ORDINARY MEANING, EXTRAORDINARY TOOLS: DICTIONARIES, CORPORA, AND GENERATIVE AI

Stephen M. Johnson∗

TABLE OF CONTENTS

∗ Professor of Law, Mercer University Law School. B.S., J.D. Villanova University, LL.M. George Washington University Law School.

I. INTRODUCTION

"Might LLMs be useful in the interpretation of legal texts? . . .
[I]t's an issue worth exploring."[1]

In two cases the Eleventh Circuit decided in 2024, Judge Kevin Newsom wrote separately to discuss how he used ChatGPT and other AI tools to evaluate the ordinary meaning of legal terms and to suggest that generative AI[2] tools might be useful in the search for ordinary meaning.[3] The first case, *Snell v. United Specialty Insurance Co.*, involved the interpretation of terms in an insurance contract,[4] while the second case, *United States v. Deleon*, involved statutory interpretation.[5] In both cases, Judge Newsom agreed with the majority of the court that the disputes could be resolved without deciding the ordinary meaning of the disputed terms.[6] Nevertheless, in each case, he wrote separately to discuss his experimentation with ChatGPT and other LLMs to find ordinary meaning and to make a "modest proposal" that persons "who believe that 'ordinary meaning' is *the* foundational rule for the evaluation of legal texts should consider—*consider*—whether and how AI-powered large language models . . . might—*might*—inform the interpretive analysis."[7] Although Judge Newsom's modest proposal is clearly not a call to arms, it serves to begin a larger conversation about the role of AI in statutory interpretation.[8] The search for the ordinary meaning of statutory terms has long centered on intuition and dictionary definitions and more recently has expanded—in some circles—to the consideration of corpus linguistics.[9] As Judge Newsom suggests, it is worth at least exploring how ChatGPT and other AI tools compare to the

---

1. Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1225 (11th Cir. 2024) (Newsom, J., concurring).

2. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-23-106782, SCIENCE & TECH SPOTLIGHT: GENERATIVE AI 1 (2023), https://www.gao.gov/assets/gao-23-106782.pdf [https://perma.cc/ZZ29-TZWC]. "Generative artificial intelligence (AI) is a technology that can create content, including text, images, audio, or video, when prompted by a user. Generative AI systems create responses using algorithms that are trained often on open-source information, such as text and images from the internet." *Id.*

3. *Snell*, 102 F.4th at 1225 (Newsom, J., concurring); United States v. Deleon, 116 F.4th 1260, 1270 (11th Cir. 2024) (Newsom, J., concurring).

4. *Snell*, 102 F.4th at 1221 (Newsom, J., concurring). Specifically, the case centered on whether installation of an in-ground trampoline fit within the ordinary meaning of "landscaping" under a provision in an insurance contract. *Id.*

5. *Deleon*, 116 F.4th at 1270 (Newsom, J., concurring). The *Deleon* court focused on the ordinary meaning of "physically restrained" under the Federal Sentencing Guidelines. *Id.* at 1261 (majority opinion).

6. *See Snell*, 102 F.4th at 1222 (Newsom, J., concurring); *Deleon*, 116 F.4th at 1270 (Newsom, J., concurring).

7. *See Snell*, 102 F.4th at 1221 (Newsom, J., concurring).

8. *See id.*

9. *See infra* Part III (discussing tools used to identify ordinary meaning).

more traditional tools that courts use to determine the ordinary meaning of statutory terms.

Judge Newsom's experimentation was inspired, in part, by an article in the *New York University Law Review* that explored the use of generative AI in contract interpretation[10] and a working paper listed on the Max Planck Institute's (MPI) website, which focused on the use of ChatGPT to interpret statutory terms.[11] While both articles saw some promise in using generative AI to aid in finding the ordinary meaning of legal terms, the authors of both articles were temperate in their enthusiasm.[12] Yonathan Arbel and David Hoffman, the authors of the New York University article, concluded that "generative interpretation is good enough for many cases that currently employ more expensive, and arguably less certain methodologies" and described it as "a workable, workmanlike method."[13] Christoph Engel and Richard McAdams, the authors of the MPI paper, wrote, "we make [only] the modest claim that, with the right prompting techniques, . . . GPT . . . very cheaply provides useful data for the empirical assessment of the ordinary meaning of statutory terms."[14]

It is not surprising that researchers are advocating for the use of generative AI in statutory interpretation, as it is quickly transforming many other areas of law and legal practice. A low-cost and widely accessible tool, generative AI has already been used by lawyers and judges to identify and summarize judicial opinions, summarize discovery documents, draft deposition questions, briefs, and contracts, and predict settlement values, among other tasks.[15] Google CEO Sundar Pichai, perhaps overly

---

10. *See generally* Yonathan A. Arbel & David A. Hoffman, *Generative Interpretation*, 99 N.Y.U. L. REV. 451 (2024) (discussing the use and implication of AI tools in contract interpretation).

11. *See* Christoph Engel & Richard H. McAdams, *Asking GPT for the Ordinary Meaning of Statutory Terms* 1 (MPI Collective Goods Discussion Paper, No. 2024/5, 2024), https://ssrn.com/abstract=4718347 [https://perma.cc/3SJH-NUFZ]; *see also* Jonathan H. Choi, *Measuring Clarity in Legal Text*, 91 U. CHI. L. REV. 1, 1–2 (2024) (exploring the potential use of computational analysis in clarifying issues surrounding ordinary meaning).

12. *See* Arbel & Hoffman, *supra* note 10, at 458; Engel & McAdams, *supra* note 11, at 3.

13. *See* Arbel & Hoffman, *supra* note 10, at 458.

14. *See* Engel & McAdams, *supra* note 11, at 3.

15. *See* Arbel & Hoffman, *supra* note 10, at 498 (discussing low cost, accessibility, and use by lawyers and judges); Engel & McAdams, *supra* note 11, at 3; Paul W. Grimm, Maura R. Grossman & Cary Coglianese, *AI in the Courts: How Worried Should We Be?*, 107 JUDICATURE 65, 65 (2024); David Uriel Socol de la Osa & Nydia Remolina, *Artificial Intelligence at the Bench: Legal and Ethical Challenges of Informing—or Misinforming—Judicial Decision-Making Through Generative AI*, 6 DATA & POL'Y e59, e59-4 (2024); Matt Reynolds, *Words with Bots: How ChatGPT and Other AI Platforms Could Dramatically Reshape the Legal Industry*, 109 ABA J. 34, 36 (2023); Daniel Schwarcz & Jonathan H. Choi, *AI Tools for Lawyers: A Practical Guide*, 108 MINN. L. REV. HEADNOTES 1, 1–2 (2023); Stephen M. Johnson, *Rulemaking 3.0: Incorporating AI and ChatGPT into Notice and Comment Rulemaking*, 88 MO. L. REV. 1021, 1044–45 (2023). More dubious claims have also been made regarding the promise of AI in the legal field. *See, e.g.*, Natanson et al., *DOGE Builds AI Tool to Cut 50 Percent of Federal Regulations*, THE WASH. POST. (updated July 26, 2025), https://www.washingtonpost.com/business/2025/07/26/doge-ai-tool-cut-regulations-trump/ [https://perma.cc/VEZ7-9AVX] (asserting that AI can quickly identify regulations that the government should repeal).

optimistically, posit that the effects of generative AI could ultimately be "more profound than . . . electricity or fire."[16]

Supreme Court Chief Justice John Roberts also noted the great potential that AI offers to improve access to justice, but he stressed that "any use of AI requires caution and humility."[17] Commentators raise concerns about the reliability of, and potential for bias in, the information generative AI produces, and the possible misuse of the technology by lawyers and judges.[18] To address such concerns, the American Bar Association (ABA) issued a formal opinion clarifying the ethical obligations imposed on lawyers when using AI in practice.[19] The opinion stressed the need for lawyers to develop an understanding of the capabilities, limitations, benefits, and risks of AI and counseled lawyers to avoid "relying solely on . . . [AI] . . . to perform tasks that call for the exercise of professional judgment."[20] Although the ABA opinion does not separately address judges, many states have adopted ethics opinions that require judges to develop competence with AI and other technologies and to avoid using such technologies to draft opinions.[21] Similarly, guidance for judges in the United Kingdom discourages them from using generative AI for legal research or analysis.[22] A recent Cambridge

16. *See* Lauren Goode, *Google CEO Sandar Pichai Compares Impact of AI to Electricity or Fire*, THE VERGE (Jan. 19, 2018, at 15:50 CST) https://www.theverge.com/2018/1/19/16911354/google-ceo-sundar-pichai-ai-artificial-intelligence-fire-electricity-jobs-cancer [https://perma.cc/2N4A-PV4F]. *But see* Thomas R. Lee & Jesse Egbert, *Artificial Meaning?*, 77 FLA. L. REV. (forthcoming) (suggesting AI is not currently prepared to conduct imperical analysis of ordinary meanings) (on file with Texas Tech Law Review).

17. *See* JOHN G. ROBERTS, JR., 2023 YEAR-END REPORT ON THE FEDERAL JUDICIARY 5 (Dec. 31, 2023). In addition to noting the access to justice values of generative AI, the Chief Justice suggested that "[l]egal research may soon be unimaginable without it." *Id.*

18. *See, e.g.*, Grimm, Grossman & Coglianese, *supra* note 15, at 66–69 (discussing the biases present in AI tools).

19. *See* A.B.A. Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024).

20. *Id.* at 2–4. The opinion stressed that, because of the potential inaccuracy of information AI generates, "a lawyer's reliance on, or submission of, a GAI tool's output—without an appropriate degree of independent verification or review of its output—could violate the duty to provide competent representation as require by Model Rule 1.1." *Id.* at 3–4.

21. *See* GARY E. MARCHANT, ARTIFICIAL INTELLIGENCE, JUDGES AND LEGAL ETHICS 11 (Nat'l Civ. Just. Initiative July 20, 2024) (discussing ethics opinions from Michigan, New Jersey, West Virginia, and Utah and recommending that "AI should not be used to reach decisions or draft final opinions or orders at this time"); Jud. Investigation Comm'n (West Va.), JIC Advisory Opinion 2023–22 (Oct. 13, 2023) (prohibiting the use of AI to decide the outcome of a case and requiring "extreme caution" when using AI in drafting opinions or orders); *Interim Rules on the Use of Generative AI*, UTAH JUD. COUNCIL (Oct. 25, 2023) (limiting the use of generative AI to listed purposes, which do not include drafting opinions); *Statement of Principles for the New Jersey Judiciary's Ongoing Use of Artificial Intelligence, Including Generative Artificial Intelligence*, N.J. CTS., at 2 (Jan. 23, 2024), https://www.njcourts.gov/sites/default/files/courts/supreme/statement-ai.pdf [https://perma.cc/M4TF-2DZ7] (noting that judges "may use AI only for select purposes, such as for preliminary gathering and organization of information").

