A Patch with the Dev-il: How the Development of Autonomous Vehicle Software Will Challenge Attorney-Client Privilege

Dawson Honey*

INTRODUCTION

As Hurricane Irma approached the Florida coast, Tesla released a software update that temporarily boosted the range of all of their vehicles in the area to help those fleeing the storm.[1] The proliferation of software features in traditionally mechanical vehicles poses novel ethical questions to developers, engineers, and lawyers.[2] Software patches can be applied remotely, allowing manufacturers to instantaneously shift the behavior of entire fleets of vehicles.[3] However, not all vehicle software updates proceed from good intentions. In 2015, Volkswagen circumvented pollution limits by installing software that temporarily improved the engine's emissions whenever it detected that the vehicle was being tested.[4]

Autonomous Vehicle ("AV") software is on a collision course with complicated ethical questions that will entangle the moral choices of developers with the legal obligations of the lawyers who represent them.[5] The hardware to control vehicles already exists, but the software for truly independent driving is an ongoing effort that has seen steady improvement through regular patches.[6] The Hurricane Irma incident demonstrates that an AV company could introduce an unannounced software patch to all of their vehicles that would instantly change the way their autonomous features operate.

A central issue in the morality of AVs is how software will be programmed to deal with conflicting or ambiguous situations. Computers are not limited by

---

* J.D., Georgetown University Law Center (expected May 2022); B.A., The College of Wooster (2019).

1. Brian Fung, *As Hurricane Irma bore down, Tesla gave some Florida drivers more battery juice. Here's why that's a big deal*, WASH. POST (Sept. 11, 2017), https://www.washingtonpost.com/news/innovations/wp/2017/09/11/as-hurricane-irma-bore-down-tesla-gave-some-florida-drivers-more-battery-juice-heres-why-thats-a-big-deal/ [https://perma.cc/95NQ-LZFY].

2. *See generally* Sven Nyholm & Jilles Smids, *The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem?*, ETHIC THEORY MORAL PRAC. (Jul. 28, 2016).

3. *See id.*

4. Russell Hotten, *Volkswagen: The scandal explained*, BBC (Dec. 10, 2015), https://www.bbc.com/news/business-34324772 [https://perma.cc/3SNX-PVKV].

5. Nyholm & Smids, *supra* note 2 at 1281.

6. *See* Steven Loveday, *Elon Musk Says Tesla Is 'Very Close' To Level 5 Self-Driving Technology*, INSIDE EEVs (Jul. 9, 2020), https://insideevs.com/news/433141/elon-musk-tesla-level-5-autonomous-driving/ [https://perma.cc/7DNR-KP6C].

instinct or reflex and have the ability to evaluate their surroundings in ways a human driver never could.[7] This allows AVs to make statistical, calculated choices in life-or-death situations.[8] What should an AV do if it encounters an obstacle on a city street and the only way around it is to break the law by crossing a double yellow line? If an AV carrying one passenger is about to hit two pedestrians, should it swerve to avoid them, endangering the life of the passenger instead?

Some states have already passed laws about AVs.[9] California has an extensive legislative scheme mandating that AV technology comply with federal law, safety regulations, testing procedures, and performance standards.[10] Though less comprehensive than the California statute, Nevada, Florida, and the District of Columbia have all passed similar laws providing rudimentary guidance for the manufacture, testing, and sale of AVs.[11] But these states are in the minority, as most states have no laws that explicitly govern AVs, and none have yet taken a stance on the ethical issues of programming.[12]

This Note evaluates how lawyers' attorney-client privilege responsibilities interact with the moral dilemmas of AV software programming. Rule 1.6 of the *Model Rules* obligates lawyers to retain confidentiality with their clients unless expressly permitted otherwise.[13] However, this obligation is not absolute, allowing disclosure for instances of imminent harm and ongoing illegal conduct.[14] This Note explores a hypothetical where an AV company consults its lawyer about its intention to remotely install a software patch to all of its vehicles with explicit instructions to give preference to the passenger's life. The question analyzed in this Note is whether the act of giving a machine instructions to choose one life over another falls within the scope of imminent harm or illegal conduct. If so, then the lawyer could break attorney-client privilege and disclose that information.

