section compares and contrasts the three different agencies approach to regulation, with specific focus on regulation for automated and autonomous technologies.

**Lessons Learned Across Aviation, Automotive, and Medical Device Regulation**

Given that the FAA, FDA, and NHTSA are all agencies that regulate safety-critical systems, it is instructive to compare how these agencies go about their approval processes and also examine some of their past difficulties to shed light on how they are likely to deal with autonomous systems in the future. To this end, the following sections compare when the three agencies intervene in the regulatory process, how they have dealt with past automation surprises, and how a bias towards regulating technologies based on equivalence could set a dangerous precedent.

**Point of First Contact**

FAA, FDA, and NHTSA clearly have very different approaches in regulating new technologies, however they all, in principle, share the same primary mission which is to promote safety. As illustrated in Figure 2, one of the major differences in their approaches to new technology regulation is the point-of-first-contact between manufacturers and the regulatory agencies, which is overlaid on the V-Model from Figure 1 to gain a better understanding of both the time and engineering maturity stage differences.

The timing of these regulatory interventions should be viewed through the lens of how the notion of *safe enough* has developed in each of the agencies over time. Political and public discourse have significantly shaped each agency's culture, and often major events in an agency's history often shape new policy. Moreover, each agency has had very different experiences and timelines in terms of dealing with advanced computer systems, which will be discussed below.

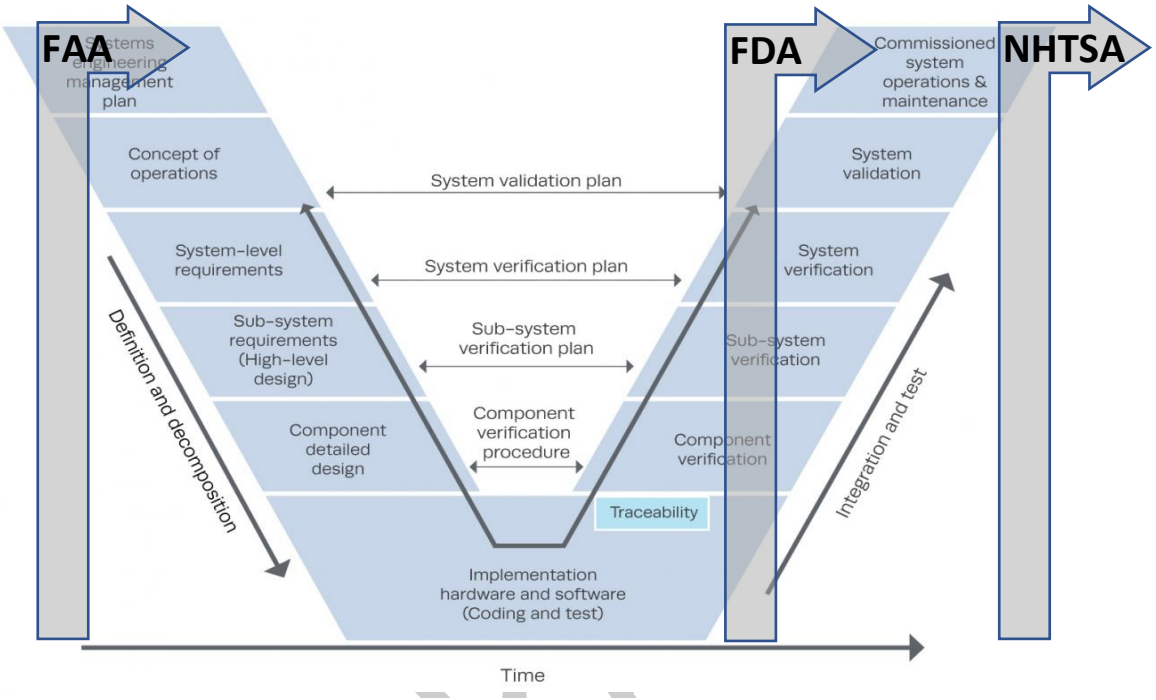


**Figure 2: Regulatory agency first formal point of contact with manufacturers**

The FAA, as depicted in Figure 2, is the earliest intervening agency in terms of advancing new technologies, especially those that affect flight safety. The FAA's early intervention in the systems engineering process of a technology and continuous relationship building with aerospace companies, while time consuming and expensive, also helps companies to avoid major post-market problems. This approach, which often takes longer than companies would like, has led to continued decreasing accident rates, with 2017 as the safest year on record for commercial air travel (Shepardson 2018).