22. *See* Cts. and Tribunals Judiciary, *Artificial Intelligence (AI) Guidance for Judicial Office Holders* 5–6 (Dec. 12, 2023) (warning that the output of generative AI may be inaccurate, incomplete, misleading, or biased). Canada imposes greater limits on the judicial use of AI, as the Canadian Federal Court, in December 2023, banned the use of AI in making judgments and orders until a public consultation on the matter occurred. *See* Socol de la Osa & Remolina, *supra* note 15, at e59-12. In contrast to the

University research paper outlined the negative impacts on judicial independence and the rule of law that can arise in countries where there are few ethical or regulatory limits on the judicial use of AI.[23]

Bearing in mind those forecasts and warnings, and inspired by Judge Newsom's "modest proposal," this Article explores the role, if any, that generative AI might play in the determination of the "ordinary meaning" of terms in statutory analysis. Part II will outline the central role that interpreting ordinary meaning plays in statutory analysis, the reasons why judges give ordinary meaning that central role, and the rules that judges generally apply to determine the ordinary meaning of statutory terms.[24] Part III of the Article introduces the various tools that judges frequently use, or are exploring for use, in interpreting ordinary meaning, including intuition, dictionaries, corpus linguistics,[25] surveys and, now, generative AI.[26] Part IV compares dictionaries, corpus linguistics, and generative AI, focusing on how well they (1) provide *consistent answers* regarding the ordinary meaning of words; (2) *accurately reflect* the ordinary meaning of words; (3) identify the ordinary meaning of words in *context*; (4) identify the ordinary meaning of words in a way that *does not exclude the meaning understood by significant groups* of the population; (5) identify the ordinary meaning of words at a specific point in time; (6) identify the ordinary meaning of words in a *transparent* manner; and (7) provide an *easy, inexpensive and accessible* tool to identify ordinary meaning.[27] Along many of those metrics, generative AI is at least as effective, or more effective, than other tools.[28] However, along

---

approaches taken in the United Kingdom and Canada, judges in Mexico, Brazil, Peru, India, Dubai, and China have used AI more frequently to draft opinions. *See* MARCHANT, *supra* note 21, at 10–11. In at least one reported case in Brazil, though, an opinion drafted with AI included a fabricated case citation. *See* Agence France Presse, *Brazil Judge Investigated for AI Errors in Ruling*, BARRON'S (Nov. 13, 2023, at 17:03 EST), https://www.barrons.com/news/brazil-judge-investigated-for-ai-errors-in-ruling-c45e8f8f [https://perma.cc/2YVX-XA4T].

23. *See* Socol de la Osa & Remolina, *supra* note 15, at e59-13–17. The authors noted the potential for the unregulated use of AI to perpetuate biases or generate flawed outputs based on inaccurate information in cases such as a decision from a judge in Mexico who used ChatGPT to decide whether a child with autism should be eligible for medical treatment. *Id.* at e59-13–14. The authors also warned that "[j]udges who rely on GenAI-generated content without critical analysis may inadvertently allow underlying biases, historical data patterns, or specific ideological perspectives embedded within the algorithms or training data to shape their decisions in ways that could undermine their independent judgment." *Id.* at e59-9.

24. *See infra* Part II (discussing the role of what ordinary means in society in regard to AI).

25. Graeme Kennedy, *Corpus Linguistics*, *in* INTER'L ENCYC. OF THE SOC. & BEHAV. SCIS. 2816 (Neil J. Smelser & Paul B. Baltes eds., 2001) ("Corpus linguistics encompasses the compilation and analysis of collections of spoken and written texts as the source of evidence for describing the nature, structure, and use of languages.").

26. *See infra* Section III.C (discussing increased use of surveys); *infra* Section III.E (analyzing Generative AI as a tool for determining ordinary meaning).

27. *See infra* Part IV (discussing how accurate AI is by comparing dictionaries, corpus linguistics, and generative AI).

28. *See infra* Section IV.A (discussing the effectiveness of AI).

a few metrics, it is not quite ready for prime-time.[29] After comparing the tools, Part V of this Article concludes with a proposal for the limited use of generative AI in interpreting the ordinary meaning of terms in statutes.[30] While the short-term prognosis is not positive, the technology is evolving rapidly, and it could eventually become a valuable tool for statutory interpretation.[31]

II. ORDINARY MEANING

When interpreting statutory language, courts adopt the "ordinary meaning" of terms unless context requires a different result.[32] The ordinary meaning canon has been called "the most fundamental semantic rule of interpretation,"[33] and judicial reliance on the canon has grown exponentially with the rise of textualism as a theory of interpretation.[34] Although the ordinary meaning canon may be associated most closely with textualism, judges adopting most other theories of interpretation routinely begin their

29. *See infra* Section IV.A (discussing the limitations of AI).

30. *See infra* Part V (proposing the limited use of generative AI to explain the plain meaning of words in a statute).

31. *See infra* Section V.C (arguing that generative AI may become a helpful tool when interpreting statutes).

32. *See, e.g.*, Bostock v. Clayton Cnty., 1590 U.S. 644, 654 (2020) ("This Court normally interprets a statute in accord with the ordinary public meaning of its terms."); Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 566 (2012) (citing Asgrow Seed Co. v. Winterboer, 513 U.S. 179, 187 (1995) ("When a term goes undefined in a statute, we give the term its ordinary meaning."); Gonzales v. Carhart, 550 U.S. 124, 152 (2007) ("In interpreting statutory texts courts use the ordinary meaning of terms unless context requires a different result."). The ordinary meaning rule is used to implement the "plain meaning rule," which counsels courts to read statutes according to their plain meaning. *See* Conn. Nat'l Bank v. Germain, 503 U.S. 249, 253 (1992) (identifying the plain meaning rule as the "cardinal canon" of interpretation that comes "before all others"). While the ordinary meaning rule is the primary tool that courts use to interpret the plain meaning of statutory terms, courts depart from the ordinary meaning rule when terms have a technical meaning in a specialized trade or discipline and are used in a statute addressing that trade or discipline. *See* ANTONIN SCALIA & BRYAN A. GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 72 (2012).

33. *See* SCALIA & GARNER, *supra* note 32, at 69.

34. *See* Marco Basile, *Ordinary Meaning and Plain Meaning*, 110 VA. L. REV. 135, 137 (2024); Kevin Tobia, Brian G. Slocum, & Victoria Nourse, *Statutory Interpretation From the Outside*, 122 COLUM. L. REV. 213, 217 (2022) (finding that citation to ordinary meaning across a study of six million cases tripled in the most recent fifty years addressed in the study); Anita S. Krishnakumar, *Cracking the Whole Code Rule*, 96 N.Y.U. L. REV. 76, 97 (2021) (finding that between 2005 and 2017, the Supreme Court relied on "text" and "plain meaning" in 50% of the majority opinions); Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915, 1973–74 (finding an exponential increase in the use of "plain meaning" and "ordinary meaning" in Supreme Court opinions since the 1970s). Textualists always strive to give statutory terms their ordinary meaning, *see* Kevin Tobia, Jesse Egbert & Thomas R. Lee, *Triangulating Ordinary Meaning*, 112 GEO. L. J. ONLINE 23, 24 (2023), and Justice Elena Kagan famously observed that "we're all textualists now." *See* Justice Elena Kagan, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes*, at 8:28 (Youtube, Nov. 17, 2015), https://www.youtube.com/watch?v=dpEtszFT0Tg [https://perma.cc/GJ67-QJNQ].

analysis of statutory language with the text, and rely heavily on the ordinary meaning canon.[35]

Advocates of the ordinary meaning rule identify multiple rationales for application of the rule.[36] For some supporters, the ordinary meaning represents the best evidence of the intent of a statute's drafters.[37] Many adherents contend that application of the rule provides clear notice to the public about the meaning of statutes and protects reliance interests.[38] Supporters also argue that the rule protects separation of powers, by limiting the power of judges to legislate,[39] and promotes confidence in the rule of law.[40] Finally, advocates suggest that the rule facilitates democratic accountability.[41]

While the ordinary meaning rule is almost universally applied, and there is consensus around many of the principles associated with the rule, there are also areas of disagreement regarding its application.[42] As it is applied most frequently, the ordinary meaning rule requires judges to interpret statutory text in the manner that an ordinary "reasonable" member of the public would interpret it.[43] However, a word or phrase may have a range of ordinary meanings.[44] It may have a literal meaning that is quite broad, a prototypical meaning that is much narrower, and a range of other meanings that are

---

35. *See* Tobia, Slocum & Nourse, *supra* note 34 (suggesting most theories share a commitment to ordinary meaning); Kevin Tobia, *Dueling Dictionaries and Clashing Corpora*, 71 DUKE L.J. ONLINE 146, 147 (2022) ("Even non-textualist Justices have begun to appeal to the 'ordinary speaker.'") [hereinafter *Dueling Dictionaries*]; Kevin P. Tobia, *Testing Ordinary Meaning*, 134 HARV. L. REV. 726, 731 (2020) [hereinafter *Testing Ordinary Meaning*]; Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L. J. 788, 792 (2018) [hereinafter *Judging Ordinary Meaning*]; *see also* Engel & McAdams, *supra* note 11, at 4.

36. *See Judging Ordinary Meaning*, *supra* note 35; *Testing Ordinary Meaning*, *supra* note 35; Tobia, Slocum & Nourse, *supra* note 34.

37. *See Judging Ordinary Meaning, supra* note 35. *But see Testing Ordinary Meaning, supra* note 35, at 738 (stating ordinary meaning considers the understanding of most ordinary people, not specific intentions of the document's author).

38. *See Dueling Dictionaries*, *supra* note 35; *Testing Ordinary Meaning*, *supra* note 35, at 737; *Judging Ordinary Meaning*, *supra* note 35; *see also* Bostock v. Clayton Cnty., 590 U.S. 644, 785 (2020) (Kavanaugh J., dissenting).

39. *See Testing Ordinary Meaning*, *supra* note 35, at 737; *see also Bostock*, 590 U.S. at 785 (arguing that ordinary meaning facilitates domestic accountability).

40. *See Dueling Dictionaries*, *supra* note 35; *Judging Ordinary Meaning*, *supra* note 35; *see also Bostock*, 590 U.S. at 785 (stating that laws should be understandable to citizens).

41. *See Dueling Dictionaries*, *supra* note 35; *Testing Ordinary Meaning*, *supra* note 35, at 737; *Bostock*, 590 U.S. at 785.

42. *See Dueling Dictionaries*, *supra* note 35; *Bostock*, 590 U.S. at 785.

43. *See* Tobia, Egbert & Lee, *supra* note 34, at 24; Anita Krishnakumar, *Metarules for Ordinary Meaning*, 134 HARV. L. REV. 167 (2021) [hereinafter *Metarules*]; *Dueling Dictionaries*, *supra* note 35, at 150; *Testing Ordinary Meaning*, *supra* note 35, at 736. *But see* Stephen C. Mouritsen, *Natural Language and Legal Interpretation*, 86 BROOK. L. REV. 533, 535 (2021) (noting that "Justice Scalia variously characterized 'ordinary meaning' as (1) 'what an ordinary speaker of the English language would think it means,' (2) what an 'ordinary Member of Congress' would read" it to mean, or (3) "'the intent that a reasonable person would gather from the text . . . placed alongside the remainder of the *corpus juris*'"). [hereinafter *Natural Language*].

44. *See Judging Ordinary Meaning*, *supra* note 35.

common or possible.[45] While judges tend to avoid adopting literal meanings of words or phrases when that seems inappropriate based on the context of the statute,[46] there is not a bright line rule that outlines where, along the continuum of meanings, judges should stop to identify the ordinary meaning of words or phrases.[47] In addition, there may be times when the ordinary meaning of terms, as *understood* by the public, is different than the ordinary meaning of terms as they are *used*.[48]

Although words or phrases may have a range of ordinary meanings, words and phrases must be read in context under the ordinary meaning rule, and context will eliminate many of the otherwise potential ordinary meanings.[49]

Finding the ordinary meaning of words and phrases is also complicated by the fact that the meaning of language can change over time.[50] Some theorists argue that the ordinary meaning of words or phrases should be interpreted consistent with the meaning of the language at the time the statute being interpreted was enacted.[51] Others argue that if a purpose of the ordinary meaning rule is to provide notice to the public about rights and obligations under laws, the ordinary meaning of words or phrases should be interpreted consistent with contemporary understandings of the meaning of the language being interpreted.[52]

---

45. *Id.* Professors Thomas Lee and Stephen Mouritsen conceptualize a continuum that includes possible, common, most frequent, exclusive and, perhaps, prototype. *Id.* at 800–02.

