


5-10-2024

## Arbitrary and Capricious x Artificial Intelligence

Zoe E. Niesel

Follow this and additional works at: <https://scholarship.law.umn.edu/mjlst>

 Part of the [Administrative Law Commons](#), and the [Science and Technology Law Commons](#)

---

### Recommended Citation

Zoe E. Niesel, *Arbitrary and Capricious x Artificial Intelligence*, 25 MINN. J.L. SCI. & TECH. 1 (2024).  
Available at: <https://scholarship.law.umn.edu/mjlst/vol25/iss2/7>

# Arbitrary and Capricious x Artificial Intelligence

Zoe E. Niesel\*

## INTRODUCTION

In administrative law, there is perhaps no more misunderstood phrase than the “arbitrary and capricious” standard of judicial review. American administrative law’s cornerstone is the judicial review of agency decision-making—a review which has created multiple high-profile legal doctrines, like *Chevron* deference,<sup>1</sup> arbitrary and capricious review,<sup>2</sup> substantial evidence evaluation,<sup>3</sup> *Auer* deference,<sup>4</sup> and more. The American system of checks on administrative agencies relies on these doctrines with the view that an out-of-control agency

---

© 2024 Zoe E. Niesel

\* Albert Herrman Professor of Law and Associate Dean for Academic Affairs, St. Mary’s University School of Law. Thank you to the team at the Minnesota Journal of Law, Science & Technology for the hard work on this piece and for hosting an interest and insightful symposium.

1. *Chevron U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 842–43 (1984) (describing the still-evolving doctrine of when a court will defer to an agency’s interpretation of a statute which the agency administers and establishing that agency interpretations of statutory gaps or ambiguities will be accepted by the reviewing court if that interpretation is reasonable).

2. The subject of this paper, arbitrary and capricious review, refers to standard which will be used to review discretionary agency decisions and factual findings. *Association of Data Processing Serv. Orgs., Inc. v. Bd. of Govs. of the Fed. Rsrv. Sys.*, 745 F.2d 677, 684 (D.C. Cir. 1984).

3. *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 477 (1951). *See also* Matthew J. McGrath, *Convergence of the Substantial Evidence and Arbitrary and Capricious Standards of Review During Informal Rulemaking*, 54 GEO. WASH. L. REV. 541, 541 (1986) (“The APA designates the ‘arbitrary and capricious’ standard for judicial review of informal agency actions and the ‘substantial evidence’ standard, which is theoretically more rigorous, for review of formal, record-producing agency actions.”).

4. *Auer v. Robbins*, 519 U.S. 452, 461 (1997) (describing the doctrine by which the courts will defer to an agency’s interpretation of its own regulations unless the agency’s position is “plainly erroneous”).

can be called to account for its interpretations and decisions in the federal courts.<sup>5</sup>

The Administrative Procedure Act, in Section 706, establishes the standards of review for challenges to agency decisions. In reviewing the agency, the court has several flavors of review available. The court can review, for example: whether the agency appropriately reached conclusions of law;<sup>6</sup> whether the agency has appropriately interpreted statutory and regulatory language;<sup>7</sup> and whether the agency has made appropriate discretionary and factual determinations based on the information available to it in the administrative record.<sup>8</sup> Each of these types of judicial review comes with their own standards of evaluation. First, courts evaluate agency conclusions of law de novo when the agency does not administer the statute in question.<sup>9</sup> Second, when reviewing an agency's construction of the statute it administers, a more deferential standard applies under *Chevron v. Natural Resources Defense Council, Inc.*<sup>10</sup> If the agency interprets statutory gaps or ambiguities in a statute it administers, courts defer to reasonable agency interpretations per *Chevron*.<sup>11</sup>

Finally, discretionary agency judgments and factual findings are reviewed under Section 706's arbitrary and capricious standard. Specifically, Section 706 of the Administrative Procedure Act provides the relevant standard for judicial review of agency action by stating that a reviewing court will:

*[H]old unlawful and set aside agency action, findings, and conclusions found to be—*

---

5. See Katrine B. MacGregor, *Kennecott Utah Copper Corp. v. United States Department of the Interior: The Validity of Interior's Interpretation of "Promulgated" Within the Statute of Limitations Provisions of CERCLA*, 83 CORNELL L. REV. 1383, 1425 (1998).

6. David Zaring, *Reasonable Agencies*, 96 VA. L. REV. 135, 143–46 (2010).

7. *Id.*

8. William W. Buzbee & Robert A. Schapiro, *Legislative Record Review*, 54 STAN. L. REV. 87, 139–40 (2001).

9. Zaring, *supra* note 6, at 146 (“De novo review is appropriate when agencies are interpreting laws that they do not have a special responsibility to administer . . .”).

10. *Chevron U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837 (1984).

11. Catherine M. Sharkey, *Cutting in on the Chevron Two-Step*, 86 FORDHAM L. REV. 2359, 2362 (2018).

(A) arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law . . .<sup>12</sup>

The “arbitrary and capricious” review standard has been interpreted to govern the judicial review of agency factual findings that are produced through informal rulemaking or informal adjudication.<sup>13</sup> In contrast, agency factual findings that are produced through formal rulemaking or adjudication will be reviewed under the Administrative Procedure Act, Section 706(2)(E), which states that such findings can be set aside if they are “unsupported by substantial evidence.”<sup>14</sup> While the standards technically vary—factual findings that are the product of informal procedures will warrant arbitrary and capricious review, while factual findings that are the product of formal procedures will warrant substantial evidence review—most courts consider the standards to have converged, and for there to be little difference in what is analytically required between the two.<sup>15</sup> As such, for ease of discussion, this paper will focus on the arbitrary and capricious standard, but with the contextual background that a similar analytical approach is employed for the substantial evidence standard.<sup>16</sup>

So, what exactly is the arbitrary and capricious standard of review? The courts use a bouquet of phrases to describe what exactly they are doing in an arbitrary and capricious review of agency factual findings and discretionary decisions.<sup>17</sup> For example, courts will note that such a review requires “agency action be reasonable and reasonably explained” and that “the agency has acted within a zone of reasonableness and, in particular, has reasonably considered the relevant issues and reasonably explained the decision.”<sup>18</sup> Judicial review of agency actions under the arbitrary and capricious standard is meant to

---

12. 5 U.S.C. § 706(2).

13. Sharkey, *supra* note 11, at 2378–79.

14. 5 U.S.C. § 706(2)(E).