Part I of this Note recounts the ethical issues of AVs and the extent of current regulations on the topic. Part II covers the scope of attorney-client privilege and when a lawyer is permitted or compelled to reveal confidential communications. Part III explores how attorney-client privilege responsibilities are bent and broken by AV software. Finally, Part IV argues that a national solution is required because applying existing state laws to AV software ethics would result in perverse outcomes.

---

7. *Id.* at 1278.

8. *Id.*

9. *See* F. Patrick Hubbard, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. 1803, 1843 (2014).

10. CAL. VEH. CODE. § 38750 (2014).

11. NEV. REV. STAT. § 482A.100 (2013); FLA. STAT. § 316.86 (2013); D.C. CODE § 50-2352 (2014).

12. *See* Andrew R. Swanson, *"Somebody Grab the Wheel!": State Autonomous Vehicle Legislation and the Road to a National Regime*, 97 MARQ. L. REV. 1085, 1096−98 (2014).

13. MODEL RULES OF PROF'L CONDUCT R. 1.6 (2018) [hereinafter MODEL RULES].

14. MODEL RULES R. 1.6.

I. Morality and AVs

The trajectory of modern vehicle development aims to minimize human error through sophisticated driver-assistance software.[15] Current technology has not yet reached the point where the driver can be entirely removed from the equation.[16] However, modern vehicles are becoming more independent from human intervention using technology like lane assistance and crash avoidance.[17] Experimental AVs have already traveled millions of miles on the road and completed cross country journeys with almost no faults.[18] However, AVs are still imperfect, and if they are implemented across the country, even one-in-a-million faults become a near certainty.[19] This makes it practically inevitable that AVs will have to make tough ethical choices in dangerous situations. Part A discusses the thought experiment that is most commonly applied to discussions of ethics and AVs. Part B discusses how this thought experiment has played out in current AV laws and how regulators have addressed the question of when a vehicle truly becomes autonomous.

A. THE TROLLEY PROBLEM

A common thought experiment applied to the ethics of AVs is the "Trolley Problem."[20] In its most basic form, this problem asks the reader to imagine a railroad worker positioned at the junction of two tracks.[21] Suddenly, a runaway train car rounds the corner that, on its current course, will kill five other workers.[22] The worker at the track switch has the choice to let the train kill the five workers or divert it to a branch line where only one worker will be killed.[23] The key question is whether acting to kill one is worse than, through inaction, allowing five to die.[24]

---

15. *See* Hubbard, *supra* note 9 at 1843.

16. *See id.*

17. *See id.*

18. Keith Naughton, *Humans Are Slamming into Driverless Cars and Exposing a Key Flaw*, Bloomberg (Dec. 17, 2015), http://www.bloomberg.com/news/articles/2015-12-18/humans-are-slamming-into-driverless-cars-and-exposing-a-key-flaw [https://perma.cc/4AT5-DUMC]; Alex Davies, *This is Big: A Robo-Car Just Drove Across the Country*, Wired (Apr. 3, 2015), https://www.wired.com/2015/04/delphi-autonomous-car-cross-country/ [https://perma.cc/Y7EW-7U7T].

19. *See* Hubbard, *supra* note 9 at 1812−13.

20. Judith Jarvis Thompson, *The Trolley Problem*, 94 Yale L. J. 1395, 1395−96 (1985) (expanding on the work of Phillipa Foot, who first theorized the problem in 1967).

21. *Id.*

22. *Id.*

23. *Id.* at 1396.

24. *Id.* at 1399. While not relevant to this Note, this phrasing of the question is considered fighting words in some circles. There is considerable debate about whether the question ought to be "is killing one worse than letting five die?" or "is killing five worse than killing one?" *See id.*

B. AV SOFTWARE ETHICS

No bright line rule exists for when a vehicle with autonomous features becomes an AV.[25] There is a spectrum of features in vehicles that utilize sophisticated cameras and algorithms.[26] Even more ubiquitous features like anti-lock brakes utilize computers to some extent.[27] The trend towards automation has not escaped regulators. In 2016, the National Highway Traffic Safety Administration ("NHTSA") released a guidance document that established five ascending levels of automation.[28] At level one, a vehicle's automation will "sometimes assist" human conduct during "some parts" of the driving task.[29] At level five, the automated systems can perform all driving tasks under the same conditions as a human.[30] There are no current vehicles that have reached level five.[31] Most autopilot features are at levels two to four, where the autonomous software still requires significant human intervention.[32]

The Trolley Problem applies to AVs when the vehicle has to play the role of the railroad switch operator. Unlike a human driver, AV software is not hampered by the slowness of human reactions nor shielded by the interest of self-preservation.[33] When an AV reaches an impasse where either the pedestrian or the vehicle occupant must be severely injured or die, it is presented with a Trolley Problem dilemma.