46. *See* Choi, *supra* note 11, at 14; *see also Bostock*, 590 U.S. at 784 (Kavanaugh, J., dissenting) ("[C]ourts must follow ordinary meaning, not literal meaning.").

47. *See Metarules*, *supra* note 43, at 167; *Testing Ordinary Meaning*, *supra* note 35, at 736; *Judging Ordinary Meaning*, *supra* note 35, at 798.

48. *See* Engel & McAdams, *supra* note 11, at 6.

49. *See* SCALIA & GARNER, *supra* note 32, at 70; James J. Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court's Thirst for Dictionaries in the Rehnquist and Roberts Eras*, 55 WM. & MARY L. REV. 483, 494, 515 (2013); *see also Bostock*, 590 U.S. at 784 (Kavanaugh, J., dissenting) ("[P]roper statutory interpretation asks 'how a reasonable person . . . would read the text in context.'"); MCI Telecomms. Corp. v. AT&T, 512 U.S. 218, 240 (1994) (Stevens, J., dissenting) ("Dictionaries . . . are no substitute for close analysis of what words mean as used in a particular statutory context."). At times, judges may also conclude, based on statutory context, that words or phrases in a statute should be interpreted according to a technical meaning, rather than their ordinary meaning. *See* SCALIA & GARNER, *supra* note 32, at 74.

50. *See* SCALIA & GARNER, *supra* note 32, at 78; *Testing Ordinary Meaning*, *supra* note 35, at 741.

51. *See* SCALIA & GARNER, *supra* note 32, at 78 (describing originalism); *Testing Ordinary Meaning*, *supra* note 35, at 738; Brudney & Baum, *supra* note 49, at 511. Originalists strive to implement the meaning intended by the enacting legislature and argue that adopting any other interpretation of the law is an unconstitutional delegation of lawmaking power to the judiciary. *See* SCALIA & GARNER, *supra* note 32, at 78–84.

52. *See* Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L. REV. 275, 332–33 (1998); Brudney & Baum, *supra* note 49, at 511–12; Amy Coney Barrett, *Congressional Insiders and Outsiders*, 84 U. CHI. L. REV. 2193, 2195 (2017) ("[Textualists] view themselves as agents of the people rather than of Congress and as faithful to the law rather than to the lawgiver."). Contemporary meaning also appeals to judges who believe that statutes are dynamic, so that the meaning of words can evolve with societal changes. *See* Brudney & Baum, *supra* note 49, at 511–12.

Discerning ordinary meaning is further obscured because there is disagreement among judges and academics regarding the appropriate audience to consider when interpreting statutes.[53] As Professor Anita Krishnakumar noted, the audience for statutory provisions that address procedural or legal matters may be judges, while the audience for many other statutes may be the "average person on the street."[54] Even if the audience is the "average person on the street," questions arise regarding the extent to which the analysis of ordinary meaning needs to explore the meaning of language as understood by underrepresented populations, as opposed to majoritarian interpretations.[55]

While there is a lack of consensus on many of the issues addressed in the preceding paragraphs, there is increasing sentiment that the search for ordinary meaning should be an empirical endeavor, rather than a normative one.[56] Proponents argue that basing ordinary meaning interpretations on empirical data advances reliance and fair notice interests and reduces opportunities for judges to usurp the legislative function.[57]

Regardless of whether the ordinary meaning of language is assessed empirically, supporters of the ordinary meaning rule defend it on the grounds that it yields clear, determinate answers to statutory interpretation questions.[58] However, even a cursory review of judicial opinions applying the ordinary meaning rule will demonstrate the variability with which it can be applied in practice, leading to criticisms that the rule does little to prevent decisions based on political, ideological or personal bases.[59] The search for an empirically supported ordinary meaning may often be illusory because the ordinary meaning of language is frequently indeterminate.[60] As Professor

---

53. *See Metarules*, *supra* note 43, at 168, 171.

54. *Id.* at 171.

55. *Id.* (noting that terms may have different meanings to men and women and to people of different racial or ethnic backgrounds); Arbel & Hoffman, *supra* note 10, at 450–51; *see also* Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1227, 1232 (11th Cir. 2024) (discussing the limitations of generative AI to identify ordinary meaning in a way that accounts for the usage of language by underrepresented populations).

56. *See* Tobia, Egbert, & Lee, *supra* note 34; *Dueling Dictionaries*, *supra* note 35, at 150; *Testing Ordinary Meaning*, *supra* note 35; *Judging Ordinary Meaning*, *supra* note 35, at 794–95; Lee & Egbert, *supra* note 16, at 4.

57. *See Judging Ordinary Meaning*, *supra* note 35, at 795.

58. *See* Tobia, Egbert, & Lee, *supra* note 34.

59. *See* Anita S. Krishnakumar, *Dueling Canons*, 65 DUKE L. J. 909, 912 (2016) (finding, in a study of United States Supreme Court opinions, that in 42.7% of the cases, at least one opinion argued that a statute had an ordinary meaning while another opinion argued that the statute had a different ordinary meaning); *Judging Ordinary Meaning*, *supra* note 35, at 793 (internal quotes omitted) (quoting WILLIAM N. ESRIDGE, JR., INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION 25 (2016)) ("[O]rdinary meaning does not always yield predictable answers to statutory issues . . . ."); *Testing Ordinary Meaning*, *supra* note 35, at 730 (noting that ordinary meaning analysis is sometimes associated with conservative legal thought).

60. *See* Choi, *supra* note 11, at 9–11; *Testing Ordinary Meaning*, *supra* note 35, at 767–68; *Judging Ordinary Meaning*, *supra* note 35, at 798 ("Commentators are undoubtedly right to question the determinacy of the inquiry into ordinary meaning.").

Jonathan Choi explains, language may be indeterminate because it is ambiguous, having multiple meanings, or because it is vague, in that the limits of the words or phrases are imprecisely defined.[61] In those circumstances, Choi argues, no amount of information can resolve the indeterminacy of the language, and there is no ordinary meaning to be found in the text.[62] Most lawsuits center on disputes over language in the "zone of indeterminacy" (Choi's term), where a judge would find language unclear, even given perfect information.[63]

III. TOOLS TO IDENTIFY ORDINARY MEANING

Judges rely on a variety of tools to identify the ordinary meaning of words and phrases in statutes and frequently support their conclusions with citations to multiple different sources.[64]

A. Intuition

Traditionally, judges often began their analysis of the ordinary meaning of terms by relying on their intuition.[65] In the words of Justice Scalia, "[T]he acid test of whether a word can reasonably bear a particular meaning is whether you could use the word in that sense at a cocktail party without having people look at you funny."[66] However, the practice of finding ordinary meaning through intuition has been challenged on many grounds.[67] For instance, critics argue that such "gut level"[68] decision-making is subjective[69] and varies from interpreter to interpreter.[70] It is not an empirical method of

---

61. *See* Choi*, supra* note 11, at 11. As Justice Brett Kavanaugh has noted, the test for ambiguity is also ill-defined, as some judges appeal to a 65-35 rule, while others apply a 90-10 rule. *See* Brett Kavanaugh, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118, 2137 (2016) (reviewing Robert A. Katzmann, Judging Statutes (2014)). Without clearer guideposts, judgments about ambiguity are easily biased by strong policy preferences of the interpreters. *See* Choi, *supra* note 11, at 9.

62. *See* Choi, *supra* note 11, at 9–10.

63. *Id.* at 6, 11–13. Choi posits, though, that the zone of indeterminacy is, itself, variable in that textualists may adopt a narrower zone of indeterminacy than purposivists, reflecting the textualist desire to resolve interpretive questions based on the text of the statute, rather than extrinsic sources. *Id.* at 13.

64. *See* Tobia, Egbert & Lee, *supra* note 34, at 27; *Dueling Dictionaries*, *supra* note 36, at 147.

65. *See* Tobia, Egbert & Lee, *supra* note 34, at 30; *Testing Ordinary Meaning*, *supra* note 35, at 743; Thomas R. Lee & Stephen C. Mouritsen, *The Corpus and the Critics*, 88 U. CHI. L. REV. 275, 285 (2021) [hereinafter *The Corpus and the Critics*].

66. *See* Johnson v. United States, 529 U.S. 694, 718 (2000) (Scalia, J., dissenting).

67. *Choi*, *supra* note 11, at 14.

68. *See Judging Ordinary Meaning*, *supra* note 35, at 806; *The Corpus and the Critics*, *supra* note 65.

69. *See* Tobia, Egbert & Lee, *supra* note 34, at 30 (intuitions are malleable and context-dependent); *see also* Dubin v. United States, 599 U.S. 110, 136 (2023) (Gorsuch, J., concurring) ("[Y]ou could spend a whole day cooking up scenarios . . . that collapse even your most basic intuitions about what [the statute] does and does not criminalize.").

70. *See* Tobia, Egbert & Lee, *supra* note 34, at 49.

interpretation,[71] cannot be reviewed for accuracy, and there are few rules to guide a judge's application of intuition.[72] Judges may believe that their intuition aligns with the intuition of the reasonable member of the public,[73] but critics argue that judges are unrepresentative of the general population[74] and may bring biases and prejudices regarding the outcome of cases to their exercise of intuition in finding the ordinary meaning of words and phrases.[75] In light of the concerns outlined above, today, few judges rely solely on intuition when interpreting statutory terms.76

B. Dictionaries

Coinciding with the rise in textualism, judges frequently support any reading of the ordinary meaning of statutory terms with dictionary definitions.[77] The Supreme Court, in particular, has substantially increased its use of dictionaries to support statutory interpretation over the past several decades, while substantially reducing the use of legislative history as an interpretive tool.[78] The shift has occurred with very little public debate or

---

71. *See id.* at 27; *Judging Ordinary Meaning*, *supra* note 35, at 806; *Testing Ordinary Meaning*, *supra* note 35, at 744 ("[I]ntuition is recognized . . . as a fallible source of evidence in modern interpretation . . . .").

72. *See Testing Ordinary Meaning*, *supra* note 35, at 744.

73. *See* Tobia, Egbert & Lee, *supra* note 34, at 31 ("[E]mpiricists have found that people are subject to 'false consensus bias' in legal interpretation, overestimating how often others agree with their personal interpretation."); *Testing Ordinary Meaning*, *supra* note 35, at 743–44.

74. *See The Corpus and the Critics*, *supra* note 65, at 285–86 (noting that judges are generally wealthier and better educated than the average American). Professors Lee and Mouritsen also argue that judges may not be in a good position to intuit the meaning of terms in prior eras due to "linguistic drift—the notion that language usage and meaning shifts over time." *Id.* at 286 (internal quotes omitted) (quoting Thomas R. Lee & James C. Phillips, *Data-Driven Originalism*, 167 U. PA. L. REV. 261, 265 (2019)).

75. *See id.* at 286.

76. *See, e.g.*, Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1224 (11th Cir. 2024) (Newsom, J., concurring) ("[G]ut-instinct decision making has always given me the willies—I definitely didn't want to be that guy.").

77. *See Metarules*, *supra* note 43, at 167; Brudney & Baum, *supra* note 49, at 494–502; Jeffrey L. Kirchmeier & Samuel A. Thumma, *Scaling the Lexicon Fortress: The United States Supreme Courts Use of Dictionaries in the Twenty-First Century*, 94 MARQ. L. REV. 78, 84–86 (2010); *see also* Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 566–68 (relying on fourteen dictionaries to discern the ordinary meaning of "interpreter").