15. Association of Data Processing Serv. Orgs., Inc. v. Bd. of Govs. of the Fed. Rsrv. Sys., 745 F.2d 677, 684 (D.C. Cir. 1984). That said, the information to be reviewed will necessarily come from different places depending on whether formal or informal procedures were used. For formal procedures, the review will be limited to the administrative record, while for informal procedures the review encompasses any information that was available to the agency at the time it made its decision.

16. Zaring, *supra* note 6, at 167.

17. Louis J. Virelli III, *Deconstructing Arbitrary and Capricious Review*, 92 N.C. L. REV. 721, 737–38 (2014).

18. FCC v. Prometheus Radio Project, 592 U.S. 414, 423 (2021).

be deferential, but not entirely “toothless.”<sup>19</sup> Courts consistently state that the reviewing court does not simply rubber stamp the agency’s policy decision; instead, the standard has “serious bite.”<sup>20</sup> This includes verifying that the agency acted within the bounds of reasonableness and considered all the relevant issues and factors.<sup>21</sup> Despite such significant definitional treatment in the case law, many scholars have noted that the actual applied meaning of arbitrary and capricious review remains unclear.<sup>22</sup> Further, inconsistencies in application have been consistently noted.<sup>23</sup>

Enter artificial intelligence (AI). Starting in late 2022, there has been a veritable explosion of interest in AI, specifically in generative platforms, as a vehicle for legal reasoning, analysis, and understanding.<sup>24</sup> Such interest is well-placed, and it raises interesting questions about how AI can help us better understand legal standards.<sup>25</sup> Through increased explainability, AI has the potential to eliminate black box-style legal reasoning wherein decisionmakers apply umbrella-style legal standards without clarifying exactly how the standard operates.<sup>26</sup> Further, by processing dozens or hundreds of precedential applications of standards like “deliberate indifference” or “equitable considerations” in light of differing facts, AI models could infer traits that are applied in various legal standards across rulings.<sup>27</sup> Statistical relationships within the case data could be identified and enable clearer legal rules that specifically set

---

19. *Sw. Elec. Power Co. v. EPA*, 920 F.3d 999, 1013 (5th Cir. 2019).

20. *Wages & White Lion Invs., LLC v. FDA*, 16 F.4th 1130, 1136 (5th Cir. 2021).

21. *Texas v. United States*, 40 F.4th 205, 227 (5th Cir. 2022).

22. Sidney A. Shapiro & Richard E. Levy, *Judicial Incentives and Indeterminacy in Substantive Review of Administrative Decisions*, 44 DUKE L.J. 1051, 1065–66 (1995) (“[T]he arbitrary and capricious standard is relatively open-ended, and the Supreme Court has not given it more precise content.”).

23. Gary Lawson, *Outcome, Procedure and Process: Agency Duties of Explanation for Legal Conclusions*, 48 RUTGERS L. REV. 313, 322–23 (1996).

24. See generally Melissa E. Love Koenig & Colleen Mandell, *A New Metaphor: How Artificial Intelligence Links Legal Reasoning and Mathematical Thinking*, 105 MARQ. L. REV. 559 (2022).

25. *Artificial Intelligence for Lawyers Explained*, BLOOMBERG L. (Aug. 1, 2023), <https://pro.bloomberglaw.com/insights/technology/ai-in-legal-practice-explained>.

26. *Id.*

27. *Id.*

standards for thresholds of misconduct or balancing of interests.<sup>28</sup>

This Article subjects the murky and sometimes maligned<sup>29</sup> arbitrary and capricious standard of review to AI treatment to see if this technology shows promise for elucidating vague legal standards in more concrete and comprehensible ways (while still acknowledging the interpretive uncertainties). Part I examines the current state of arbitrary and capricious review in the courts, while Part II examines the current view regarding AI's ability to engage in judicial discretion and standard setting. Finally, Part III explores whether a generative AI platform like ChatGPT can clarify aspects of the arbitrary and capricious standard of review. The Article concludes that current AI may be helpful in a number of areas like determining procedural compliance, generating analytical checklists, or helping organize component pieces of various legal standards, but it fails to conduct the deep scrutiny required of an arbitrary and capricious review.

### I. A RIGOROUS BUT NARROW BUT SEARCHING BUT DEFERENTIAL “HARD LOOK”-STYLE ARBITRARY AND CAPRICIOUS REVIEW

The birth of the modern application of judicial review of agency factual and discretionary decisions comes from the 1970s and 1980s, and the creation of the Supreme Court doctrine known as “hard look review”<sup>30</sup> in 1983.<sup>31</sup> Prior to this period, judges on the D.C. Circuit had begun referring to judicial review of agency decisions as requiring a hard look, with the language explained as having the court require the agency to provide “explanations for their conclusions, to respond to counterarguments, to justify departures from past practices, and

---

28. Love Koenig & Mandell, *supra* note 24, at 598.

29. Zygmunt J. B. Plater & William Lund Norine, *Through the Looking Glass of Eminent Domain: Exploring the “Arbitrary and Capricious” Test and Substantive Rationality Review of Governmental Decisions*, 16 B.C. ENV'T AFFS. L. REV. 661, 665–66 (2022).

30. The term “hard look review” was initially coined by Judge Leventhal of the D.C. Circuit in 1974. Harold Leventhal, *Environmental Decisionmaking and the Role of the Courts*, 122 U. PA. L. REV. 509, 514 (1974) (“The court does not make the ultimate decision, but it insists that the official or agency take a ‘hard look’ at all relevant factors.”).

31. Nikol Oydanich, *Chief Justice Roberts’s Hard Look Review*, 89 FORDHAM L. REV. 1635, 1640 (2021).

to give careful consideration to alternatives.”<sup>32</sup> For example, in *Greater Boston v. FCC* in 1970, the D.C. Circuit discussed a substantial evidence review— which again is largely considered commensurate with an arbitrary and capricious review<sup>33</sup>—in the context of the judiciary’s role with respect to the federal agencies. Specifically, the D.C. Circuit noted that “[i]ts supervisory function calls on the court to intervene . . . if the court becomes aware, especially from a combination of danger signals, that the agency has not really taken a ‘hard look’ at the salient problems.”<sup>34</sup> Per the *Greater Boston* court, if “the agency has not shirked this fundamental task, however, the court exercises restraint and affirms the agency’s action even though the court would on its own account have made different findings or adopted different standards.”<sup>35</sup> As such, the scope of review described meant that as long as the agency demonstrates that it has rigorously examined the relevant issues and alternatives, using sound data and standards to inform its policy judgments, judicial deference remains warranted.<sup>36</sup> Thus, courts will uphold agency conclusions, even if suboptimal in clarity, as long as the path of decision-making can be reasonably discerned from the record.<sup>37</sup>

In 1971, the Supreme Court in *Citizens to Preserve Overton Park v. Volpe* used similar language to the *Greater Boston* court in describing a review under the APA’s Section 706(2)(A):

“[t]o make this finding the court must consider whether the decision was based on a consideration of the relevant factors and whether there has been a clear error of judgment . . . . Although this inquiry into the facts is to be searching and careful, the ultimate standard of review is a narrow one. The court is not empowered to substitute its judgment for that of the agency.”<sup>38</sup>

In 1983, the Court was seen to have formalized the idea of a hard look review of agency discretionary decisions and factual findings in *Motor Vehicle Manufacturers Association v. State*

---

32. *Id.* (quoting Thomas J. Miles & Cass R. Sunstein, *The Real World of Arbitrariness Review*, 75 U. CHI. L. REV. 761, 761 (2008)).