In contrast to the FAA, the FDA only becomes involved in medical device regulation once the system test validation stage is reached, as seen in Figure 2. The FDA's middle ground approach likely reflects the nature of regulating potential life-or-death medical devices but also an understanding that many medical devices are eliminated internally due to a lack of funding or poor bench testing results. The FDA regulatory process also adheres to a systems engineering process which helps to mitigate the risk, especially for unproven technologies. However, as will be discussed in more detail in a following section, allowing companies to bypass a more rigorous systems engineering approach by claiming equivalence to an existing technology moves the point of contact to the right. This movement allows companies to bypass more expensive test and evaluations and shifts the burden of testing from the company to the public.

Lastly, when looking at the motor vehicle regulatory process, NHTSA has almost no involvement in a manufacturer's systems engineering process and only actively regulates technologies after problems emerge post-market. Clearly the primary advantage to this approach is that it allows companies to get technologies to market faster and more cheaply. Whether this is an effective regulatory approach is terms of safety is a debated topic, especially as systems in cars grow more complex. While annual motor vehicle fatality rates steadily fell from the late 1970s, reaching their lowest point in 2011, they have been trending upwards since that time (NSC 2018), suggesting that the current status quo is not sufficient.

One example of how increasing complexity in automotive designs is at odds with current post-market regulatory practices is the Toyota unintended acceleration problem in the 2009-2010 timeframe. During this time period, at least 89 people were killed when Toyota cars reportedly accelerated in an uncontrollable manner, causing catastrophic accidents (Associated Press 2010). In 2011, NTHSA and the US Department of Transportation absolved Toyota of any technical liability for these instances, and instead blamed drivers (US DOT 2011).

However in 2013, Toyota paid 1.6B to settle a class action lawsuit over these incidents (Trop 2013). Then a few months later a jury found Toyota responsible for two unintended acceleration deaths, with expert witnesses citing bugs in the software and throttle fail safe defects. In e-mails released as part of the legal proceedings, Toyota executives admitted that they were not familiar with systems engineering practices like the V model, and that considering deign of failsafe technologies "is not part of the Toyota's engineering division's DNA (Koopman 2014)." In a related finding, in 2014 the US Department of Justice fined Toyota $1.2B for covering up known safety problems with the system which were related to mechanical issues with the mats (Douglas and Fletcher 2014).

This case highlights that waiting to regulate until post-market problems emerge for complex systems with embedded software is a dangerous strategy, especially when manufacturers do not understand the importance of engaging in well-established risk mitigation strategies for safety-critical systems. In the end, the cost of not regulating the insertion of advanced automated technologies in this case were many deaths and billions of dollars in settlements and fines. Unless changes are made in how NHTSA currently regulates such technologies, this problem is likely only to get worse with the introduction of autonomous cars, which are far more complex with no standards for testing or safety.

**Automation Surprises**

Since the 1970s, commercial airline accident rates and fatalities have fallen due to improvements in air traffic control and aircraft technologies, as well as crew training. However, in the early 1990s, there was an uptick in accidents that can be partially attributed to insertion of new automated flight control technologies which caused a series of episodes often termed as "automation surprises (Sarter, Woods et al. 1997)". More specifically, these accidents were caused by a disconnect in the design of the advanced automation and pilot interfaces, with a series of airline accidents worldwide caused by pilot misunderstanding of what new advanced fly-by-wire technologies were doing (Ben-Yosef 2005). This phenomenon is termed "mode

confusion", which indicates that an operator does not realize what mode the automation is in and may take actions counterproductive actions as a result (Bredereke and Lankenau 2002).

Significant research into this phenomenon followed this realization and resulting improvements have been made in aviation human-automation interfaces, as well as improved training. However, mode confusion continues to be a problem even in commercial airline operations today. Given that accident rates have decreased despite continued problems with mode confusion, one could argue that the FAA's early and frequent interactions with industry has provided checks and balances to this emergent problem, despite the fact that the FAA has not issued any specific regulations to guard against mode confusion.

The FDA has similarly struggled with understanding the impact of a similar phenomenon in the medical devices they regulate. One such example is the da Vinci robotic surgical assistant, which allows surgeons to remotely manipulate very small surgical tools through teleoperation. Such systems are automated and not autonomous, meaning that the computer only translates a surgeon's commanded inputs into specific actions and no probabilistic reasoning occurs.