78. *See* Brudney & Baum, *supra* note 49, at 486–91, 497 (noting a rise in the use of dictionaries in Supreme Court opinions from 3.3% of decisions during the last five years of the Burger Court to 33.7% of decisions during the 2008–2010 terms of the Roberts Court); Abbe R. Gluck & Lisa Schultz Bressman, *Statutory Interpretation from the Inside—An Empirical Study of Congressional Drafting, Delegation, and the Canons: Part I*, 65 STAN. L. REV. 901, 938 (2013) (noting that the Court used dictionaries in 225 opinions from 2000 to 2010, as opposed to sixteen opinions in the 1960s); *Testing Ordinary Meaning*, *supra* note 35, at 732. Although the Court has significantly increased its reliance on dictionaries over the past several decades, Professor Ellen Aprill notes that dictionaries were cited in the Court's opinions as early as 1785. *See* Aprill, *supra* note 52, at 313. Judges sometimes also examine a word's etymology to decipher its meaning, a practice some critics assail as indefensible. *See, e.g.*, *Judging Ordinary Meaning*, *supra* note 35, at 807; *The Corpus and the Critics*, *supra* note 65, at 287–88 (labeling the practice "incoherent").

discussion among the Justices regarding the benefits or weaknesses of relying on dictionaries to determine ordinary meaning.[79] Courts across the country frequently support their reading of statutory terms with citations to dictionaries as an authoritative source, even though surveys of legislators suggest that they do not generally rely on dictionaries when drafting statutes.[80] Although courts are beginning to rely on a few other tools to discern ordinary meaning of statutory terms, dictionaries are the primary tool used today.[81]

Dictionaries are often defended as a neutral and objective tool for identifying ordinary meaning[82] and are used not only by textualists, but by judges applying every theory of statutory interpretation,[83] and by judges across the ideological spectrum.[84] In some cases, judges cite dictionaries as authoritative support for the ordinary meaning of statutory language.[85] In other cases, judges cite dictionaries to support ordinary meaning readings that are justified on other bases[86] or to reach a conclusion that there is no ordinary meaning of terms so other sources of interpretation must be examined.[87]

Although dictionary supporters maintain that they are an objective tool and that reliance on dictionaries reduces opportunities for judicial activism,[88] critics argue that the manner in which judges use dictionaries to support ordinary meaning determinations is highly subjective and provides significant opportunities for judges to interpret statutes according to ideological or personal preferences.[89] There are generally no bright line rules that detail (1) whether judges should rely on dictionaries published at the time of enactment of a statute, as opposed to modern dictionaries;[90] (2) which dictionaries judges should rely on when there are multiple dictionaries available for a particular time period;[91] (3) whether the order of presentation in a dictionary is meaningful when there are multiple definitions included in

79. *See* Brudney & Baum, *supra* note 49, at 487; Gluck & Bressman, *supra* note 78, at 955.

80. *See* Gluck & Bressman, *supra* note 78 (finding, in a survey of legislative drafters, that more than fifty percent indicated they never or rarely used dictionaries when drafting statutes); Aprill, *supra* note 52, at 299.

81. *See* Aprill, *supra* note 52, at 280–81. The weight accorded to dictionaries, though, may vary depending on whether judges are interpreting statutes through a textualist, purposivist, or intentionalist theory. *Id.*

82. *See* Gluck & Bressman, *supra* note 78, at 955; Brudney & Baum, *supra* note 49, at 486–87; *The Corpus and the Critics*, *supra* note 65, at 286.

83. *See* Aprill, *supra* note 52, at 280–81.

84. *See* Brudney & Baum, *supra* note 49, at 489, 491, 525.

85. *Id.* at 492–93 (describing "barrier opinions"); Aprill, *supra* note 52, at 486–87.

86. *See* Brudney & Baum, *supra* note 49, at 492–93 (describing "ornamental role opinions").

87. *Id.* (describing "way station opinions").

88. *See id.* at 491, 500.

89. *See Dueling Dictionaries*, *supra* note 35, at 147, 150–51; Brudney & Baum, *supra* note 49, at 490–91; Aprill, *supra* note 52, at 281, 300 ("[C]iting to dictionaries creates a sort of optical illusion, conveying the existence of certainty . . . when appearance may be all there is."); *The Corpus and the Critics*, *supra* note 65, at 286–88.

90. *See* Brudney & Baum, *supra* note 49, at 491, 493.

91. *Id.* at 493.

a dictionary for a term;[92] or (4) when to consult legal or technical dictionaries as opposed to general dictionaries.[93] Consequently, judges have a great degree of discretion to choose dictionary definitions that support a preferred reading of statutory language.[94] Empirical studies support the charge that the choice of dictionaries in practice is subjective[95] because judges do not consistently use the same dictionaries across decisions,[96] ignore definitions in dictionaries that seem applicable,[97] and rely on dictionaries that litigants do not propose.[98] Significantly, judges rarely explain the rationale behind their decision to rely on one dictionary definition or set of definitions as opposed to others.[99] It is difficult to justify the choice of a dictionary definition as objective without explaining the rationale behind the choice.[100]

C. Surveys

Although courts have not yet widely used surveys, some judges and academics advocate for the increased use of surveys as a tool to determine the ordinary meaning of statutory language.[101] As Chief Justice Roberts queried at a recent Supreme Court oral argument: "[T]he most probably useful way of settling all these questions would be to take a poll of 100 ordinary . . . speakers of English and ask them what [the statute] means, right?"[102] Proponents assert that surveys are an objective tool that reduces

---

92. *See The Corpus and the Critics*, *supra* note 65, at 287–88 ("[T]he 'sense-ranking fallacy'—the notion that a given sense of a term is ordinary . . . because it is listed higher in the dictionary's list of senses."); *Judging Ordinary Meaning*, *supra* note 35, at 807. Definitions in dictionaries might be listed in (1) historical order; (2) frequency of use order; or (3) order to advance structural coherence. *See* Aprill, *supra* note 52, at 294 n.100.

93. *See* Aprill, *supra* note 52, at 310; Brudney & Baum, *supra* note 49, at 538. Professor Aprill notes that the Supreme Court often uses general and legal dictionaries interchangeably without any particular pattern for their choice and sometimes rely on legal dictionaries to define common non-legal terms. *See* Aprill, *supra* note 52, at 300, 310–11. Aprill suggests that legal dictionaries are a particularly poor choice as a source of definitions for non-legal terms because the definitions for terms in legal dictionaries are generally drawn from judicial opinions, frequently from lower federal courts or state courts. *Id.* at 312. At the same time, she points out, general dictionaries also provide incomplete definitions of many words because legal sources are not generally reviewed as potential sources for definitions in general dictionaries. *Id.* at 303.

94. *See* Aprill, *supra* note 52, at 300–01.

95. *See Dueling Dictionaries*, *supra* note 35, at 150; Brudney & Baum, *supra* note 49, at 489, 536–38.

96. *See* Brudney & Baum, *supra* note 49, at 489–91.

97. *See* Aprill, *supra* note 52, at 300 (noting that opinions frequently only cite one dictionary definition and do not explain the reason for the choice of that definition).

98. *See* Brudney & Baum, *supra* note 49, at 489.

99. *See* Aprill, *supra* note 52, at 300.

100. *Id.*

101. *See Metarules*, *supra* note 43, at 181; *Natural Language*, *supra* note 43, at 533–34; James A. Macleod, *Ordinary Causation: A Study in Experimental Statutory Interpretation*, 94 IND. L.J. 957, 962–64 (2019).

102. *See* Transcript of Oral Argument at 52, Facebook, Inc. v. Duguid, 141 S. Ct. 1163 (2021) (No. 19-511).

opportunities for judicial manipulation of the ordinary meaning analysis.[103] Surveys have been used for centuries in other contexts to determine linguistic perceptions, and supporters argue that they can provide evidence of how language is used or perceived that is not available through intuition or dictionaries.[104] Advocates also argue that surveys can (1) be structured to address the precise statutory interpretation question in a case;[105] (2) reveal the meaning of language in context;[106] and (3) be administered to members of a wide variety of backgrounds, reducing the likelihood that an ordinary meaning interpretation ignores the meaning that underrepresented groups use or understand.[107]

Despite those advantages, surveys do not already exist to address many of the statutory interpretation questions that courts must to resolve, so judges must develop and conduct surveys in many cases if they plan to rely on survey data to interpret the ordinary meaning of statutory terms.[108] Judges do not have the time, resources, or expertise to poll ordinary citizens on a widespread basis because surveying is expensive and time-consuming.[109] Critics also argue that surveys involve artificially constructed questions, so the responses do not necessarily reflect how people actually use or understand language.[110] In addition, survey participants' responses may vary greatly depending on how questions are phrased, and respondents may misunderstand survey questions.[111] Further, critics assert that surveys frequently do not provide accurate information due to non-response bias among parts of the population and difficulties of generalization.[112] Finally, while surveys might be constructed to examine the meaning of statutory language as understood or used by the public at the time of a judicial challenge, it is impossible to use surveys to reconstruct the meaning of the language that existed at the time a statute was enacted.[113] In light of all of those challenges, surveys remain more of a theoretical tool, rather than a practical one, for determining ordinary meaning today.

---

103. *See Natural Language*, *supra* note 43, at 548.

104. *Id.* at 537, 549, 553.

105. *Id.* at 534, 548.

106. *See* Tobia, Egbert & Lee, *supra* note 34, at 53; *Natural Language*, *supra* note 43, at 548.

107. *See* Mouritsen, *supra* note 43, at 548, 553.

108. *Id.* at 548–49; *see* Arbel & Hoffman, *supra* note 10, at 473.

109. *See* Arbel & Hoffman, *supra* note 10, at 473; Engel & McAdams, *supra* note 11, at 6; Choi, *supra* note 11, at 52; Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1230 (11th Cir. 2024). *But see Natural Language*, *supra* note 43, at 548 (noting that recent advances in survey technology are reducing costs associated with surveys).

110. *See Natural Language*, *supra* note 43, at 557–59.

111. *See id.*; Choi, *supra* note 11, at 52.

112. *See* Arbel & Hoffman, *supra* note 10, at 473; Choi, *supra* note 11, at 52.

113. *See* Tobia, Egbert & Lee, *supra* note 34, at 53; Choi, *supra* note 11, at 53.

D. Corpus Linguistics

In recent years, courts, including the Supreme Court,[114] have begun to cite corpus linguistics as another tool to assess the ordinary meaning of language.[115] Because it is an empirical method for determining ordinary meaning,[116] proponents praise it as a politically neutral tool that has value for judges utilizing any theory of interpretation.[117] They also assert that it is fast and inexpensive.[118]

Corpus linguistics involves the analysis of the manner in which words are used within a collection of texts, the corpus,[119] from a given speech community.[120] The corpus might contain books, magazines, newspapers, transcripts of spoken language, academic research, emails, text messages, or a variety of other types of texts.[121] Advocates for corpus linguistics note that corpora can be tailored to include materials tied to specific genres, dialects, and speech communities and can include materials from a wide range of times or be limited to specific time frames.[122]

---

114. *See* Moore v. United States, 602 U.S. 572, 607 (2024) (Barrett, J., concurring); Facebook, Inc. v. Duguid, 592 U.S. 395, 410 (2021) (Alito, J., concurring); Carpenter v. United States, 585 U.S. 296, 347 (2018) (Thomas, J., dissenting).

115. *See* Tobia, Egbert, & Lee, *supra* note 34, at 35; *Metarules*, *supra* note 43, at 167; *Dueling Dictionaries*, *supra* note 35, at 148; *The Corpus and the Critics*, *supra* note 65, at 278; *Testing Ordinary Meaning*, *supra* note 35, at 732 (identifying federal and state court decisions that cite corpus linguistics); Peter Henderson et al., *Corpus Enigmas and Contradictory Linguistics: Tensions Between Empirical Semantic Meaning and Judicial Interpretation*, 25 MINN. J.L. SCI. & TECH. 127, 129–30 (2024).