33. *Pac. Legal Found. v. DOT*, 593 F.2d 1338, 1343 (D.C. Cir. 1979).

34. *Greater Bos. Television Corp. v. FCC*, 444 F.2d 841, 851 (D.C. Cir. 1970).

35. *Id.*

36. *Id.*

37. *Id.*

38. *Citizens to Pres. Overton Park, Inc. v. Volpe*, 401 U.S. 402, 416 (1971).

*Farm Automobile Mutual Insurance Co.*<sup>39</sup> *Motor Vehicles* examined the National Highway Traffic Safety Administration's (NHTSA's) 1981 rescission of Motor Vehicle Safety Standard 208, a requirement that new vehicles be equipped with passive restraints, like automatic seatbelts or airbags, to protect occupants in crashes.<sup>40</sup> The key issue on appeal was whether NHTSA's rescission of this long-debated regulation was arbitrary and capricious under Section 706(2)(A) of the Administrative Procedure Act.<sup>41</sup>

The Supreme Court outlined principles for arbitrary and capricious review under the APA.<sup>42</sup> It emphasized that this review applied to revocations of existing regulations, rejecting the parties' suggestions that rescissions deserved more deference.<sup>43</sup> The court explained that while arbitrary and capricious review was meant to be narrow, courts had to ensure agencies examined relevant data and alternatives and coherently justified their policy choices by connecting facts to decisions.<sup>44</sup> As the Court made clear, agencies must "articulate a satisfactory explanation" with a "rational connection between the facts found and the choice made."<sup>45</sup> Courts must "consider whether the decision was based on a consideration of the relevant factors" or manifests clear judgmental error.<sup>46</sup> Typically, rules would be arbitrary and capricious

"if the agency has relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise."<sup>47</sup>

---

39. *Motor Vehicle Mfrs. Ass'n v. State Farm Auto. Mut. Ins. Co.*, 463 U.S. 29, 43 (1983).

40. *Id.* at 37–38.

41. *Id.* at 41.

42. *Id.* at 42.

43. The Court of Appeals had struggled with the question of the appropriate scope of judicial review for an agency's rescission of a rule, but the Supreme Court had no such issue, noting that "[t]he agency's action in promulgating such standards therefore may be set aside if found to be 'arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law.' . . . [T]he rescission . . . of an occupant protection standard is subject to the same test." *Id.* at 41.

44. *Id.* at 43.

45. *Id.*

46. *Id.*

47. *Id.*



Judges should “uphold a decision of less than ideal clarity if the agency’s path may reasonably be discerned.”<sup>48</sup> In sum, deferential, arbitrary and capricious review requires judicially confirming that the agency analyzed relevant inputs and alternatives and transparently explained its policy decisions in a rational manner.<sup>49</sup>

Turning to NHTSA’s rescission of Standard 208, the court faulted the agency for failing to consider the possibility of keeping a passive restraint rule focused solely on mandating airbags.<sup>50</sup> Given previous findings upholding airbag effectiveness, NHTSA had an obligation to address modifying the standard to achieve its safety objectives through the remaining technologically and economically feasible option before abandoning the regulation entirely.<sup>51</sup> The court also concluded NHTSA failed to adequately justify its reversal regarding detachable automatic seatbelts.<sup>52</sup> The agency determined these would not yield significant safety gains due to easy detachability, reversing its prior view that detachable belts could still improve usage rates and occupant protection.<sup>53</sup> But the court held that doubts over achieving precise predicted benefits did not automatically allow abandoning the standard, particularly given contrary real-world evidence of increased seatbelt usage by the passive design.<sup>54</sup>

The majority opinion<sup>55</sup> concluded by vacating the rescission of Standard 208 as arbitrary and capricious and remanding for

---

48. *Id.*

49. *Id.*

50. *Id.* at 46–47.

51. *Id.* at 47–51.

52. *Id.* at 51–52.

53. *Id.* at 52–54.

54. *Id.* at 54–57. Overall, the court ruled that NHTSA fell short of its APA obligations to confront alternatives clearly within reach under the existing standard and evidence that might sustain its regulation. *Id.* A reasoned analysis would require addressing why the identified problems could not be solved short of complete rescission of the regulation. *Id.* The agency also failed to adequately justify departures from prior findings in the record. *Id.*

55. In a concurring opinion, Justice Rehnquist wrote separately to emphasize the limited nature of arbitrary and capricious review. *Id.* at 58 (Rehnquist, J., concurring). Specifically, he did not agree that the NHTSA’s view of automatic seatbelts was arbitrary and capricious. *Id.* He noted that the majority had rejected the “agency’s explanation for its conclusion that there is uncertainty whether requiring installation of detachable automatic belts would substantially increase seatbelt usage.” *Id.* The agency’s conclusion was based

reconsideration consistent with its analysis of the proper application of the APA standard of review.<sup>56</sup> The court clarified it expressed no view on the ultimate merits but required further explanation and analysis from the agency bound by its statutory mandate.<sup>57</sup>

Although *Motor Vehicles* never uses the term “hard look review,” it is considered to have formalized the modern formulation of the standard and ushered in a new era of increased judicial supervision of agency decision-making.<sup>58</sup> The standard has persevered to this day, with little disruption by the Court.<sup>59</sup> But since *Motor Vehicles*, hard look review has been criticized for its impacts on agency decision-making.<sup>60</sup> Specific critiques focus on the incentive structure it creates for the agencies.<sup>61</sup> Since a hard look review will turn over the agency’s reasoning for its decision, the agency is incentivized to bloat its record with research, data, and reasoning in order to ensure success in a later judicial challenge.<sup>62</sup> Such ossification concerns

---

on the agency’s rejection of a study that showed an increase in seatbelt usage was achieved with passive belts when there was an interlock to prevent the car from operating unless the seatbelt was engaged. *Id.* Justice Rehnquist would have found that the agency’s decision to reject the study as supporting mandatory passive seatbelts more globally (i.e., in situations without an interlock) as adequate. *Id.* Justice Rehnquist noted that the “agency acknowledged that there would probably be some increase in belt usage, but concluded that the increase would be small and not worth the cost,” and that this was a “rational connection between the facts found and the choice made.” *Id.*

56. *Id.* at 57.