Even rudimentary driver assistance software has the potential to create Trolley Problem issues. Theoretically, even automation levels two to four could still be programmed to make consequential decisions. For example, lane assistance software could be programmed to override the controls of the driver and swerve the vehicle off the road to avoid a head-on collision. NHTSA's policy does not take a hard stance about the ethical dilemmas that AV software will have to navigate.[34] Instead, NHTSA's policy only states that manufacturers should ensure that "such ethical judgments and decisions are made consciously and intentionally."[35] The vagueness of NHTSA's policy gives each manufacturer the discretion to decide

---

25. *Id.* at 1844.

26. *Id.*

27. *Id.*

28. NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., FEDERAL AUTOMATED VEHICLES POLICY 1.0, 9 (2016) [hereinafter NHTSA POLICY] (the definitions of automation were adopted from SAE International and are often referred to as "SAE levels").

29. *Id.* (lane departure detection and blind spot warnings are examples of a level one automation).

30. *Id.*

31. Brenda Goh & Yilei Sun, *Tesla 'very close' to level 5 autonomous driving technology, Musk says*, REUTERS (July 9, 2020), https://www.reuters.com/article/us-tesla-autonomous/tesla-very-close-to-level-5-autonomous-driving-technology-musk-says-idUSKBN24A0HE [https://perma.cc/KSU2-SF7C].

32. *Id.*

33. *See, e.g.*, Hubbard, *supra* note 9 at 1860−61.

34. NHTSA POLICY at 26.

35. *Id.*

whether to preference the life of the pedestrian or the passenger, setting the stage for potential conflicts with attorney-client privilege.

II. ATTORNEY-CLIENT PRIVILEGE AND EXCEPTIONS

The relationship between lawyer and client demands a higher standard of accountability that imparts a specific set of responsibilities on the former. Among the most hallowed of these responsibilities is the duty to retain strict confidentiality.[36] Confidentiality between lawyers and clients exists to create confidence in the relationship and, in some cases, to prevent defense counsel from becoming an agent of the prosecution.[37]

This relationship is strained when an attorney's loyalty to their client becomes an adversary of their obligation to the truth.[38] Because of this, attorney-client privilege is not impermeable. There are many exceptions, and this Note's hypothetical implicates two. First, client information may be revealed when the lawyer reasonably believes that disclosure is necessary to prevent imminent death or severe bodily harm.[39] Second, client information may be revealed, either through disclosure or through discovery, when the lawyer's services are in furtherance of an ongoing crime or fraud.[40] The key word in both of these rules is *may*, meaning that if either exception applies to this hypothetical, the lawyer could decide whether to disclose the software patch.

A. IMMINENT HARM

Rule 1.6 states that a lawyer may reveal client information "to prevent reasonably certain death or substantial bodily harm" or "to prevent the client from committing a crime or fraud that is reasonably certain to result in substantial injury to the financial interests or property of another and in furtherance of which the client has used or is using the lawyer's services."[41] Prior to 2002, Rule 1.6 only permitted a lawyer to break confidentiality when their client's "criminal" act was reasonably likely to result in imminent bodily harm.[42] Rule 1.6 has since been broadened, authorizing disclosure to prevent harm that "will be suffered immediately or . . . at a later date if the lawyer fails to take action necessary to eliminate the threat."[43] The example given in the *Model Rules* is where a lawyer becomes

---

36. *See* Note, *The Future Crime or Tort Exception to Communications Privileges*, 77 HARV. L. R. 730, 730 (1964).

37. *Id.*

38. *See* Emiley Zalesky, *When Can I Tell a Client's Secret? Potential Changes in the Confidentiality Rule*, 15 GEO. J. LEGAL ETHICS 957, 957−58 (2002).