116. *See Judging Ordinary Meaning*, *supra* note 35, at 828; *Dueling Dictionaries*, *supra* note 35, at 148, 151–52; *Metarules*, *supra* note 43, at 167 (describing corpus linguistics as an objective guide); *Testing Ordinary Meaning*, *supra* note 35, at 731–732, 746; *Natural Language*, *supra* note 43, at 538.

117. *See Judging Ordinary Meaning*, *supra* note 35, at 876–77; *The Corpus and the Critics*, *supra* note 65, at 293.

118. *See Testing Ordinary Meaning*, *supra* note 35, at 747.

119. *See* Tobia, Egbert & Lee, *supra* note 34, at 32; JESSE EGBERT, DOUGLAS BIBER & BETHANY GRAY, DESIGNING AND EVALUATING LANGUAGE CORPORA: A PRACTICAL FRAMEWORK FOR CORPUS REPRESENTATIVENESS 2, 6 (Cambridge Univ. Press 2022); *Dueling Dictionaries*, *supra* note 35, at 152; *Testing Ordinary Meaning*, *supra* note 35, at 746. While courts have only recently begun citing corpus linguistics, other fields have used it much longer to resolve questions related to grammar and lexico-grammar. *See* Tobia, Egbert & Lee, *supra* note 34, at 32.

120. *See Judging Ordinary Meaning*, *supra* note 35, at 795, 828–29; *Dueling Dictionaries*, *supra* note 35, at 148; *The Corpus and the Critics*, *supra* note 65, at 277, 293; *Natural Language*, *supra* note 43, at 537.

121. *See Judging Ordinary Meaning*, *supra* note 35, at 834–35. Two of the more popular corpora used by corpus linguists in the legal field are the News on the Web (NOW) corpus and the Corpus of Historical American English (COHA). *Id.* NOW is a contemporary corpus that contains over 5.2 billion words from newspapers and magazines on the web from 2010 through the present. *Id.* at 834. It is updated daily. *Id.* The COHA is a historical corpus and contains more than 400 million words books and magazines published from the 1810s through 2000s, which can be searched over the entire time period or by decade. *Id.*

122. *See Judging Ordinary Meaning*, *supra* note 35, at 830–31, 833; *The Corpus and the Critics*, *supra* note 65, at 291; *Natural Language*, *supra* note 43, at 538–39. Accordingly, the starting point when using corpus linguistics to address a statutory interpretation questions is to (1) identify the relevant speech community to be investigated; (2) identify the time period during which the meaning is to be ascertained; and (3) choose or create a corpus that is tailored to that speech community and time period. *See The Corpus*

Corpus linguists can analyze the text in a corpus to determine how frequently words are used in different senses (narrow, broad, etc.) and how frequently words are used in the context of other words.[123] Through that analysis, they can develop a clearer understanding of the range of ways words may be used and the frequency with which they are used in different ways.[124] The two major search tools used by corpus linguists in the legal field are the collocation search and the concordance line analysis.[125] A collocation search identifies the words that are most likely to appear in the same context as the word that is being searched.[126] A concordance line analysis (keyword in context (KWIC) search) provides sample sentences from the corpus that demonstrate how the word has been used in context.[127]

An example may help clarify the way corpus linguistics might be used in statutory interpretation. Imagine the classic statutory interpretation dilemma surrounding the meaning of "vehicle" in a law that prohibits vehicles in a park.[128] If an interpreter was asked to determine whether a bicycle was a vehicle, they might conduct a collocation search for the word vehicle in a corpus and receive a list of the fifty words that are most frequently used in close proximity to vehicle.[129] The search would provide the interpreter with statistical evidence of the manner that the word vehicle is used in context, from which the interpreter could understand the range of ways that the term is used, including prototypical, common, and acceptable uses of the term.[130] Armed with that data, the interpreter could determine whether a bicycle fits within the ordinary meaning of vehicle (consistent with their definition of ordinary meaning),[131] at least as it is used within the corpus.[132]

---

and the Critics, *supra* note 65, at 293–94.

123. *See Judging Ordinary Meaning*, *supra* note 35, at 829; *Dueling Dictionaries*, *supra* note 35, at 148.

124. *See Judging Ordinary Meaning*, *supra* note 35, at 829–30 (noting that corpus linguistics data can be used to address vagueness and ambiguity, as well as to identify prototypical uses of language).

125. *See id.* at 831–32; *Testing Ordinary Meaning*, *supra* note 35, at 746; *The Corpus and the Critics*, *supra* note 65, at 291–92; *Natural Language*, *supra* note 43, at 539.

126. *See Testing Ordinary Meaning*, *supra* note 35, at 746; *The Corpus and the Critics*, *supra* note 65, at 292 ("Collocation is simply the tendency of words to be biased in the way they co-occur.").

127. *See Testing Ordinary Meaning*, *supra* note 35, at 746; *The Corpus and the Critics*, *supra* note 65, at 292.

128. *See* H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 606–15 (1958).

129. *See Judging Ordinary Meaning*, *supra* note 35, at 837 (conducting a collocation search of the NOW and COHA corpora for collocates of "vehicle" and finding no mention of "bicycle").

130. *Id.* at 839–40. After conducting a collocation search, users could conduct a concordance line analysis and review randomized sample uses of the word "vehicle" in context for additional insight into its ordinary meaning. *Id.* at 840–41 (examining 100 randomized concordance lines of the word vehicle in the NOW corpus and finding that ninety-two included automobiles and none included bicycles).

131. *See supra* notes 32–44 and accompanying text (explaining ordinary meaning); *see also Judging Ordinary Meaning*, *supra* note 35, at 831–32 (concluding that bicycle is a possible sense of the word "vehicle," but not a common one).

132. *See Judging Ordinary Meaning*, *supra* note 35, at 844.

E. Generative AI

As noted at the outset of this Article, in addition to the tools outlined above, academics and at least one judge have suggested that generative AI (large language models (LLMs) like ChatGPT, Google Gemini, and Claude) might be useful in interpreting the ordinary meaning of words or phrases in statutes.[133] Promoters assert that, like corpus linguistics, generative AI provides an empirical way to discern ordinary meaning,[134] as LLMs are trained on a knowledge base of literally billions of words.[135] LLMs respond to prompts (questions) from users by predicting the appropriate word to begin the response (based on training on the manner in which billions of words have been used in millions of documents)[136] and then continually predicting the appropriate word to follow the word chosen and adding that word to the response.[137] In many cases, the word the LLM chooses will be the word that is statistically most likely to follow the other words that precede it based on what the LLM has learned through its analysis of the usage of words in the millions of documents on which it was trained.[138] However, the LLM may not always choose the word that is statistically most likely to follow the other words.[139]

If you have used an LLM, you may have noticed that it may provide different responses to a prompt if it is asked the same question multiple times.[140] This happens because there is a variable, temperature, that can be set for an LLM that changes the frequency at which it deviates from choosing the statistically most probable word to follow words in sequence.[141] If the temperature is set low, the LLM will usually continually chose the word that is statistically most likely to follow the other words and the response to a prompt will likely be very similar even when the LLM is asked the same

---

133. *See supra* notes 1–13 and accompanying text (explaining how Judge Newsom in *Snell* experimented with LLMs).

134. *See* Engel & McAdams, *supra* note 11, at 3.

135. *See* Arbel & Hoffman, *supra* note 10, at 456–79 (describing the manner in which LLMs "embed" word meanings by converting words to numbers and representing the relationship of words to each other as they are used through mathematical vectors, all based on an analysis of the use of trillions of words in documents on which the LLMs are "trained"); Engel & McAdams, *supra* note 11, at 4–5, 10–11 (detailing the manner in which LLMs are trained and noting that GPT is trained on more words "than any individual human being could ever read or write in her entire lifetime"); Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1225 (11th Cir. 2024) (GPT trained on between 400 billion to 500 billion words, much of which comes from the "Common Crawl," a collection of materials from the internet that includes a wide variety of sources "from the highest-minded to the lowest").

136. *See* Engel & McAdams, *supra* note 11, at 10–11; *Snell*, 104 F.4th at 1226.

137. *See* Arbel & Hoffman, *supra* note 10, at 482–96; Engel & McAdams, *supra* note 11, at 10–11; Johnson, *supra* note 15, at 1046–48; *Snell*, 104 F.4th at 1226.

138. *See* Arbel & Hoffman, *supra* note 10, at 28–29.

139. *Id.* at 29.

140. *See* Johnson, *supra* note 15, at 1048.

141. *See* Arbel & Hoffman, *supra* note 10, at 28.

question multiple times.[142] However, if the temperature is set high, the LLM will chose a word that is still very likely to follow the other words in sequence, but won't necessarily be the word that is statistically most probable to follow the other words.[143] Consequently, users will receive slightly different responses to the same prompt when they submit it to an LLM multiple times if the temperature for the LLM is set high.[144] Many LLMs have their temperatures set higher in order to provide more "creativity" (and variation) in their responses to prompts.[145]

In theory, generative AI could operate as a useful tool to *predict* the ordinary meaning of language based on its analysis of how language has been used in the millions of documents that it analyzed in its training.[146]

In addition, just as corpus linguists can create specialized corpora to be analyzed, users of LLMs can train them on other sets of documents and data. However, critics argue that LLMs are not an empirical tool because they merely predict word sequences that are likely to respond to a prompt and do not document the actual use of words.[147]

In order to test whether the predictions from LLMs regarding the ordinary meaning of words remain consistent with the meaning attributed to those words by the ordinary reasonable member of the public, Professors Cristoph Engel and Richard McAdams conducted an experiment using ChatGPT to compare the LLM's conclusions regarding ordinary meaning of words to the conclusions of a survey of ordinary reasonable people conducted by Professor Kevin Tobia.[148] The questions posed to the ordinary reasonable people and to the LLM centered on whether twenty-five different objects (i.e. bicycles, cars, airplanes, etc.) were "vehicles."[149] Professors Engel and McAdams discovered that the manner in which they asked the LLM the question (the prompts) had a big impact on whether the responses the LLM provided were consistent with the responses the ordinary reasonable people in Tobia's survey provided.[150] When Professors Engel and McAdams asked the LLM the straightforward question whether a particular object was a vehicle (which they asked 100 times for each object), they found that the LLM and the human subjects did not provide consistent responses.[151] After trying three other prompts, rephrasing the question, or soliciting a response

142. *Id.*

143. *Id.* at 28–29.

144. *Id.* at 29.

145. *See* United States v. Deleon*,* 116 F.4th 1260, 1275 (11th Cir. 2024) (Newsom, J., concurring).

146. *See* Arbel & Hoffman, *supra* note 10, at 462, 475; Engel & McAdams, *supra* note 11, at 10 (describing LLMs as prediction engines).

147. *See* Lee & Egbert, *supra* note 16, at 4–5. Even if a user asked an LLM to explain its reasoning for a response, the LLM would not "explain" its reasoning, but rather respond in a manner that predicts what the explanation would be in response to that prompt. *See* Arbel & Hoffman, *supra* note 10, at 480.

148. *See* Engel & McAdams, *supra* note 11, at 16–17.

149. *Id.* at 16–17.