57. *Id.* at 56–57.

58. See Ganesh Sitaraman, *Foreign Hard Look Review*, 66 ADMIN. L. REV. 489, 500 (2014) (“Though it again did not use the ‘hard look’ phrasing, *State Farm* established that courts would play a significant role in reviewing agency action as arbitrary and capricious.”).

59. *Id.* at 501 (describing the hard look review doctrine as “remarkably stable”).

60. See, e.g., Bijal Shah, *Judicial Administration*, 11 U.C. IRVINE L. REV. 1119, 1130–31 (2021); Jacob Gersen & Adrian Vermeule, *Thin Rationality Review*, 114 MICH. L. REV. 1355, 1356 (2016).

61. Sydney A. Shapiro & Richard W. Murphy, *Arbitrariness Review Made Reasonable: Structural and Conceptual Reform of the “Hard Look”*, 92 NOTRE DAME L. REV. 331, 334 (2016).

62. Heath A. Brooks, *American Trucking Associations v. EPA: The D.C. Circuit’s Missed Opportunity to Unambiguously Discard the Hard Look Doctrine*, 27 HARV. ENV’T L. REV. 259, 269 (2003) (“As a result of hard look review, an already financially strapped agency must devote much more time and resources toward enacting a rule than it did a generation ago. Because an arbitrary and capricious challenge to any link or omission in the chain of

critique the increased cost, time, and resources spent on a rulemaking process that is ultimately a response to heightened judicial scrutiny.<sup>63</sup> Additionally, some have criticized the standard for creating a shield for courts to engage in policymaking, by providing a standard that allows the court a cover for giving itself the authority to make agency decisions.<sup>64</sup>

Finally, there is real concern that we still do not really know what it means to take a hard look at agency decision-making.<sup>65</sup> “Hard look” review has thus been described in all these different ways (and more):

- “It presumes agency action to be valid.”<sup>66</sup>
- Agency action must be “reasonable and reasonably explained” to withstand judicial review.<sup>67</sup>
- The agency must act within a “zone of reasonableness” and consider relevant issues and explain its decision.<sup>68</sup>
- Agencies “must examine relevant data” and articulate a satisfactory, rational explanation connecting facts to their policy choice.<sup>69</sup>

---

reasoning could invalidate the entire rule, agencies must prepare for the worst[.]”).

63. William S. Jordan, III, *Ossification Revisited: Does Arbitrary and Capricious Review Significantly Interfere with Agency Ability to Achieve Regulatory Goals Through Informal Rulemaking?*, 94 NW. U. L. REV. 393, 395 (2000) (identifying the theory that “hard look review results in excessive data gathering, analysis, and long-winded explanations, often of marginal points, all of which imposes unnecessary costs and delays upon the agencies’ regulatory program”); Sidney A. Shapiro, *Rulemaking Ossification and the Debate Over Reforming Hard Look Review*, ADMIN. & REG. L. NEWS 13, 13 (2015) (“Since agencies cannot know whether they will be subject to aggressive hard look review, they produce encyclopedic rulemaking justifications, seeking to dot every ‘I’ and cross every ‘T.’”).

64. Shah, *supra* note 60, at 1195.

65. Carrie Rosenbaum, *Arbitrary Arbitrariness Review*, 100 DENV. L. REV. 775, 796 (2002) (“There is ample controversy regarding what *State Farm* hard look mandates, if it mandates anything in particular, and its implications.”).

66. *Ethyl Corp. v. EPA*, 541 F.2d 1, 34 (D.C. Cir. 1976).

67. *FCC v. Prometheus Radio Project*, 592 U.S. 414, 423 (2021).

68. *Id.*

69. *Motor Vehicle Mfrs. Ass’n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983).

- Courts must set aside agency actions based on reasoning that fails to account for all relevant factors or evinces “a clear error of judgment.”<sup>70</sup>
- Agency action is generally arbitrary and capricious if: the agency relied on improper factors, entirely failed to consider “an important aspect of the problem,” offered an explanation “counter to the evidence,” or was implausible/not grounded in expertise.<sup>71</sup>
- Courts evaluate the agency’s contemporaneous justifications, not post-hoc rationalizations.<sup>72</sup>
- Courts may uphold less-than-clear agency decisions where the path is reasonably discernible.<sup>73</sup>
- An arbitrary and capricious standard of review requires that the court engages in a “searching and careful” process, but the court also cannot “substitute its judgement for that of the agency.”<sup>74</sup>
- Courts must ensure that “variables be accounted for, that the representativeness of test conditions be ascertained, that the validity of tests be assured and the statistical significance of results determined.”<sup>75</sup>
- “[T]he rejection of alternate theories or abandonment of alternate course of action [must] be explained[.]”<sup>76</sup>
- An agency cannot reach an outcome that contradicts the underlying record,<sup>77</sup> issue a rule based on political considerations rather than scientific study,<sup>78</sup> or allow private parties to provide the agency’s reasoning.<sup>79</sup>

The volume of words used across the case law to describe the hard look-style of arbitrary and capricious review shows that the

---

70. U. of Tex. M.D. Anderson Cancer Ctr. v. DHHS, 985 F.3d 472, 475 (5th Cir. 2021).

71. *Motor Vehicle Mfrs.*, 463 U.S. at 43.

72. *Michigan v. EPA*, 576 U.S. 743, 758 (2015).

73. *Motor Vehicle Mfrs.*, 463 U.S. at 43.

74. *Citizens to Pres. Overton Park, Inc. v. Volpe*, 401 U.S. 402, 416 (1971).

75. *Nat’l Lime Ass’n v. EPA*, 627 F.2d 416, 452–53 (D.C. Cir. 1980).

76. *Id.* at 453.

77. *Petroleum Commc’ns, Inc. v. FCC*, 22 F.3d 1164, 1172 (D.C. Cir. 1994).