In addition to the significant number of technical and engineering problems with the da Vinci, research has demonstrated that almost 9% of all problems with da Vinci can be attributed to a form of mode confusion called unintended operation (Alemzadeh, Raman et al. 2016). Despite a significant number of lawsuits, both past and present, the FDA has not specifically addressed these mode confusion problems, which also occur in other computerized medical devices like patient-controlled analgesic pumps (Obradovich and Woods 1997) and in radiation therapy (Leveson and Turner 1995). Given that autonomous systems are even more difficult to understand, it is likely these mode confusion problems will lead to even more problems unless specifically acted upon by regulatory agencies.

Automation can produce additional surprises in the form of emergent latent software failures, which has been experienced by the both commercial airplanes (Johnson and Holloway 2007) as well as medical devices like the da Vinci (Alemzadeh, Raman et al. 2016). NHTSA is particularly susceptible to these problems since the Federal Motor Vehicle Safety Standards do not address software safety. In 2015, software bugs in general were responsible for 15% of automotive recalls, and the number of software-related components involved in recalls grew by almost 600% between 2011-2015 (Steinkamp 2016). NHTSA's inexperience in understanding the systems engineering principles behind software testing and safety assurances means that NHTSA is ill-prepared to understand what safety regulations are needed for autonomous cars, which present an exponential leap in software complexity.

**Regulating Technologies vis-à-vis equivalence**

One commonly-used regulatory tool for approving technologies that are deemed low risk are to provide exemptions and expedited approval processes. As discussed previously, the FDA has attempted to streamline the regulatory process of medical devices by providing a shorter and cheaper path to market though the 510(k) process for technologies deemed to be of lower risk. A key element of such approvals is the fact that new technologies can be shown to be equivalent to existing technologies in some way. This test of equivalence is not clearly defined and can vary widely across devices, and this practice has recently been widely criticized.

An example of one such decision for an automated medical device is the approval process for the da Vinci surgical robot (Rabin 2013). The da Vinci received approval in 2000 from the FDA through the 510(k) process (Jones 2000), claiming equivalence to existing laparoscopic surgical technologies. In 2014, Intuitive, the manufacturer of da Vinci, earmarked $67 million dollars for settlements for over 3000 product liability claims (Baron 2017).

The substantial settlements which are on-going with at least 52 pending lawsuits across 22 states suggest that the FDA failed to accurately assess the risk of this new technology, which clearly was not low. As discussed in the previous section on automation surprises, technology with embedded software, even if it performs the same function as an analog device, cannot be considered equivalent because of the unique and opaque nature of software code. If and when true autonomy is inserted in the da Vinci or other robotic surgery systems, it is not clear whether the FDA will understand the substantial increase in risk or continue to let manufacturers use the less rigorous 510(k) process which only requires limited testing, effectively allowing manufacturers to bypass well-established systems engineering practices.

Autonomous technologies are new and unproven, with no accepted software testing practices. Even the Department of Defense is struggling to determine what processes and technologies are needed to test autonomous systems (Ahner and Parson 2016). Given the significant unknowns and the fact that these are safety-critical systems, it seems as if the "precautionary principle (Wiener 2017)" should prevail. Where the risks of a new technology are unproven or unpredictable, the precautionary principle suggests (or in alternative formulations, demands) that technology be banned from public use until scientific methods show that the risks are acceptably low. When the risks include long-term, irreversible damage, of the three regulatory approaches featured here, it appears that the FAA's approach is the most precautionary.

This is not to say that the FDA and NHTSA should necessarily adopt those practices of the FAA, but rather that these and other agencies need to work together to define an approach to the regulation of autonomous systems. If a technology, whether it is autonomous or not, is truly low risk, then perhaps NHTSA's post market approach and even the FDA's 510(k) approach is warranted. But all agencies need to recognize that autonomous systems are not equivalent to any systems on the market and new regulatory approaches to these high risk, probabilistic reasoning safety-critical systems are needed.

**Conclusion**

When considering three different regulatory agencies (FAA, FDA, NHTSA) that certify technologies in safety-critical systems, the FAA's approach to regulation is the most conservative and precautionary in nature, in that a new product is not allowed to be put to real-world use without extensive review and testing. NHTSA's approach is exactly the opposite, namely that it assumes a new technological feature on a car is safe enough to put on the market unless shown to pose an unreasonable risk after introduction into the marketplace. The FDA is somewhere in the middle, as it leans more towards the precautionary approach by requiring extensive testing for new medical devices unless equivalency can be shown with an existing device, in which case a more limited pre-market review is applied.