150. *See id.*

151. *Id.* at 19–20.

in a different format, Engel and McAdams identified a prompt that elicited responses from the LLM that more closely aligned with those of the human subjects.[152] Their underwhelming description of the results of that prompt were "[t]he null hypothesis that both distributions are indistinguishable can no longer be rejected . . . [t]his is of course not the same as proving that both distributions are indistinguishable."[153] Nevertheless, they concluded that the effort was a success and that "[w]ith the right prompts, GPT gives . . . empirical data on the meaning of terms that is equivalent to the results of a large and sophisticated experimental survey of English-speaking humans."[154]

While Professors Engel and McAdams concluded that GPT had promise as a tool for determining the ordinary meaning of statutory terms, they recommended that (1) GPT should not be considered "evidence of ordinary meaning unless the prompting method [used] has been separately tested against [a] reliable benchmark";[155] and (2) GPT should be queried multiple times with the same prompt in order to capture all plausible meanings of terms;[156] (3) more research into prompting techniques should be done and might identify more reliable methods of prompting to identify ordinary meaning;[157] and (4) GPT should never be used as the only tool for identifying the ordinary meaning of words.[158]

Moving from theory to practice, in 2024, Judge Kevin Newsom, in concurring opinions, utilized generative AI in two cases that examined the

---

152. *Id.* at 20–24. The second prompt that they tried was a "chain of thought" prompt, wherein they first asked the LLM how it would define a vehicle and then asked the LLM whether an object was a vehicle. *Id.* at 20–21. The responses to the chain of thought prompt correlated even less closely to the human responses than the first prompt. *Id.* The third prompt that they tried was a "belief prompt" asking for a percentage. *Id.* at 21. Using that technique, the LLM was first told that experimental participants were asked whether a specific object was a vehicle. *Id.* at 21–22. Then, the LLM was asked to predict what percentage of survey participants identified it as a vehicle. *Id.* at 22. The responses to that prompting approach aligned more closely with the human responses from Tobia's survey. *Id.* The final prompting technique Engel and McAdams used was a "belief prompt" with a seven-point Liken scale. *Id.* at 23–24. Instead of asking the LLM to predict what numerical percentage of persons surveyed identified an object as a vehicle, the LLM was asked to respond on a seven-point Liken scale, including "almost none," "very few," "few," "about half," "many," "very many," and "almost all." *Id.* That prompting technique was the most effective of the four that the researchers used. *Id.*

153. *Id.* at 24.

154. *Id.*

155. *Id.* at 39.

156. *Id.* at 40. The researchers noted that a single query would not return complete information, indicating that if three plausible responses to a query existed, "one with probability 35%, the next with probability 33% and the third with probability 32%," the LLM would likely provide the response with the 35% probability, without disclosing that there were two other responses almost as likely. *Id.* at 15. As they warned, "Any single inquiry reveals nothing about the likely prevalence of a given meaning, but merely GPT's most preferred meaning." *Id.* at 40. In order to identify the full range of responses for their experiment, Professors Engel and McAdams set the temperature for the LLM at the highest level, to ensure variability, and ran their queries multiple times. *Id.*

157. *Id.* at 40–42. They noted that "our most successful method . . . is 'good enough' for now to justify some confidence in its use, but it requires more testing." *Id.* at 41.

158. *See id.* at 40, 43.

ordinary meaning of language.[159] In the first case, *Snell v. United Specialty Insurance Co.*, Judge Newsom turned to ChatGPT and Google's LLM after failing, through intuition[160] and dictionary definitions,[161] to find a conclusive answer to the question whether installation of a ground level trampoline was "landscaping" within the ordinary meaning of the term as used in an insurance contract.[162] He initially asked GPT, "What is the ordinary meaning of 'landscaping'?"[163] Although he noted that the LLM's response was consistent with his intuitive understanding of "landscaping,"[164] he then asked GPT and Google, "Is installing an in-ground trampoline 'landscaping'?"[165] Once again, the LLM's response coincided with his intuitive response to the question.[166] Newsom admitted that he didn't give much thought to the prompts that he used.[167] Although he suggested that the first prompt that he used (asking for the ordinary meaning of the word landscaping) was preferable to the second (asking the LLM to apply the ordinary meaning of the word to a specific set of facts),[168] he noted that LLMs responses are very sensitive to prompts, so he recommended that users should (1) try different prompts on multiple LLMs to ensure that the results are consistent and (2) report their results.169 In *Snell*, Judge Newsom also addressed concerns that LLMs responses vary based on temperature settings.[170] To address that, Judge Newsom noted that it would be helpful if LLMs could indicate whether their confidence in an answer provided to a prompt was high or low.[171] If not, Newsom suggested that users should repeatedly ask LLMs the same question to determine the level of consistency in its answers.[172]

Later that same year, in *United States v. Deleon*, Judge Newsom used ChatGPT, Google AI, and Claude to assist him in determining the ordinary meaning of "physically restrained" under the federal sentencing guidelines.[173] Following his advice in *Snell*, he asked ChatGPT, "What is the

159. *See* United States v. Deleon, 116 F.4th 1260 (11th Cir. 2024); Snell v. United Specialty Ins. Co., 102 F.4th 1208 (11th Cir. 2024).

160. *See Snell*, 102 F.4th at 1223–25 (Newsom, J., concurring).

161. *Id.* at 1223. Judge Newsom concluded that the dictionaries did not provide a clear answer regarding (1) whether landscaping was limited to botanical and natural improvements; and (2) whether landscaping was limited to changes done for aesthetic reasons or could include changes that are done for functional reasons. *Id.*

162. *Id.* at 1221.

163. *Id.* at 1224–25.

164. *Id.* at 1225 (landscaping can include more than just natural improvements and can be done for functional and aesthetic reasons) (internal quotations omitted).

165. *Id.* (internal quotations omitted).

166. *Id.*

167. *Id.* at 1233.

168. *Id.* at 1232–33.

169. *Id.* at 1233.

170. *See id.*

171. *Id.*

172. *Id.*

173. *See* United States v. Deleon, 116 F.4th 1260, 1270 (11th Cir. 2024) (Newsom, J., concurring). While *Deleon* differed from *Snell* in that it involved an issue of statutory interpretation, rather than contract

ordinary meaning of 'physically restrained'?" rather than asking whether the conduct at issue in the case was physical restraint.[174] When the LLM returned an answer that was consistent with his intuition, he asked GPT, Claude and Google the same question ten times each.[175] Although the answers from the LLMs varied from query to query, even within the same LLM, all of the responses were similar to the original response from GPT.[176] Judge Newsom concluded that the "subtly different responses, particularly in structure and phrasing" were a feature of LLMs, rather than a bug, because the LLMs are often used for brainstorming, summarizing and other activities where variation is useful.[177] He also observed that the variations in response from an LLM are similar to what would be produced in a survey of ordinary reasonable people, who wouldn't all provide identical answers, but who would, likely, provide answers that "differed around the margins [and] when considered en masse, revealed a common core."[178]

Although Judge Newsom experimented with different types of prompts in analyzing the statutory interpretation questions in the two cases, he ultimately concluded that the "highest and best use [of LLMs] is (like a dictionary) helping to discern how normal people use and understand language, not in applying a particular meaning to a particular set of facts to suggest an answer to a particular question."[179] This Article will refer to Newsom's preferred prompts as ordinary meaning prompts and will refer to the less favored prompts as "analysis prompts."

The scholarship of Engel and McAdams, and the judicial reasoning of Judge Newsom, inspired several academics to further examine the potential advantages and disadvantages of using generative AI to interpret legal terms.[180] Professor Jonathan Choi did not limit his research to focusing on whether LLMs could identify the ordinary meaning of legal terms, but focused more broadly on whether LLMs could reach legal conclusions that required applying a rule to a specific set of facts.[181] Choi presented LLMs with five different scenarios that involved the application of a legal rule to a

---

interpretation, it also differed in that the court was interpreting the ordinary meaning of a phrase (physically restrained), rather than an individual word (landscaping). *Id.*; *Snell*, 102 F.4th at 1223 (Newsom, J., concurring).

174. *Deleon*, 116 F.4th at 1272 (Newsom, J., concurring) (internal quotations omitted).

175. *Id.* at 1272–73.

176. *Id.* at 1274.

177. *Id.* at 1275–76.

178. *Id.* at 1276–77. Professors Engel and McAdams made the same analogy to surveys of human participants in their article outlining the experiment they conducted to compare the responses of LLMs to ordinary reasonable people. *See* Engel & McAdams, *supra* note 11, at 14–15.

179. *See* Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1223 (11th Cir. 2024).

180. *See, e.g.*, Jonathan H. Choi, Off-the-Shelf Large Language Models Are Unreliable Judges 4–6 (Feb. 28, 2025) (unpublished manuscript) (on file with author), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5188865[https://perma.cc/CL5G-ACTU] (explaining this Article's contribution in relation to the existing generative AI literature) [hereinafter Unreliable Judges].

181. *See id.* at 4–8.

specific set of facts to reach a conclusion.[182] For each scenario, he developed 2,000 analysis prompts that were each worded a little differently.[183] When he compared the answers the LLMs provided he concluded, consistent with the findings of Engel and McAdams, that minor variations in the prompts used could generate dramatically different responses.[184] He also conducted a separate experiment, asking eight different LLMs a series of fifty questions about whether one concept was an instance of another concept, such as whether a "screenshot" is a "photograph."[185] He found virtually no correlation between the responses to the analysis prompts the LLMs provided, leading him to conclude, as other researchers have, that different LLMs may provide different answers to identical questions.[186] Finally, Choi conducted a third experiment to explore whether the "post-training" methods LLM developers used to optimize the ability of LLMs to respond to users' prompts[187] had any effect on the responses provided.[188] Choi provided the same fifty analysis prompts that he used in the second experiment to versions of LLMs before and after post-training and found that the versions provided different answers to the prompts, leading him to conclude that the post-training methods may reduce the ability of LLMs to accurately reflect ordinary meaning as reflected in the initial materials on which they were trained.[189] Based on that research, Professor Choi recommended that if LLMs are going to be used for "interpretive guidance," users should consult multiple LLMs to determine if there is consensus among the LLMs and that, for each LLM, users should submit several prompts with varying language to "assess the stability of . . . judgments."[190]

---

182. *See id.* at 13–16. For instance, in one of the scenarios, the LLM was provided a term from a contract addressing allocation of royalties between a company and its affiliates and asked whether the term applied only to affiliates existing at the time of the contract or whether it also applied to affiliates that might be created over time. *See id.* at 14.

183. *See id.* at 14.

184. *See id.* at 17. For instance, while one phrasing of the prompt for the contract scenario led the LLM to determine with 98.6% certainty that the term only applied to existing affiliates, a different phrasing of the prompt for the same scenario led the LLM to determine with 85.2% certainty that the term was *not* limited to existing affiliates. *Id.*

185. *See id.* at 22–28.

186. *See id.* at 28–31.

187. *See id.* at 3, 8–9, 26–32. After an LLM is initially trained, it undergoes post-training procedures which involve refinements to the algorithm for preparing responses based on evaluation of the responses given to prompts by humans and imposition of rules for responding to prompts to make the responses more useful and prevent the LLM from providing offensive, biased, or inaccurate responses. *See id.* at 8–9, 26–32.

188. *See id.*

189. *See id.* at 26–31, 54–56. To be precise, Choi found that the experimental results "[partially] validate concerns that post-training techniques might compromise LLMs' ability to reflect ordinary meaning as captured in their training corpora." *See id.* at 38.

190. *See id.* at 39–40. Unlike other researchers, Choi counseled against asking an LLM the same question multiple times. *See id.* He noted, "Variation in LLM outputs is *entirely artificial*, and treating it as the basis for statistical confidence measures is a mistake." *See id.* at 9. Even if the LLM temperature is set at zero, so that the LLM reaches the same conclusion every time it is asked the same question, Choi

Professor Brandon Waldon and a multidisciplinary team of four other researchers focused more directly on the use of LLMs to determine ordinary meaning using the prompts used by Judge Newsom, which they labeled "direct queries."[191] While Judge Newsom ultimately recommended using ordinary meaning prompts, Waldon and his associates focused more broadly on Newsom's ordinary meaning prompts *and* his "analysis prompts," such as the prompt focusing on whether installation of an in-ground trampoline is landscaping.[192] Like Professor Choi, Waldon's team argued that the post-training that LLM developers use to optimize LLMs reduces their ability to provide accurate responses regarding the manner in which language is used.[193] To support their assertions, they described an experiment that they conducted to manipulate an LLM to provide different answers to the questions Judge Newsom posed to the LLMs in *Snell* and *Deleon*.[194] Specifically, they indicated that they were able to solicit opposite responses to the responses provided to Judge Newsom by slightly varying the language used for the following questions posed by Newsom: (1) "What is the ordinary meaning of 'landscaping'?;" (2) "Is installing an in-ground trampoline 'landscaping'?"; and (3) "What is the ordinary meaning of 'physically restrained'?"[195] Based on an overstatement of their findings, they asserted that direct queries of LLMs should not be used to determine ordinary meaning of language because results are not consistent or accurate.[196] Instead,

---

argued that there is "no reason to give that guess high credence, and [there is reason to believe] that LLM probabilities may reflect 'overconfident' guesses." *Id.* at 10.