78. *Midwater Trawlers Co-operative v. Dep’t. of Com.*, 282 F.3d 710, 720 (9th Cir. 2002).

79. *Texas Off. of Pub. Util. Couns. v. FCC*, 265 F.3d 313, 328 (5th Cir. 2001).

concept has abstract borders.<sup>80</sup> In some ways, the murkiness may suggest that the review is purposefully established as a framing reference for the judge rather than a strict test.<sup>81</sup> Ultimately, *State Farm* leaves us with an imprecise approach to what exactly this review is intended to accomplish.<sup>82</sup> Although the Court gave additional information about when something would be arbitrary and capricious—if the agency relies on factors which Congress has not intended it to consider, entirely fails to consider an important aspect of the problem, offers an explanation for its decision that runs counter to the record, or is implausible<sup>83</sup>—there is no evidence in the case law that these examples have helped clarify the standard.<sup>84</sup>

The *State Farm* Court suggested both a deferential and probing inquiry, and its actual analysis appeared to harken to the 1970's D.C. Circuit approach to hard look.<sup>85</sup> While the court warned that the judiciary was not to substitute its own judgment for that of the expert agency, it also had no problem with combing through the NHTSA's studies and disagreeing with its interpretation of those studies.<sup>86</sup> Specifically, the court found it arbitrary that the NHSTA had rejected a study showing that detachable passive belts would meaningfully boost seatbelt usage in a scenario where the vehicles had interlock devices preventing cars from starting unless belts were engaged.<sup>87</sup> But as the concurring Justice Rehnquist pointed out, the agency was only obligated to provide "a rational connection between the facts and the choice made," and it could certainly be rational for

---

80. Shapiro & Murphy, *supra* note 61, at 333.

81. Sidney A. Shapiro, *Rulemaking Ossification and the Debate Over Reforming Hard Look Review*, 41 ADMIN & REG. L. NEWS 13, 14 ("A judge who is to take a 'hard look' at the agency's explanation is likely to be in a different frame of mind than one who approaches a case as a 'pass-fail' professor.").

82. See Virelli III, *supra* note 17, at 728.

83. *Motor Vehicle Mfrs. Ass'n, Inc. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983).

84. Shapiro & Levy, *supra* note 22, at 1067 (finding that "[b]ecause the Supreme Court has not promoted use of the *State Farm* criteria, the definition of 'arbitrary and capricious' remains relatively indeterminate," and that in 118 surveyed cases, circuit courts only cited *State Farm* forty-five times, and only thirteen of those forty-five actually mentioned the *State Farm* criteria for arbitrary and capricious).

85. Shapiro & Murphy, *supra* note 61, at 334 n.16.

86. *Motor Vehicle Mfrs.*, 463 U.S. at 46–50.

87. Shapiro & Levy, *supra* note 22, at 1066 (calling this aspect of the majority's decision "aggressive substantive review").

the agency to refuse to generalize the study to situations where vehicles did not have mechanisms preventing movement without the seat belt being engaged.<sup>88</sup>

The majority's substantive review appears to be of the exact kind it had previously rejected the same year as *State Farm in Baltimore Gas & Electric*.<sup>89</sup> In *Baltimore Gas*, the Court deferred to an agency's technical expertise in evaluating data by noting that the agency's data or scientific determinations are the special purview of the agency, and that the Court's only role was to examine for a "rational connection between the facts found and the choice made."<sup>90</sup> As such, the deferential but searching tension remains taut.

## II. BRINGING AI INTO THE MIX

Because this article is interested in how AI will interpret and apply judicial review standards, it makes sense to examine how AI may perform in other judicial contexts. The actual use of AI to analyze facts and apply law in the same method as a judge remains a fantastical concept in the United States, outside some limited cases.<sup>91</sup> But that is not the case everywhere. For example, one real-world use case comes from China, where software is being used to analyze past cases with similar fact patterns to recommend potential criminal sentences to judges.<sup>92</sup> The goal is not to replace human decision making, but to provide additional guidance and normalization of sentences (i.e., deciding like cases alike).<sup>93</sup> As such, while the actual legal application will still come from humans and the human judge will still retain full discretion, the computer model will provide

---

88. *Motor Vehicle Mfrs.*, 463 U.S. at 58–59 (Rehnquist, J., concurring).

89. *Baltimore Gas & Elec. Co. v. Nat. Res. Def. Council, Inc.*, 462 U.S. 87 (1983).

90. *Id.* at 105.

91. Richard C. Kraus, *Artificial Intelligence Invades Appellate Practice: The Here, The Near, and The Oh My Dear*, ABA: APP. ISSUES (Feb. 5, 2019), [https://www.americanbar.org/groups/judicial/publications/appellate\\_issues/2019/winter/artificial-intelligence-invades-appellate-practice-the-here-the-near-and-the-oh-my-dear](https://www.americanbar.org/groups/judicial/publications/appellate_issues/2019/winter/artificial-intelligence-invades-appellate-practice-the-here-the-near-and-the-oh-my-dear) (noting that while AI has gained use in legal research applications, "more fantastic ideas such as using AI to objectively decide cases by analyzing facts and applying law—satirized in a *Daily Show* skit about a trial with the Honorable Amazon Alexa, presiding—are still figments of creative imaginations").

92. Rachel E. Stern et. al., *Automating Fairness? Artificial Intelligence in the Chinese Courts*, 59 COLUM. J. TRANSNAT'L L. 515, 526–27 (2021).

93. *Id.* at 527.

benefits like efficiency, speed, and alignment (with it up for grabs whether having judges align their outcomes with an algorithm is a good or bad thing).<sup>94</sup> In 2019, Estonia announced plans to have an AI adjudicate small claims decisions.<sup>95</sup> The plan was in keeping with the general Estonian pro-technology approach; the country already deployed thirteen artificial intelligence projects to replace government workers.<sup>96</sup>

In the United States, there has been squeamishness about AI judicial decision making, and with good reasons, like concerns about bias, ethical implications, and more.<sup>97</sup> The prospect of substituting human judges with AI adjudicators promises greater efficiency and neutrality, but also raises concerns about replacing human discretion and judgment cultivated through lived experience.<sup>98</sup> An AI judge may not appropriately take equitable considerations, emotional intelligence, human values, and motives into account.<sup>99</sup> But despite these concerns, interest in AI as legal decisionmaker remains high.<sup>100</sup>

Areas where AI decision-making has received the most attention are in bail and sentencing in criminal cases. With regards to sentencing, the judicial discretion employed carries high stakes and under existing human decision making there is a significant risk of inequitable outcomes and bias based on the views of beliefs of the individual human judges.<sup>101</sup> As such, a positive view of AI argues that sentencing is an area where AI

---

94. See *id.* (“In practice, what is billed as computer-assisted judging is likely to edge the Chinese courts toward a world in which judges seek to align their decisions with an outcome recommended by an algorithm.”).