Past history of regulation of advanced automated safety-critical systems, specifically the da Vinci and the Toyota unintended acceleration case studies, demonstrate that the FDA and NHTSA approaches to regulation may not accurately assess the risk to public safety particularly for embedded software systems. While aviation has not been immune to similar problems, the FAA's precautionary regulatory process has provided a layer of safety checks and balances not afforded by the other agencies' expedited or post-market regulation approaches.

Given that autonomous systems will be much more difficult to assess in terms of risk to the public than previous automated technologies, regulatory agencies need to reassess when and how they make their first points of contacts with manufacturers of autonomous systems, as well as understanding the importance of tracking a company's use of established systems engineering practices that help to reduce risk. Moreover, these agencies need to realize that because of the complexities of probabilistic reasoning software that is at the heart of autonomous systems, traditional approaches to regulation may not support their overall mission of public safety.

To this end, new regulatory approaches are needed that likely involve not just reassessing current practices, but also incorporate expertise from a larger group of stakeholders. In order to develop a set of regulatory best practices that permit innovation while ensuring public safety for autonomous systems, regulatory agencies will need to include company representatives like software engineers, much like the FAA currently does for its type certifications. In addition, because of the nascent and unproven nature of autonomous technologies, these regulatory agencies should routinely call on academics and government experts from places like federally funded research and development centers (FFRDCs) as independent reviewers.

Autonomous systems across medicine and transportation have the potential to usher in a new age of personalized services, which could dramatically improve the quality of life for a broad cross-section of society. However, autonomous systems represent a significant increase system complexity, with engineers and computer scientists still struggling to understand their own creations in this space. Given that there is no consensus on how to test embedded probabilistic reasoning in autonomous software systems to ensure equivalent or better levels of safety, it seems prudent that regulatory agencies, at this point in time, should take a more precautionary approach until more confidence is gained, especially for application in safety-critical systems.



**Acknowledgments**

This paper was supported in part by the US Department of Transportation and the University of North Carolina's Collaborative Sciences Center for Road Safety (CSCRS).

.

## References


Ahner, D. K. and C. R. Parson (2016). Workshop Report: Test and Evaluation of Autonomous Systems. STAT T&E Center of Excellence. Wright-Patterson AFB, OH, US Air Force.

AIA, AEA, GAMA and FAA Aircraft Certification Service and Flight Standards Service (2017). The FAA and Industry Guide to Product Certification. FAA. Washington DC, US Department of Transportation**:** A1-30.

Alemzadeh, H., J. Raman, N. Leveson, Z. Kalbarczyk and R. Iyer (2016). "Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data." PLoS One **11**(4).

Andrews, A., M. Abdelgawad and A. Gario (2016). World Model for Testing Autonomous Systems Using Petri Nets. IEEE 17th International Symposium on High Assurance Systems Engineering (HASE). Orlando, FL**:** 65-69.

Ashford, N. A. and R. P. Hall (2012). "Regulation-Induced Innovation for Sustainable Development." Administrative & Regulatory Law News **21**(3).

Associated Press (2010). Sudden Acceleration Death Toll Rises. New York Times. New York, New York Times**:** B2.

Bar-Yam, Y. (2003). When Systems Engineering Fails --- Toward Complex Systems Engineering. 2003 IEEE International Conference on Systems, Man and Cybernetics, Washington, DC, IEEE.

Baron, E. (2017). Robot-surgery firm from Sunnyvale facing lawsuits, reports of death and injury The Mercury News. San Jose, Bay Area News Group.

Ben-Yosef, E. (2005). The Evolution of the US Airline Industry: Theory, Strategy and Policy. New Tork, Springer US.

Blanchard, B. S. and W. J. Fabrycky (1998). Systems Engineering and Analysis. Upper Saddle River, NJ, Prentice Hall.

Boehm, B. (1988). "A Spiral Model of Software Development and Enhancement." Computer: 61-72.

Boeing Commercial Airplanes. (2013). "Certifying Boeing Airplanes."

Bredereke, J. and A. Lankenau (2002). A Rigorous View of Mode Confusion. SafeComp, Bremen, Germany, Springer Verlag.