39. MODEL RULES R. 1.6.

40. MODEL RULES R. 1.6(b)(2).

41. MODEL RULES R. 1.6(b)(1), (2).

42. MODEL RULES R. 1.6(b)(1); *see also* RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 66 (2000).

43. MODEL RULES R. 1.6 cmt. 6 (twenty-six states have adopted this rule and twenty-three retain a criminal conduct requirement).

aware that their client has spilled toxic waste into a town's water supply and disclosure is necessary to "eliminate the threat or reduce the number of victims."[44] In this situation, the lawyer is permitted but not required to disclose the incident.[45]

Even in situations where the danger does not come to fruition, the exception can still apply.[46] If the attorney has a reasonable belief that bodily harm is imminent, disclosure is permitted.[47] In *McClure v. Thompson*, the Ninth Circuit held that an attorney who had a reasonable, but erroneous belief that two murder victims were still alive and in imminent danger could validly disclose their locations conveyed by a client.[48]

On the other end of the spectrum, a harm may be too generalized to be considered imminent. In *R.L.R. v. State*, a Florida appellate court found that arguing a runaway juvenile was "a danger to himself" was *not* sufficient to invoke the imminent harm exception.[49] Many jurisdictions also do not consider the harms of execution as sufficiently "imminent" to permit disclosure of a client's confession that proves the condemned person innocent.[50] There is no consistent rule for when a harm becomes sufficiently imminent to invoke the exception, nor for when the probability of harm becomes reasonably certain.[51]

B. THE CRIME-FRAUD EXCEPTION

Even when the attorney-client relationship is clearly established, confidence does not extend to "communications made for the purpose of getting advice for the commission of a fraud or crime."[52] In courts, this is often expressed as the crime fraud-exception, where the "seal of secrecy" is removed for attorney-client communications and a lawyer's testimony can be lawfully compelled.[53] To overcome the attorney-client privilege through the crime-fraud exception, the party seeking information must make a *prima facie* case that (1) the client was committing or intending to commit a crime or fraud, and (2) the attorney-client communications were in furtherance of that alleged crime or fraud.[54] The proof required to demonstrate a *prima facie* case varies by jurisdiction but there are five

---

44. MODEL RULES R. 1.6 cmt. 6

45. MODEL RULES R. 1.6 cmt. 6

46. *See* McClure v. Thompson, 323 F.3d 1233, 1242−43 (9th Cir. 2003).

47. *Id.* at 1242.

48. *Id.* at 1243.

49. R.L.R v. State, 116 So. 3d 570, 574 (Fla. 3d Dist. Ct. App. 2013).

50. *See* Colin Miller, *Ordeal by Innocence: Why There Should be a Wrongful Incarceration/Execution Exception to Attorney-Client Confidentiality*, 102 Nw. U. L. REV. COLLOQUY 391, 391 (2008).

51. *See* Ken Strutin, *Preserving Attorney-Client Confidentiality at the Cost of Another's Innocence: A Systemic Approach*, 17 TEX. WESLEYAN L. REV. 499, 546−48 (2011).

52. United States v. Zolin, 491 U.S. 554, 563 (1989).

53. Drummond Co. v. Conrad & Scherer, LLP, 885 F.3d 1324, 1335 (11th Cir. 2018) (citing *Zolin*, 491 U.S. at 563).

54. *In re* Grand Jury Subpoena, 745 F.3d 681, 687 (3d Cir. 2014) (quoting *In re* Grand Jury, 705 F.3d 133, 151 (3d Cir. 2012)).

principles on which courts agree:[55] (1) a bare allegation of a crime or ongoing fraud is not sufficient; (2) mere suspicion that the lawyer's services are furthering a crime or fraud is not sufficient; (3) the movant does not need to conclusively prove the existence of a crime or fraud to trigger the exception; (4) a movant cannot rely on the communications at issue to demonstrate the alleged crime or fraud; (5) an unsuccessful movant may try again after additional discovery and testimony.[56]