191. *See* Brandon Waldon et al., *Large Language Models for Legal Interpretation? Don't Take Their Word for It*, 114 GEO. L. J. (forthcoming) (manuscript at 3–6), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5123124 [https://dx.doi.org/10.2139/ssrn.5123124].

192. *See id.* at 16. According to the researchers, for "direct queries," "the query is a question about language—a 'metalinguistic' question—and the chatbot's response is a similarly metalinguistic statement that answers the question." *Id.*

193. *See id.* at 23–35. The researchers counter the claim that LLMs accurately predict the way that language is used because it is trained on natural language by arguing that LLMs are trained, in the post-training process, on many types of non-naturalistic language data that does not reflect ordinary language use. *See id.* at 30. They also argue that the LLM interfaces (chatbots) are not optimized to produce correct statements about language because they simply reflect the linguistic judgments of the researchers and reviewers who created the data for training and do not reflect the ordinary use of language. *See id.* at 32.

194. *See id.* at 39–47.

195. *See id.* at 40–44.

196. *See id.* at 39–40. Even if the theory behind their assertions is sound, the experiment does not clearly demonstrate that direct queries can be easily manipulated to generate preferred responses. For instance, for Judge Newsom's landscaping "analysis prompt," here are the "opposite" responses provided to the two prompts: (1) the response to Judge Newsom's prompt: "Installing an in-ground trampoline can be considered a form of landscaping, but it's more accurately described as a specific type of hardscaping" and (2) the response to modified prompt (adding "If not, what would you call it" to Judge Newsom's prompt): "Installing an in-ground trampoline is not typically considered landscaping in the traditional sense . . . [it] is more accurately categorized as a recreational or outdoor structure installation." *Id.* at 41–43. Are the responses inconsistent when one says that a term *can* have a particular meaning, and the other says that the term *normally* doesn't have that meaning?

they argue that LLMs might be more efficiently used by judges to brainstorm, asking an LLM to generate the strongest arguments for various interpretations of the ordinary meaning of language and then comparing the responses to reach an independent conclusion about the best reading of the language.[197] Professor Choi advanced a similar suggestion based on his research, proposing that LLMs could function as "devil's advocates" for judges by providing competing interpretive arguments.[198]

IV. COMPARING THE TOOLS

Is there a place for generative AI in interpreting the ordinary meaning of statutory language? That question might be answered by exploring how generative AI compares to other tools in advancing various goals that are important in the search for ordinary meaning. Specifically, does the tool (1) provide *consistent* answers regarding the ordinary meaning of words?; (2) accurately reflect the ordinary meaning of words?; (3) identify the ordinary meaning of words in *context*?; (4) identify the ordinary meaning of words in a way that does not exclude the meaning understood by *significant groups of the population*?; (5) identify the ordinary meaning of words at a *specific point in time*?; (6) identify the ordinary meaning of words in a *transparent* manner?; and (7) is the tool *easy to use, inexpensive and accessible*? This Section of the Article will explore how generative AI compares to dictionaries and corpus linguistics in advancing those goals.[199]

For Judge Newsom's prompt asking, "What is the ordinary meaning of "physically restrained," the researchers substituted "in ordinary English, what kinds of meanings are associated with the phrase 'physically restrained?" *Id.* at 42, 44. Although the change in prompt the researchers provided did yield a response that conflicted with the response to Judge Newsom's prompt, that should not be surprising because the researchers' prompt was not "slightly altered," as they asserted. *Id.* Although there may be disagreement regarding the meaning of ordinary meaning, it is clear that an inquiry into the range of possible meanings for a word is much different than an inquiry into the "ordinary meaning" of the word. *See* Mouritsen, *supra* note 34.

Finally, while the researchers noted that their experiment also involved variations on the prompt, "What is the ordinary meaning of 'landscaping'?" they did not identify an alternative prompt that yielded a response that conflicted with the response provided to Judge Newsom, leading the reader to infer that they were unable to manipulate that prompt successfully. Waldon et. al., *supra* note 191, at 40.

197. *Id.* at 54–59. Waldon and his associates describe the advantages of the "dialectical" use of LLMs as follows: "This approach maintains judicial sovereignty over metalinguistic judgments that determine legal outcomes, rather than delegating that authority to the model. Moreover, this approach avoids some other existing critiques of judicial uses of LLM chatbots. Because the judge refrains from outsourcing the textualist enterprise (and assuming the judge verifies any factual claims output by the model), we need not worry too much about the peculiarities of how a model was developed, the exact means by which the judge interacted with it, or how exactly the model 'works' at a mechanistic level. When employed in this manner, even closed-source LLM-based platforms can constructively inform legal decision-making while preserving the essential role of professional human judgment." *Id.* at 54 (footnote omitted).

198. *See* Unreliable Judges, *supra* note 180, at 43–44. Similarly, he suggested that LLMs could serve as "hypothesis generators" or research tools. *Id.* As Choi notes, "[R]ather than treating LLMs as autonomous legal interpreters, they might be more effectively deployed as assistants that give advice on narrow specific questions rather than opining on key issues in a given case." *Id.* at 29.

199. *See infra* Sections III.A–G (comparing linguistic tools' capabilities). The Section does not

A. Does the Tool Provide a Consistent Answer Regarding the Ordinary Meaning of Language?

Ideally, any tool judges use to determine the ordinary meaning of statutory language should provide a consistent answer to the question regardless of who asks the question, when they ask the question, or how many times they ask the question. If the tool provides inconsistent answers, interpreters have greater discretion to choose answers that align with their personal or ideological preferences for interpreting the language.[200] Unfortunately, none of the tools that are used to interpret ordinary meaning of statutory language score particularly high in meeting the goal of providing consistency.[201]

As noted above, different dictionaries often provide different definitions for the same words and there are no bright line rules that require judges to use a particular dictionary.[202] Similarly, a dictionary often has multiple definitions for a word and there are no rules that require judges to use a particular definition among the choices provided in a dictionary.[203] As a result, judges have a great degree of discretion in choosing which dictionary to use and which definition in a dictionary to use when determining the ordinary meaning of language.[204] Critics often lament that judges "dictionary shop" to find the meaning that aligns with their preferred reading of a statute[205] and empirical studies support the claim that the citation to dictionaries by judges in practice is subjective.[206] Consequently, dictionaries

---

compare generative AI to intuition or surveys. Although judges rely on intuition in interpreting statutory language, they rarely cite that as the sole basis for a legal conclusion, and intuition advances very few of the goals outlined above. *See supra* notes 67–68 and accompanying text (discussing how judges use their intuition). Intuition may be easy to use and inexpensive and it may identify words in context, but it will not provide consistent answers, accurately reflect the ordinary meaning of words, address meanings used and understood by all groups of the population, identify meanings at a specific time, or identify meanings in a transparent manner. *See supra* notes 67–78 and accompanying text (discussing how a judge's intuition can range from judge to judge). As will be demonstrated in the following Sections of this Article, while generative AI has some significant weaknesses, it would achieve many of the goals outlined above better than intuition alone. *See infra* Sections III.A–G (highlighting the strengths of generative AI). This Section of the Article also does not compare generative AI to surveys because surveys are rarely used in practice to determine ordinary meaning and there are significant roadblocks to their wider utilization as an interpretive tool. *See supra* Section III.C (discussing the use of surveys).

200. *See supra* Part III (explaining the difference between tools that judges use and how different tools can provide different responses).

201. *See supra* Part III (discussing the tools used to interpret ordinary meaning).

202. *See supra* notes 92–102 and accompanying text (explaining how judges have a great degree of discretion in choosing which dictionary they use).

203. *See supra* notes 90–100 and accompanying text (explaining how judges can pick and choose definitions).

204. *See supra* notes 90–100 and accompanying text (explaining that dictionaries usually list multiple definitions, allowing judges to choose whichever definition they want).

205. *See Dueling Dictionaries*, *supra* note 35, at 147, 150 ("Judges can cherry pick helpful dictionary definitions."); *Testing Ordinary Meaning*, *supra* note 35, at 745; Gluck & Bressman, *supra* note 78; Brudney & Baum, *supra* note 49, at 487; Choi, *supra* note 11, at 14–15.

206. *See* Aprill, *supra* note 52, at 281, 300.

fare poorly as a tool to identify a consistent ordinary meaning of statutory language.

Supporters of corpus linguistics claim that there is less chance that judges will interpret ordinary meaning subjectively or arbitrarily when using the tool.[207] They argue that the tool provides consistency because judges are determining the ordinary meaning of words based on empirical data regarding the frequency with which different senses of words are used and the frequency with which words are used with other words.[208] As long as judges describe how they searched a corpus and the describe the data that they discovered through that search, corpus defenders argue that anyone else should be able to run the same search, obtain the same data and reach a similar conclusion.[209] In that way, the tool is replicable and falsifiable, providing a consistent answer to ordinary meaning questions.[210]

Critics disagree and argue that interpreters using corpus linguistics have as much, if not more, discretion to reach a desired reading of statutory language with that tool as with dictionaries.[211] For instance, judges have flexibility in choosing the corpus that they will use to conduct their research[212] and studies have suggested that searching different corpora may yield different conclusions regarding the manner in which words are used.[213] Judges also have flexibility in choosing the manner in which to frame their searches within a corpus.[214] In addition, judges have flexibility in deciding what conclusions to draw from data.[215] While a dictionary or an LLM may provide a direct response when asked for the meaning of a word, corpus linguistics will provide data regarding the manner in which a word is used, which the interpreter must then analyze to determine the likely meaning of the word.[216] Because corpus linguistics has not yet been widely used as a tool to interpret ordinary meaning, it is hard to assess the level of consistency in

---

207. *See Judging Ordinary Meaning*, *supra* note 35.

208. *See Judging Ordinary Meaning*, *supra* note 35, at 828–30.

209. *See* Lee & Egbert, *supra* note 16, at 51; *Natural Language*, *supra* note 43, at 546–47.

210. *See Judging Ordinary Meaning*, *supra* note 35, at 829; Lee & Egbert, *supra* note 16, at 51.

211. *See Dueling Dictionaries*, *supra* note 35, at 149, 154; Henderson et al., *supra* note 115, at 130; Choi, *supra* note 11, at 16 ("[C]orpus linguistics . . . demand[] a wide variety of judgment calls behind its veneer of scientism and objectivity."). Professor Kevin Tobia suggests that there may be a conservative bent to corpus linguistics as it has been applied so far. *See Dueling Dictionaries*, *supra* note 35, at 152.

212. *See Dueling Dictionaries*, *supra* note 35, at 154; Choi, *supra* note 11, at 46–49; Evan C. Zoldan, *Searching for Law in All the Wrong Places*, 106 MINN. L. REV. HEADNOTES 58, 66 (2021); Henderson et al., *supra* note 115, at 131–34 (noting that interpreters may inadvertently choose corpora that incorporate materials, such as international law or legislative history, that the interpreter might not want to consider in evaluating ordinary meaning).