95. See Eric Niiler, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, WIRED (Mar. 25, 2019), <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so>.

96. *Id.*

97. See, e.g., James Manyika et al., *What Do We Do About the Biases in AI?*, HARV. BUS. REV. (Oct. 25, 2019), <https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>.

98. Muharrem KILIÇ, *Ethical-Judicial Inquiry Regarding the Effect of Artificial Intelligence Applications on Legal Profession and Legal Practices*, 14 J. MARSHALL L.J. 202, 211 (2021).

99. *Id.*

100. Christopher Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs*, NAT’L INST. JUST. (Oct. 8, 2018), <https://www.nij.gov/journals/280/Pages/using-artificialintelligence-to-address-criminal-justice-needs.aspx>.

101. Dan Hunter et al., *A Framework for the Efficient and Ethical Use of Artificial Intelligence in the Criminal Justice System*, 47 FLA. ST. U. L. REV. 749, 782–83 (2020).

could effectively “judge” in a capacity better than humans—an appropriate program could more transparently weigh variables in sentencing, balance punishment proportionality, and lead to faster decisions.<sup>102</sup> This would rely, of course, on an algorithm that is publicly accessible in order to create transparency.<sup>103</sup>

But what scholars have noted about this pro-AI approach to sentencing is that “[c]ourts generally ‘duck’ the question of exactly how they weight the [varying] interests”<sup>104</sup> in a sentencing decision, while utilizing an AI can instead generate “explanations for their lines of reasoning and . . . provide authority for all assertions made and conclusions drawn.”<sup>105</sup> That is a particularly interesting point in relation to something like the AI’s ability to apply judicial review standards, like the arbitrary and capricious hard look review.<sup>106</sup> As noted above, even courts do not really understand what factors or inputs they are using in conducting an arbitrary and capricious review—they only know it is meant to be deferential but searching but narrowing but biting.<sup>107</sup> This has caused scholars like Professor Louis Virelli to propose deconstructing arbitrary and capricious review to better understand how it applies across various components of agency action, such as record building, reason giving, and input quality.<sup>108</sup> His 2014 paper proposes that breaking down this arbitrary and capricious review into particular decision components rather than an undifferentiated whole helps determine appropriate scrutiny levels for different agency functions.<sup>109</sup>

AI appears to be in a particularly unique place to take umbrella-style standards and turn them into more filtered rules, explain exactly how it does so, and how it applied those filtered

---

102. *Id.* at 783.

103. *Id.*

104. *Id.* at 785 (quoting Eric Engle, *Legal Interpretation by Computer: A Survey of Interpretive Rules*, 5 AKRON INTELL. PROP. J. 71, 92–93 (2011)).

105. RICHARD SUSSKIND, *TRANSFORMING THE LAW: ESSAYS ON TECHNOLOGY, JUSTICE AND THE LEGAL MARKETPLACE* 183 (2000).

106. *See generally Artificial Intelligence for Lawyers Explained*, BLOOMBERG L. (Aug. 1, 2023), <https://pro.bloomberglaw.com/insights/technology/ai-in-legal-practice-explained>.

107. Shapiro & Murphy, *supra* note 61, at 360.

108. Virelli III, *supra* note 17, at 724.

109. Virelli III, *supra* note 17, at 724–25.



rules to a given situation.<sup>110</sup> As such, it may be even more adept than human decisionmakers at understanding how to break a standard down into more concrete analytical steps or more defined input-output analysis.<sup>111</sup>

### III. CHATGPT AND THE QUESTION OF ARBITRARINESS

ChatGPT is a publicly accessible AI, specifically a generative pre-trained transformer.<sup>112</sup> The most defining characteristic of ChatGPT is its ability to generate highly coherent, detailed, and human-sounding responses to a vast range of open-ended inputs.<sup>113</sup> It does not simply retrieve answers from an existing database—it actively constructs responses word-by-word based on the patterns it has learned from training data.<sup>114</sup> This gives ChatGPT an impressive capacity to reply to many kinds of queries, explain complex concepts conversationally, and even adopt distinct voices and tones when asked.<sup>115</sup>

Since its launch, a major question in the legal industry has been whether ChatGPT, and similar programs like Anthropic's Claude or Google's Bard, have the capacity to conduct legal analysis.<sup>116</sup> Many have concluded that these programs, at least at this stage, are incapable of doing so, as they may inaccurately state the law, fail to appropriately weigh the facts, or simply

---

110. Frank Fagan & Saul Levmore, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, 93 S. CAL. L. REV. 1, 31 (2019) ("It may be apparent that AI can turn standards into rules . . . . It promises a more rule-oriented legal system.")

111. *See id.*

112. Eve Ross & Amy Milligan, *What Can ChatGPT Do, and Should We Let It?*, 34 S.C. LAW. 34, 35 (2023) ("Generative means the chatbot predicts how text can be recombined in a way that does more than merely repeat existing patterns. Pre-trained means that long passages of text were used as training data to show the chatbot what types of answers are expected. Transformer means the chatbot doesn't take everything the user says as a literal, absolute command.")

113. Harry Guinness, *How Does ChatGPT Work?*, ZAPIER (Sept. 6, 2023), <https://zapier.com/blog/how-does-chatgpt-work>.

114. *Id.*

115. *Id.*

116. Andrew M. Perlman, *The Implications of ChatGPT for Legal Services and Society* 1–2 (Suffolk U. L. Sch. Research Paper No. 22-14, 2022).

refuse to come to legal conclusions at all.<sup>117</sup> Additional concerns have been raised about ChatGPT in the legal profession, including its tendency to hallucinate information in regards to inquiries, which can include the production of non-existent cases, quotations, article references, and more.<sup>118</sup> But the future for ChatGPT in the legal profession is ultimately promising—ChatGPT has been able to pass the bar exam (in the 90th percentile of takers, no less), certainly suggesting that some type of legal analysis is within its grasp.<sup>119</sup> The very open-ended nature of applying multiple legal doctrines and precedents likely requires exposure to enormous volumes of case law for algorithms to sufficiently infer patterns of weighing factors against case specifics.<sup>120</sup> But areas like summarizing the implications of textual rulings, suggesting analogies or distinguishing cases, and surfacing relevant evidence appear more immediately attainable.<sup>121</sup>

Using the arbitrary and capricious standard of review, it is an interesting question whether ChatGPT at this time can unpack the standard, perhaps explaining its reasoning on how an arbitrary and capricious review should be conducted. To explore this idea, ChatGPT was provided with the following:

- The text of *State Farm* (including where it describes the arbitrary and capricious hard look review standard).
- Three additional cases where the arbitrary and capricious review standard was applied to agency action with varying outcomes—cases were chosen based on if they quoted the standard for hard look review directly from *State Farm*:
  - o Dep’t of Com. v. New York, 139 S. Ct. 2551 (2019);

---

117. Ross & Milligan, *supra* note 112, at 36 (“At present, ChatGPT is not even that useful . . . When asked to apply the law to the facts, ChatGPT’s responses were underwhelming.”).