Just the mere association with an attorney does not confer the protection of attorney-client relationships.[57] The role of an attorney is not to cloak illegal actions by lending them credibility.[58] Abetting illegality is so far outside the realm of a lawyer's duties that the position of attorney becomes merely incidental.[59] Thus, the existence of a privileged communication is a precondition for invoking attorney-client privilege.[60] In *United States v. Williams*, the Eighth Circuit held that the attorney-client privilege dissolved when a client involved his attorney in an ongoing conspiracy to sell drugs and launder money.[61] The conduct requested of the attorney was outside his "professional competence as an *attorney*, as opposed to a criminal enabler."[62] The statements made to the attorney were in furtherance of that criminal enterprise and were properly admitted into evidence.[63] In *United States v. Ivers*, the Eighth Circuit held that threatening statements a defendant made about a judge were not covered by attorney-client privilege.[64] The client was on a phone call with an attorney to discuss matters related to a civil lawsuit against an insurance company and expressed his intention and desire to kill the judge who was hearing his case.[65] A key fact to the court was that these threatening statements were not made "in order to obtain guidance in filing [the civil lawsuit]" and thus served no legal advisory purpose.[66] Similarly in *United States v. Matsa*, the Sixth Circuit found no protection for conversations where the defendant expressed intentions to tamper with witnesses.[67]

---

55. Douglas R. Richmond, *Understanding the Crime-Fraud Exception to the Attorney-Client Privilege and Work Product Immunity*, 70 S.C. L. Rev. 1, 24−25 (2018).

56. *Id.*

57. *See* Note, *The Future Crime or Tort Exception to Communications Privileges*, 77 Harv. L. R. 730, 730 (1964).

58. *Id.*

59. *Id.*

60. *In re* LeFande, 919 F.3d 554, 563 (D.C. Cir. 2019).

61. United States v. Williams, 720 F.3d 674, 688−89 (8th Cir. 2013) (the attorney participated in the conspiracy with the awareness of law enforcement and provided them with evidence that led to Williams's eventual conviction).

62. *Id.* at 688.

63. *Id.*

64. United States v. Ivers, 967 F.3d 709, 716 (8th Cir. 2020).

65. *Id.* at 717 (limiting the holding so that attorney-client privilege was only inapplicable to the portions of the call relating to the threat).

66. *Id.*

67. United States v. Matsa, 540 F. App'x 520, 526 (6th Cir. 2013).

These issues can become particularly tricky when an attorney represents a corporation.[68] The scope of corporate counsel privilege is complicated when an attorney's singular client can have hundreds of separate individuals all with their own autonomy. As a result, attorney-client privilege is much more porous in the context of individual employee communications but remains stronger for confidential company communications.[69] For the purposes of this hypothetical, the consultation between the lawyer and the AV company is assumed to be a privileged communication.

III. HOW AVs CHALLENGE ATTORNEY-CLIENT PRIVILEGE EXCEPTIONS

As this hypothetical is set squarely on the cutting edge, there is no direct case law on point regarding algorithms programmed to kill or injure pedestrians. This section proceeds in three parts. Part A compares this Note's hypothetical to the Ford Pinto, arguably the closest analog to the dilemma faced by AV software developers. Part B argues that laws on imminent danger likely do not extend to this hypothetical. Part C argues that laws on crime or fraud also do not neatly extend to this hypothetical, but could be analogized to even more callous applications of AV software.

A. THE FORD PINTO

Every vehicle is an amalgam of safety design compromises threading the needle between prohibitive expense and certain death.[70] Value judgment in vehicle safety is not without precedent and not every company strikes the correct balance. The Ford Pinto controversy stands as an exemplar of what happens when a manufacturer chooses savings over safety.[71] In the early 1970s, Ford designed the Pinto to be a subcompact car that would weigh fewer than 2,000 pounds and cost fewer than 2,000 dollars.[72] Throughout the design process, Ford knew the vehicle had a fatal flaw that could cause the gas tank to ignite in low speed rear-end collisions.[73] Ford chose to ignore engineers' suggestions for alternate designs, concluding that the cost of preventing fuel tank fires did not outweigh the savings of keeping the original design.[74]

The actions taken by Ford could be compared to the AV patch because both involve a utilitarian value judgment between two competing interests. However, there is a key difference that prevents the Pinto case from neatly applying to the

---

68. *See* H. Lowell Brown, *The Crime-Fraud Exception to the Attorney-Client Privilege in the Context of Corporate Counseling*, 87 KY. L. J. 1191, 1201−03 (1999).