Calo, R. (2014). The Case for a Federal Robotics Commission. Washington DC, Brookings.

Center for Devices and Radiological Health (2001). Early Collaboration Meetings Under the FDA Modernization Act (FDAMA); Final Guidance for Industry and for CDRH Staff. US Food and Drug Administration. Washington DC, Department of Health and Human Services,.

Claybrook, J. and S. Kildare (2018). "Autonomous vehicles: No driver…no regulation?" Science **361**(6397): 36-37.

Crosse, M. (2009). Shortcomings in FDA's Premarket Review, Postmarket Surveillance, and Inspections of Device Manufacturing Establishments. C. o. E. a. C. Testimony Before the Subcommittee on Health, House of Representatives,. Washingotn DC, Government Accounting Office,.

Crowder, J. A. and S. A. Friess (2013). Systems Engineering Agile Design Methodologies. New York, Springer.

Cummings, M. L. (in press). Adaptation of Licensing Examinations to the Certification of Autonomous Systems," Safe, Autonomous and Intelligent Vehicles. Unmanned System Technologies. X. Li, R. Murray, C. J. Tomlin and H. Yu, Springer.

Douglas, D. and M. A. Fletcher (2014). Toyota reaches $1.2 billion settlement to end probe of accelerator problems. Washington Post Washington DC, Washington Post.

FAA. (2018). "Mission." from https://www.faa.gov/about/mission/.

Favarò, F. M., D. W. Jackson, J. H. Saleh and D. N. Mavris (2013). "Software contributions to aircraft adverse events: Case studies and analyses of recurrent accident patterns and failure mechanisms." Reliability Engineering & System Safety **113**: 131-142.

FDA (2018). FDA approves automated insulin delivery and monitoring system for use in younger pediatric patients. Washington DC, Department of Health and Human Services.

Fraade-Blanar, L. and N. Kalra (2017). Autonomous Vehicles and Federal Safety Standards: An Exemption to the Rule? Santa Monica, CA, RAND Corporation.

Griggs, T. and D. Wakabayashi (2018). How a Self-Driving Uber Killed a Pedestrian in Arizona. New York Times. NY, NY, The New York Times Company.

Grissinger, M. (2012). "Misprogramming Patient-Controlled Analgesia Levels Causes Dosing Errors." Pharmacy and Therapeutics **37**(2): 74–75.

Habib, K. (2017). Investigation: PE 16-007. Office of Defects Investigations. Washington DC, US Department of Transportation

Henderson, P., A. Scott and T. Hepher (2013). Insight: Will Dreamliner drama affect industry self-inspection? Reuters. London, Reuters Thompson.

Hutchins, A. R., M. L. Cummings, M. Draper and T. Hughes (2015). "Representing Autonomous Systems' Self-Confidence through Competency Boundaries.". 59th Annual Meeting of the Human Factors and Ergonomics Society, Los Angeles, CA.

Jaffe, A. B., S. R. Peterson, P. R. Portney and R. N. Stavins (1995). "Environmental Regulation and the Competitiveness of U.S. Manufacturing: What does the Evidence Tell Us?" Journal of Economic Literature **33**(1): 132-163.

Johnson, C. W. and C. M. Holloway (2007). The Dangers of Failure Masking in Fault-Tolerant Software: Aspects of a Recent In-Flight Upset Event. 2nd Institution of Engineering and Technology International Conference on System Safety. London**:** 60-65.

Johnson, J. A. (2016). FDA Regulation of Medical Devices, Congressional Research Services,.

Jones, U. (2000). "FDA Clears Robotic Surgical System." Retrieved July 17, 2018.

Kalra, N. and S. M. Paddock. (2016). Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability? Santa Monica, CA, RAND.

Kerry, C. F. and J. Karsten (2017). Gauging investment in self-driving cars. Washington DC, Brookings.

Knight, W. (2017). The Dark Secret at the Heart of AI. MIT Technology Review. Cambridge, MA, MIT.

Koenig, D. (2018). US watchdog criticizes FAA oversight of American Airlines. Associated Press. Washington DC.

Koopman, P. (2014). A Case Study of Toyota Unintended Acceleration and Software Safety, Carnegie Mellon University.

Krisher, T. and J. Billeaud (2018). Police: Backup driver in fatal Uber crash was distracted. Associate Press.

Kulisch, E. (2018). Lobbying push targets holdouts on autonomous vehicle bill. Automotive News, Reuters.