118. Herbert B. Dixon Jr., *My “Hallucinating” Experience with ChatGPT*, 62 JUDGES J. 37, 37 (2023). *See also* Mata v. Avianca, Inc., No. 22-CV-01461-PKC, 2023 WL 4114965, at \*1 (S.D.N.Y. June 22, 2023) (imposing Rule 11 sanctions when attorneys in a brief “submitted non-existent judicial opinions with fake quotes and citations created by the artificial intelligence tool ChatGPT”).

119. Jeffrey Allen, *Maximizing Efficiency: Using ChatGPT While Working on the Road*, 40 GPSOLO 93, 94 (2023); Matt Reynolds, *How ChatGPT and Other AI Platforms Could Dramatically Reshape the Legal Industry*, ABA J. 34, 36 (2023).

120. Perlman, *supra* note 116, at 1.

121. Perlman, *supra* note 116, at 1.

- o Nat. Res. Def. Council v. EPA, 658 F.3d 200 (2d Cir. 2011);
- o Greater Yellowstone Coal., Inc. v. Servheen, 665 F.3d 1015 (9th Cir. 2011).

With these materials, ChatGPT was asked to create a model, framework, or approach for a judge to use in applying the arbitrary and capricious standard. It supplied the following:

*Step 1: Establish the Scope of Review*

- Narrow Review: Confirm that the review does not involve substituting the court's judgment for that of the agency.*
- Focus: Center on whether the agency has made a decision based on a consideration of the relevant factors and evidence.*

*Step 2: Examine the Agency's Decision-Making Process*

- Data and Explanation Review: Assess if the agency examined relevant data and articulated a satisfactory explanation for its decision.*
- Rational Connection: Determine if there is a logical link between the facts found and the choice made.*

*Step 3: Evaluate Consideration of Relevant Factors*

- Comprehensive Assessment: Check if the agency considered all important aspects of the issue.*
- Error in Judgment: Look for clear errors or misjudgments in the agency's decision-making process.*

*Step 4: Check for Reliance on Appropriate Factors*

- Congressional Intent: Ensure the agency's decision is based on factors that Congress intended for it to consider.*
- Prohibited Factors: Verify that the agency did not rely on factors which Congress did not intend it to consider.*

*Step 5: Assess the Plausibility and Evidence Alignment*

- Plausibility Check: Evaluate if the agency's explanation is plausible and not contradicting the evidence.*
- Evidence Consistency: Examine if the decision aligns with or runs counter to the evidence before the agency.*

*Step 6: Determine Adequacy of Explanation*

- Reasoned Basis: Ensure that the agency has provided a reasoned basis for its action.*
- Deficiency and Clarity: Check if the decision is clear enough that its path can be reasonably discerned, even if not ideal.*

*Step 7: Consider Precedent and Legal Obligations*

- *Legal Consistency: Review if the decision aligns with legal standards and past precedents.*
- *Legislative and Regulatory Context: Understand the legislative and regulatory context within which the decision is made.*

*Step 8: Final Evaluation and Decision*

- *Decision Justification: Decide if the agency's action is justified based on the above assessments.*
- *Remedy Options: If found arbitrary and capricious, determine the appropriate remedy, such as remanding to the agency for further action or explanation.*

For a complicated doctrine, this is a good starting point for an analysis. The model is by no means comprehensive, since it draws on a relatively small body of case law. A true holistic model would likely require a larger volume of cases for ChatGPT to assess, in order to discern larger themes or patterns in the application of the standard.<sup>122</sup>

ChatGPT was next provided with a new fact pattern and asked to use its model to conduct an arbitrary and capricious analysis. The fact pattern was drawn from *Owner-Operator Independent Driver's Association v. Federal Motor Carrier Safety Administration*.<sup>123</sup> In that case, the Federal Motor Carrier Safety Administration's (FMCSA) revised Hours of Service rules that prohibited truck drivers from extending their maximum fourteen-hour driving-eligible work window through off-duty breaks.<sup>124</sup> Under prior rules, short breaks for naps, meals, or rest could pause the clock towards time limits, but the updated regulation mandated rig operators shut down after fourteen hours since starting their shift.<sup>125</sup> An industry group petitioned for judicial review, arguing the FMCSA acted arbitrarily and capriciously by ignoring adverse health and safety effects of making the fourteen-hour daily on-duty limit non-extendable.<sup>126</sup>

In examining the fact pattern for an arbitrary and capricious review, ChatGPT applied its model as follows:

---

122. Guinness, *supra* note 113.

123. *Owner-Operator Indep. Drivers Ass'n v. Fed. Motor Carrier Safety Admin.*, 494 F.3d 188 (D.C. Cir. 2007).

124. *Id.* at 207.

125. *Id.* at 208.