69. *Id.* at 1209−10.

70. *See* Mark Geistfeld, *Reconciling Cost-Benefit Analysis with the Principle that Safety Matters More than Money*, 76 N.Y.U. L. REV. 114, 116−17 (2001).

71. Bryant Walker Smith, *The Trolley and the Pinto: Cost-Benefit Analysis in Automated Driving and other Cyber-Physical Systems*, 4 TEX. A&M L. REV. 197, 204 (2017).

72. *Id.*

73. *Id.*

74. *Id.* at 205 (Ford estimated that the litigation caused by the flawed design would still be cheaper than moving the gas tank).

dilemma of AV software. The central controversy of the Pinto case was Ford's choice to preference cost savings over safety.[75] In contrast, the AV problem is less about a balance between money and life and more about a balance between life and life. Pragmatic choices about the safety of vehicles and the probability of severe injury are distinct from providing the vehicle with instructions to give more weight to the lives of passengers. The former is best characterized as omission or negligence whereas the latter is a volitional action with intent. In the context of the AVs, a more apt comparison would be if Ford had implemented a self-destruct function that would detonate the gas tank if the vehicle were unable to avoid hitting a pedestrian.

B. THE SOFTWARE PATCH AS AN IMMINENT DANGER

Most jurisdictions are unlikely to consider programming AVs to kill or injure a pedestrian under rare circumstances as an imminent or reasonably certain harm.[76] Most jurisdictions are hesitant to hold that execution of an innocent defendant is an imminent harm, so the harm of AV life preferences is likely even more of a stretch.[77] However, whether a software patch is a reasonably imminent harm could depend on how many AVs are on the road. If a company has thousands of AVs nationwide and intends to apply the patch to all of them simultaneously, then the remote possibility of harm in each vehicle interaction may rise to the level of a reasonably certain, imminent harm.

Applying the imminent harm exception to life preferences in AVs results in a perverse outcome by creating a dissonance between the operative legal analysis and the goals of the *Model Rules*. This Note argues that a perverse outcome not only occurs when a set of rules produces a contradictory result, but also when applying rules makes the wrong facts dispositive. No two applications of law are alike, and rules can never be comprehensive enough to cover every minutia and edge case. Discussing the law of cyberspace in 1999, Lawrence Lessig argued that existing law should always be applied to new situations, unless that application would result in a perverse outcome.[78] The application of current legal ethics rules perversely shifts the key questions of AV liability; whether harm is imminent becomes dependent on the size and distribution of the company a lawyer represents. If the company is small, the harm would likely not be sufficiently imminent to invoke the exception. However, if the company is large and ubiquitous, remote circumstances become a near certainty. This is a perverse outcome

---

75. *Id.*

76. *See* Miller, *supra* note 50 at 391.

77. *Id.*

78. *Compare* Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 Harv. L. Rev. 501, 503 (1999) ("The choice among [legal] tools obviously depends upon their efficacy. But importantly, the choice will also raise a question about values. . . . We see something when we think about the regulation of cyberspace that other areas would not show us."), *with* Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. Chi. Legal F. 207 ("Develop a sound law of intellectual property, then apply it to computer networks . . . we do not know whether many features of existing law are optimal.").

because it shifts the question into an area that is detached from the key issues of AV software. When the dispositive facts are so drastically skewed, forwarding the policy goals of a Rule comes down to a matter of chance, making any outcome foundationally problematic even if the result aligns with the policy goals of the rule.

Under current law, programming life preferences into an AV is unlikely to fall within most definitions of imminent harm. Many have foretold about the coming conflict between law and autonomous vehicles.[79] Through the lens of tort law, the action of programming a life preference favoring passengers is likely too attenuated from any specific intent to harm to fall within most intentional torts.[80] The fact that the AV software does not know when and where these dilemmas will happen puts the conduct outside the scope of intentional torts, where generalized danger "at some undefined time and place" is insufficient.[81] If an action is too separated from a substantial harm to be considered a tort, then it is likely too separated to be considered an imminent harm as well.[82] This is not a question that the closest laws on point are well equipped to handle. As a result, the imminent harm exception would likely not apply to this Note's hypothetical.