Laris, M. (2018). Fatal Uber crash spurs debate about regulation of driverless vehicles. Washington Post. Washington DC, Nash Holdings.

Leveson, N. G. and C. S. Turner (1995). An Investigation of the Therac-25 Accidents. Computers, Ethics & Social Values. D. J. Johnson and H. Nissenbaum. Upper Saddle River, NJ, Prentice Hall**:** 474-514.

Lietzen, I. (2017). Robots: Legal Affairs Committee calls for EU-wide rules. Brussels, European Parliament.

Maisel, W. (2015). Robotic Assisted Surgical Devices Workshop keynote. FDA. Silver Spring, MD, Department of Health and Human Services.

Makower, J., A. Meer and L. Denend (2010). FDA Impact on U.S. Medical Technology Innovation: A Survey of Over 200 Medical Technology Companies. Stanford, CA, Medical Device Manufacturers Association

Maney, K. (2017). Goldman Sacked: How Artificial Intelligence Will Transform Wall Street. Newsweek Magazine. NY, NY, Newsweek Media Group.

Marcus, G. (2018). " Deep Learning: A Critical Appraisal." arXiv.

Mitre (2014). Systems Engineering Guide. McLean, VA, The Mitre Corporation.

Morgan, D. and D. Seetharaman (2015). Industry lobbyists take aim at proposed FAA drone rules. Technology News. London, Reuters Thompson.

NHTSA (2016). Budget Estimates Fiscal Year 2017. NHTSA. Washington DC, US Department of Transportation.

NHTSA (2016). Federal Automated Vehicles Policy:Accelerating the Next Revolution In Roadway Safety. Department of Transportation. Washington, DC.

NHTSA (2016). NHTSA Enforcement Guidance Bulletin 2016–02: Safety-Related Defects and Automated Safety Technologies US FDA. Washington DC, Federal Register.

NHTSA. (2018). "NHTSA's Core Values." from https://www.nhtsa.gov/about-nhtsa.

Nneji, V., A. Stimpson, M. L. Cummings and K. Goodrich (2017). Exploring Concepts of Operations for On-Demand Passenger Air Transportation,. AIAA Aviation, Denver, CO.

NSC (2018). NSC Motor Vehicle Fatality Estimates,. Itasca, IL, National Safety Council.

NTSB (2017). Highway Accident Report: Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida May 7, 2016. National Transportation Safety Board. Washington DC.

Obradovich, J. H. and D. D. Woods (1997). "Users as Designers: How People Cope with Poor HCI Design in Computer-Based Medical Devices." Human Factors **38**(4): 574-592.

Office of Vehicle Safety Compliance (1998). Compliance Testing Program. NHTSA. Washington DC, US Department of Transportation

Rabin, R. C. (2013). Salesman in the Surgical Suite. New York Times.

Sarter, N. B., D. D. Woods and C. E. Billings (1997). Automation surprises. Handbook of Human Factors and Ergonomics. G. Salvendy. New York, Wiley**:** 1926-1943.

Schwartz, A. (2017). "Hybrid closed-loop insulin delivery systems for Type 1 diabetes come of age." Stanford Medicine News Retrieved July 1, 2018, from https://med.stanford.edu/news/all-news/2017/04/hybrid-insulin-delivery-systems-for-type-1-diabetes-come-of-age.html.

Shepardson, D. (2018). 2017 safest year on record for commercial passenger air travel. Reuters. London.

Steinkamp, N. (2016). 2016 Automotive Warranty & Recall Report, Stout Risius Ross, Inc.

Trop, J. (2013). Toyota Will Pay $1.6 Billion Over Faulty Accelerator Suit. New York, New York Times**:** B3.

US DOT (2011). U.S. Department Of Transportation Releases Results From NHTSA-NASA Study Of Unintended Acceleration In Toyota Vehicles. Washington DC, US Department of Transportation.

US FDA (2018). FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Silver Spring, DM, Department of Health and Human Services.

Villasenor, J. (2014). Products Liability and Driverless Cars: Issues and Guiding Principles for Legislation. Washington DC, Brookings.

Wiener, J. B. (2017). Precautionary Principle. Principles of Environmental Law L. Krämer and E. Orlando.

Woolley, S. C. and P. N. Howard (2017). Computational Propaganda Worldwide: Executive Summary. Working Paper 2017.11. Oxford, UK, University of Oxford.