126. *Id.*

*Step 1: Establish the Scope of Review*

□ *Narrow Review Focus: The review will focus on the agency's decision-making process and its consideration of relevant factors, not on substituting the court's judgment for that of the FMCSA.*

*Step 2: Examine the Agency's Decision-Making Process*

□ *Data and Explanation Review: FMCSA based its decision on usage statistics showing most drivers do not exhaust the full 14-hour limit.*

□ *Rational Connection: There is a link between the decision and the agency's goal to prevent fatigued driving.*

*Step 3: Evaluate Consideration of Relevant Factors*

□ *Comprehensive Assessment: The agency considered the impact of nonextendable driving windows on driver behavior and fatigue.*

□ *Error in Judgment: There does not appear to be a clear error in judgment; the decision is based on usage data and safety considerations.*

*Step 4: Check for Reliance on Appropriate Factors*

□ *Congressional Intent and Prohibited Factors: The decision aligns with the general legislative intent of ensuring road safety.*

*Step 5: Assess the Plausibility and Evidence Alignment*

□ *Plausibility Check: The FMCSA's reasoning that a firm cutoff can prevent extremely fatigued driving is plausible.*

□ *Evidence Consistency: The decision aligns with the data indicating most drivers do not use the full 14-hour period.*

*Step 6: Determine Adequacy of Explanation*

□ *Reasoned Basis: FMCSA provided a reasoned basis focusing on preventing extremely fatigued driving.*

□ *Deficiency and Clarity: The explanation is clear, showing a path from evidence to decision.*

*Step 7: Consider Precedent and Legal Obligations*

□ *Legal Consistency: The decision appears consistent with FMCSA's mandate to ensure safety on roads.*

□ *Legislative and Regulatory Context: The decision aligns with the broader context of road safety regulations.*

*Step 8: Final Evaluation and Decision*

□ *Decision Justification: FMCSA's action seems justified based on safety data and the goal to prevent extremely fatigued driving.*

□ *Remedy Options: If deemed necessary, the court might remand for further clarification on the impact of the rule on drivers' break-taking behavior.*

ChatGPT reached the same outcome as the court in the underlying case, which had found that the agency had not acted arbitrarily and capriciously since it had in fact wrestled with the problem by relying on survey data showing that most drivers did not fully utilize the fourteen-hour driving window and that the non-extendable limit had benefits like keeping severely fatigued operators off the road after lengthy shifts.<sup>127</sup>

Both analyses assessed factors like evidentiary support and the coherency of the agency's explanatory reasoning. But even a surface reading of both shows major differences—ChatGPT's response takes a more mechanical, checklist-style approach, while the court's analysis in the underlying case digs more deeply into core logic gaps (such as examining alternative inferences from the statistics cited).

Additionally, ChatGPT's framing appears more deferential to the agency's plausible surface explanations because its analysis is structured around affirming certain criteria. ChatGPT took the FMCSA's claims regarding increased safety at face value without scrutinizing the agency's evidence or assumptions. By contrast, the court's actual analysis applied deeper scrutiny.<sup>128</sup> Where ChatGPT largely affirmed an issue-spotting of addressed factors, the court's analysis involved specifically drilling down into certain substantive claims rather than checking items off a list.

This suggests generative AI may be procedurally compliant in applying legal standards, but still struggle with cultivating healthy scrutiny or deeper legal analysis. However, on follow-up, ChatGPT did show a willingness to provide more information. When asked to explain why it thought the agency had reached a reasoned decision, it provided the following:

“FMCSA relied on usage statistics to inform its decision. These statistics indicated that most drivers do not use the full 14-hour

---

127. *Id.*

128. *Id.*

driving window available to them. This data suggests that a firm cutoff would not significantly impact the majority of drivers' ability to manage their driving and rest periods within the allotted time."

When asked about being skeptical of the agency, ChatGPT provided that its analysis was "more about ensuring that the FMCSA's decision was grounded in reason, evidence, and aligned with legal standards. It did not deeply delve into questioning the assumptions behind the agency's rationale[.]"

When asked to resolve the tension in conducting a deferential but searching review, ChatGPT provided the following:

*I aimed to balance deference to the agency's expertise and authority with a thorough and searching examination of its decision-making process... Here's how this balance was maintained:*

#### *Deference to the Agency*

**1. Respect for Agency Expertise:** *The analysis acknowledged the FMCSA's specialized knowledge and expertise in regulating commercial driving and road safety. This expertise grants the agency a certain latitude in making decisions that might not be immediately apparent to non-experts.*

**2. Acknowledgment of Agency Data:** *The decision was based on the FMCSA's usage statistics, indicating that most drivers do not use the full 14-hour driving window. This data was accepted as a legitimate basis for the agency's decision, respecting the agency's ability to collect and interpret relevant data.*

**3. Upholding Agency's Policy Choices:** *The agency's policy decision to implement a firm 14-hour cutoff was respected as a choice within the agency's purview to make. It was understood that agencies are often required to make policy decisions that involve complex trade-offs.*

#### *Searching Analysis*

**1. Examining Rational Connection:** *While being deferential, the analysis critically examined whether there was a rational connection between the FMCSA's data and its decision.*

*This step ensured that the decision was not just based on data but logically followed from it.*

**2. Assessing Consideration of Relevant Factors:** *The analysis involved scrutinizing whether the FMCSA had considered all relevant factors, particularly the impact of the rule change on driver behavior and safety.*

**3. Evaluating Evidence and Assumptions:** *The analysis critically looked at the assumptions underlying the FMCSA's rationale, such as whether a 14-hour nonextendable window would indeed lead to safer driving practices.*

**4. Questioning Potential Unintended Consequences:** *Although deferential to the agency's expertise, the analysis also considered the potential unintended consequences of the rule, such as drivers skipping breaks to maximize driving time within the 14-hour window.*

While the AI approach is still mechanical, it does reveal exactly what the AI relied on in making its arbitrary and capricious decision. Unlike a true black box, it can explain the exact aspects of its reasoning – a potential benefit of AI legal application that has already been identified in the literature.<sup>129</sup> AI is able to break down and explain its steps in a way human judges sometimes do not. This is, of course, in no way an endorsement of throwing out human legal application in favor of ChatGPT. In subjecting arbitrary and capricious review to AI treatment, it shows that AI has considerable promise for explaining elements of legal doctrines and turning one-dimensional legal standards into a more organized schema of individual rules. But its current capabilities are still tied to descriptive fluency.<sup>130</sup> And, it appears overly concerned with meeting items on a checklist, rather than examining alternative explanations or diving into the agency's explanations. As such, the benefit ChatGPT provides in this space is in turning a standard into discrete rules. As such, models generated by AI may be able to articulate legal tests intelligibly but may at this moment fail at applying the type of scrutiny that would reveal limitations within reasoning.

---

129. David Gewirtz, *How to Make ChatGPT Provide Sources and Citations*, ZDNET (Nov. 29, 2023, 8:13 AM), <https://www.zdnet.com/article/how-to-make-chatgpt-provide-sources-and-citations>.

130. Guinness, *supra* note 113.



So, for now, AI appears better equipped to serve as an initial legal standard clarifier versus a replacement for applying scrutiny. But the future may hold interesting developments. Should the AI be exposed to increased volumes of case law, it may be able to develop higher-order models and use previous precedent to better understand larger themes in the application of doctrine. Just as human wisdom comes partly from experiential reference, ChatGPT's legal prowess may follow suit if larger data sets can simulate enough actual application.