C. THE SOFTWARE PATCH AS A CRIME OR FRAUD

The question of whether AV life preference software is outside of attorney-client privilege depends on whether the risk caused by the software patch is itself a crime.[83] If the attorney does not disclose their client's plan to patch in life preferences, their conduct may cross the threshold of the crime-fraud exception where their services are in furtherance of ongoing unlawful conduct.[84] As there is no comprehensive set of laws about abstract moral decision-making in the context of AVs, this would put the onus on courts to decide whether such conduct fits within existing classes of crime or fraud.

This hypothetical could be compared to the situation in *Ivers*, where an imminent threat to the safety of a judge hearing a defendant's case was outside the protections of attorney-client privilege.[85] If an attorney is consulted about their client's intention to patch in instructions to endanger pedestrians in certain circumstances, the harm that might ensue if the update were implemented could be considered close enough to a true threat to grant the attorney leave to disclose it.

---

79. *See, e.g.*, Jeffrey K. Gurney, *Crashing into The Unknown: An Examination of Crash-Optimization Algorithms Through the Two Lanes of Ethics and Law*, 79 ALB. L. REV. 183, 224−28 (2016); Jack Boeglin, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J. L. & TECH. 171, 185 (2015) (suggesting shifting the costs of AV liability away from the consumer).

80. *See* Gurney, *supra* note 79 at 227.

81. RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL & EMOTIONAL HARM § 1 cmt. e (2010).

82. *See* Gurney, *supra* note 79 at 227.

83. *See* Richmond, *supra* note 55 at 24−25.

84. *Id.*

85. United States v. Ivers, 967 F.3d 709 (8th Cir. 2020).

If the company's jurisdiction makes a distinction between acting to kill and allowing to die through inaction, then whether the act is a crime might depend on which person the AV chooses. This brings up a host of other questions about whether the action-inaction distinction even applies in the context of machines unburdened by the limitations of human reflex.

Alternatively, this hypothetical could be compared to dramatically "Pinto-esque" applications of AV software that strike a callous balance between life and cost. An AV could be updated with software that chooses between multiple unavoidable collisions by evaluating what injuries will occur and opting for the collision that exposes the manufacturer to the least liability. An AV could also be programmed to kill occupants or pedestrians rather than injure them if a company calculates that wrongful death lawsuits will be cheaper than injury lawsuits. On a macro scale, such software could also be implemented across fleets of AVs to evaluate whether the danger of increasing highway speeds would be outweighed by the cost savings in overall reduced travel time.

The scienter and cruelty of the above examples might make them into instances where the crime-fraud exception would neatly apply. With clever argument or a receptive judge, one of these situations could be analogized to choosing the life of one person over another. On the other hand, one counter argument to draw from the Pinto case would be to say that the world of automobile manufacturing is an inherently bloody business and there will always need to be some safety concessions.[86] Applied to this hypothetical, programming life preferences that would only apply in rare, emergency circumstances could be no more a crime or fraud than choosing not to make seat belts out of expensive high-thread Kevlar.

IV. Recommendations

Programming life preferences into AV software is not explicitly covered by the imminent harm or crime-fraud exceptions, but it ought to be outside the discretion of each manufacturer. Programming a vehicle with instructions to kill or injure in certain circumstances puts too much power in the hands of vehicle manufacturers. Current regulations on software and legal ethics are not sufficient to abate this risk. There needs to be more comprehensive regulation dictating what programming choices are prohibited in the context of AVs. This Note concludes with a normative appeal to regulators that, as a prudential matter, life preferences in AVs should not be left up to the discretion of vehicle manufacturers and the attorneys who represent them. Programming a distinct set of choices into an AV that places weighted values on life should be an act that is tightly controlled and regulated by a consistent set of laws. Attorneys who are consulted about an imminent, unannounced patch containing such features should not have to shield that information behind attorney-client privilege.

---

86. *See* Smith, *supra* note 71 at 204.

The Pinto case required extreme carelessness, but it does not exist in a vacuum. The Pinto's failures required a culture that allowed those responsible to convince themselves that it was just business as usual.[87] The most instructive aspect of the Pinto case is the danger of attaching a set value to human life. Statistically calculating the monetary impact of lost lives is a very common and perhaps unavoidable aspect of many industries, vehicle manufacturing included. On the regulatory front, NHTSA still utilizes statistical calculations that factor in the average value of life.[88] The argument of this Note is not that these calculations are inherently unethical, but rather that lawmakers should be cognizant of the mindset that a myopic focus on these numbers can create.

Statistics and probability can obscure ethical balancing and make it easy to cross a line without knowing it.[89] Without the proper abundance of caution, an AV design culture that is steeped in abstract hypotheticals can have a similar effect. The Trolley Problem is a constrained universe. In the real world, rather than flipping the switch or staying still, the operator can just yell "WATCH OUT." Choosing between the passenger or pedestrian risks creating a false dichotomy built on hypothetical absolutes. Life or death calculations made on the assumption of a frictionless plane should be disincentivized by any means possible.

The application of the Trolley Problem takes for granted that AVs have perfect information and can accurately calculate survival chances for all persons involved. This fails to account for errors that would be unique to a machine. This Note has only addressed the errors that are predicted to arise. There may be entirely new types of errors and miscalculations that have yet to be realized. After decisively beating all-time human champions Ken Jennings and Brad Rutter on *Jeopardy!*, IBM's "Watson" made a shocking and uncharacteristic mistake.[90] The Final Jeopardy! question asked the contestants to name a United States city whose largest airport is named after a World War II hero, and whose second largest is named after a World War II battle.[91] Jennings and Rutter both put down the correct answer,[92] but Watson answered with "Toronto."[93] After dominating two trivia champions, the supercomputer answered with a city that is not even in the right country.[94] Because it was Final Jeopardy!, Watson was forced to answer

87. *See* Gurney, *supra* note 79 at 227.

88. U.S. Dep't of Transp., Guidance on Treatment of The Economic Value of a Statistical Life (VSL). In U.S. Department of Transportation Analyses (2013).

89. *See* Smith, *supra* note 71 at 204.

90. Denyse O'Leary, *Why Did Watson Think Toronto Was in the U.S.A.?*, MIND MATTERS NEWS (Aug. 3, 2019), https://mindmatters.ai/2019/08/why-did-watson-think-toronto-was-in-the-u-s-a/ [https://perma.cc/C6TA-GL96].

91. *Id.*

92. *Id.* The correct response was "What is Chicago?" (O'Hare and Midway).

93. *Id.*

94. *Id.* In its page explaining the error, IBM offers a map of several towns in the United States also named "Toronto."

even though it had an extremely low confidence score and would never have buzzed-in had the question been asked in a regular round.[95] AVs may be prone to entirely new kinds of malfunctions that are currently unknown. A mistake like the Toronto error might arise in the context of AVs when they are given explicit instructions to assign value to human life. When the AV determines it must act, it may be forced by its programming to act in a situation of low confidence, just like Watson. In dire situations where the machine determines it has no choice, it could exacerbate an emergency by acting drastically on flawed or incomplete information. Creating software with inflexible instructions and obligations to act could lead to situations where an AV is under the mistaken impression that it must swerve in order to save a life, causing it to crash into an unseen obstacle without warning.

CONCLUSION

The development of AVs is a general good that promises to save thousands of lives every year.[96] However, the moral and ethical choices that AVs will inevitably have to make are not adequately covered by current laws. The *Model Rules* have a blind spot with regard to AV software ethics that threatens to strain the protections of attorney-client privilege. Programming AVs to give more value to one life over another could result in disastrous and unforeseen consequences. Neither the crime-fraud nor the imminent harm exceptions are well equipped to govern the issues that will arise when a lawyer is presented with this Note's hypothetical. The best laws are not those that resolve disputes, but rather those that prevent disputes from happening. The gaps in vehicle ethics law create ambiguity that will result in contentious and expensive litigation. A consistent standard would be the best way to obviate these disputes. How AVs contemplate difficult ethical choices, and the consequences of those choices needs to be closely scrutinized through targeted laws, rather than being left to the discretion of each vehicle manufacturer.

---

95. *Id.*

96. *See* Matthew Blunt, *Highway to a Headache: Is Tort-Based Automotive Insurance on a Collision Course with Autonomous Vehicles?*, 53 WILLAMETTE L. REV. 107, 109–12 (2017).