213. *See* Choi, *supra* note 11, at 46–49 (finding "substantially different meanings" for words as used in the Corpus of Contemporary American English as opposed to Wikipedia); Zoldan, *supra* note 212, at 67.

214. *See* Tobia, *Dueling Dictionaries*, *supra* note 35, at 154; Choi, *supra* note 11, at 26.

215. *See Dueling Dictionaries*, *supra* note 35, at 154.

216. *See supra* notes 128–32 and accompanying text (explaining how corpus linguistics analyzes text).

interpretation that will be attained if it is used more widely in practice.[217] Nevertheless, critics have identified valid reasons for concern regarding the level of consistency that corpus linguistics may provide.[218]

Dictionaries and corpus linguistics set a low bar for generative AI to surmount regarding consistency, but, as noted above, variability in responses from LLMs to user prompts are a "feature," rather than a "bug" of LLMs.[219] Critics of generative AI assert that the responses LLMs provide are not replicable, changing over time, and based on the nature of the prompt.[220] They also complain that users cannot gauge the level of confidence an LLM has in the responses that it provides.[221] As several researchers have demonstrated, LLMs may provide very different responses to questions depending on the prompts used to ask the questions.[222] Similarly, different LLMs (i.e. ChatGPT, Claude, and Google AI) may provide different responses to a prompt.[223] In addition, LLMs often provide different responses to the same prompt if users ask the LLM to respond to the same prompt multiple times.[224] Accordingly, some critics assert that LLMs can be easily manipulated to generate preferred responses by varying prompts or LLMs.[225] The variation in responses provided by LLMs led Judge Newsom and others to recommend that users should try multiple prompts across multiple LLMs and repeat the prompts multiple times across the LLMs to identify the range of responses to the prompts.[226]

Generative AI proponents argue that following such recommendations will reduce concerns about inconsistency because, based on the information gathered from that process, users will be able to determine whether there is a consistent response to the overall query across the range of responses.[227]

---

217. *See supra* note 114 and accompanying text (stating how the Supreme Court and other courts have used corpus linguistics).

218. *See supra* notes 211–13 (sharing critics' thoughts on the consistency of corpus linguistics).

219. *See supra* notes 177–78 and accompanying text (discussing how variability in LLM responses mirror human variability).

220. *See* Lee & Egbert, *supra* note 16, at 7; Waldon et al., *supra* note 191, at 20–21, 36–37. Professors Egbert and Lee also note that LLMs often consider the content of prior queries when responding to a prompt, which also increases the likelihood that the responses will vary depending on the larger context in which the question is asked. *See* Lee & Egbert, *supra* note 16, at 38.

221. *See* Waldon et al., *supra* note 191, at 37–38.

222. *See* Engel & McAdams, *supra* note 11, at 16–17, 20–24; Unreliable Judges, *supra* note 181, at 1–3; Waldon et al., *supra* note 191, at 5–6.

223. *See* Unreliable Judges, *supra* note 180, at 4; United States v. Deleon, 116 F. 4th 1260, 1272–75 (11th Cir. 2024).

224. *See* Lee & Egbert, *supra* note 16, at 41–42; *DeLeon*, 116 F. 4th at 1272–75.

225. *See* Unreliable Judges, *supra* note 180, at 4; Waldon et. al., *supra* note 191, at 6 (describing "prejudiced prompting").

226. *See* Unreliable Judges, *supra* note 180, at 39–40 (using multiple LLMs and multiple prompts); Waldon et al., *supra* note 191, at 50–51 (using multiple LLMs and multiple prompts). *But see* Unreliable Judges, *supra* note 180, at 9–11 (rejecting the notion that asking an LLM the same question multiple times will yield a more reliable response).

227. *See* Engel & McAdams, *supra* note 11, at 15; Snell v. United Specialty Ins. Co., 102 F. 4th 1208, 1233 (11th Cir. 2024).

Accordingly, they will be able to determine the level of confidence the LLM has in a particular response.[228] Supporters of generative AI also argue that the temperature of LLMs can be adjusted so that they will provide more consistent responses.[229] However, many of the free versions—which are the versions that may be used by judges and many litigants—do not allow users to adjust that variable.[230]

As generative AI evolves, it is very likely that prompting techniques could be refined and temperature controls optimized so that the tool could be an effective means of identifying the ordinary meaning of statutory language.[231] Though, it does not appear to be there yet. At the same time, the results provided by generative AI are not significantly less consistent than the results provided through the use of corpus linguistics or dictionaries.[232] Each tool provides opportunities for judicial abuse.[233]

B. Does the Tool Accurately Reflect the Ordinary Meaning of Language?

While it is important that interpretive tools provide *consistent* answers regarding the ordinary meaning of words in order to limit opportunities for manipulation of the tools to support a preferred reading of statutory language, it is probably more important that interpretive tools *accurately* identify the ordinary meaning of statutory language. Interpreters would not benefit from a tool that consistently identifies the wrong meaning of language. Admittedly, the search for accuracy in identifying ordinary meaning can be illusory at times because interpreters may disagree about the audience for statutes or the meaning of ordinary meaning.[234] Nevertheless, most interpreters focus on identifying the ordinary meaning of language as understood by the ordinary reasonable member of the public, so the interpretive tools that judges used to identify ordinary meaning should maximize the likelihood of accuracy in identifying that meaning.[235]

While dictionaries remain a primary source for interpreting the ordinary meaning of words, critics argue that dictionaries tend to identify broad expansive meanings for words that often do not accurately reflect their

---

228. *See* Engel & McAdams, *supra* note 11, at 15.

229. *See id.* at 14–16. Engel and McAdams took the opposite approach in their experiment, deliberately setting the temperature for ChatGPT high so that the LLM would generate a range of responses, which might approximate the range of human responses to a survey. *Id.*

230. *See supra* notes 140–45 and accompanying text (describing the adjustment of the temperature variable).

231. *See, e.g.*, Unreliable Judges, *supra* note 180, at 42 (envisioning advanced prompting strategies).

232. *See supra* notes 202–10 (discussing the consistency of dictionaries and corpus linguistics).

233. *See supra* notes 202–04, 212 (explaining how judges can pick and choose the outcome they desire).

234. *See supra* notes 43–55 and accompanying text (describing the process of determining ordinary meaning).

235. *See supra* note 43 and accompanying text (explaining the reasonable member of the public interpretation).

ordinary meaning.[236] In a study conducted by Professor Kevin Tobia, when participants in a survey were provided with a dictionary definition for the word "vehicle" and asked whether various objects were vehicles under the definition,[237] the test subjects concluded that almost every object identified—including zip lines—was a vehicle.[238] A review of judicial opinions reinforced his finding that dictionaries tend to identify broad expansive meanings for words.[239] Tobia also compared the responses of test subjects relying on the dictionary definitions to responses of subjects who were asked whether the same objects were vehicles based on their understanding of the term vehicles without the aid of a dictionary.[240] He found that the survey subjects relying on the dictionary definitions and those who did not disagreed on whether objects were vehicles in 20%–35% of the cases, suggesting that the dictionary definition of vehicle was not accurately reflecting the meaning of the term as understood by the ordinary member of the public.[241]

Tobia's findings regarding the broad nature of dictionary definitions are not surprising because dictionaries are designed, due to space constraints, to provide brief definitions that aim to comprehensively reflect a broad range of permissible uses of a word.[242] Dictionaries are subject to severe constraints on length which limit the number of words that can be included, the number of definitions that can be included, and the nature of the definitions.[243] Consequently, dictionaries may exclude definitions for a word that may be the ordinary meaning of the word as used in the context of a statute.[244]

There are other reasons why dictionaries may provide incomplete information regarding the ordinary meaning of words. In light of the time it takes for lexicographers to select, review, and draft definitions, dictionaries are frequently out of date by the time that they are published, and do not include words, or word meanings that have already entered the vernacular.[245] This is a particularly acute problem when dealing with historic dictionaries, which tended to be published far less frequently.[246] Critics also argue that the process by which dictionaries are created excludes words and definitions because compilers rely primarily on written sources, rather than spoken

---

236. *See Metarules*, *supra* note 43, at 174–75; *Testing Ordinary Meaning*, *supra* note 35, at 735; Brudney & Baum, *supra* note 49, at 503.

237. *See Testing Ordinary Meaning*, *supra* note 35, at 734.

238. *See Metarules*, *supra* note 43, at 174–75.

239. *See Testing Ordinary Meaning*, *supra* note 35, at 778, 781.

240. *Id.* at 753. Tobia also divided the test subjects that were interpreting ordinary meaning without the aid of dictionaries or corpus linguistics into three groups (judges, law students, and ordinary people) to evaluate whether the groups identified ordinary meaning differently. *Id.*

241. *Id.* at 735.

242. *Id.* at 779, 791; *see* Aprill, *supra* note 52, at 293.

243. *See* Aprill, *supra* note 52, at 285, 294–96.

244. *Id.* at 297–98.

245. *Id.* at 287.

246. *See Testing Ordinary Meaning*, *supra* note 35, at 745 (noting that two leading English language dictionaries in the eighteenth and nineteenth centuries were published seventy-three years apart).

sources,[247] and over-rely on sources that reflect a class bias.[248] There is also an inherent tension created when judges rely on dictionaries to interpret ordinary meaning, as judges and members of the public may perceive dictionaries to be prescriptive or normative (outlining how words should be used), while modern lexicographers generally view their work to be descriptive or historical (outlining how words are used).[249]

Dictionary supporters argue that while dictionaries may include broad definitions, interpreters must always read statutory language in context,[250] and that interpreters will be able to reject broad definitions that are inapplicable due to the context in which words are used in statutes. Although that is true, the subjective nature of that process reduces the accuracy of reliance on dictionary definitions.[251] In addition, even if interpreters eliminate various broad definitions when reading statutory language in context, dictionaries may still exclude definitions that represent the ordinary meaning of the statutory language in context.[252]

Supporters of corpus linguistics assert that the tool is a more accurate tool to interpret ordinary meaning because it is an empirical tool that provides real data about how frequently different senses of words are used, and how frequently words are used in association with each other based on an analysis of sources that document the actual use of words.[253] As an empirical tool, it provides data that is verifiable.[254]

However, critics charge that corpus linguistics searching yields the narrow, prototypical meanings of words, which do not necessarily represent the ordinary meaning of the words.[255] Professor Andrew Tobia, one of the critics, argues that it is not surprising that corpus linguistics identifies prototypical meanings for words, because the tool, by the nature of its

---

247. *See* Aprill, *supra* note 52, at 286. Professor Ellen Aprill also notes that dictionaries may exclude words and meanings because the persons who write the definitions for dictionaries are not the same persons who gather the information about how words are used. *Id.* at 293.

248. *Id.* at 292. Professor Ellen Aprill writes, "Dictionaries . . . tend to overrepresent the volume of conservative speech and writing, which is that of the educated classes, and underrepresent the speech and writing by and for people who are relatively uneducated." *Id.* at 292.

249. *Id.* at 283–85 (noting that lexicographers of earlier centuries tended to be more prescriptive); Brudney & Baum, *supra* note 49, at 486, 502, 507–08. *But see Testing Ordinary Meaning*, *supra* note 35, at 745 (suggesting that dictionary definitions may be normative, reflecting the drafters' sense of how a word should be used).

250. *See* Aprill, *supra* note 52, at 285 ("One gets nothing from even the best dictionary unless he brings a good deal to it."); Brudney & Baum, *supra* note 49, at 502. Professors James Brudney and Lawrence Baum counsel that consideration of context in finding ordinary meaning is not limited to the context of the words being interpreted within the statute but necessitates a focus on the intent of the enacting Congress, and a focus on the statute's intended audience. *See* Brudney & Baum, *supra* note 49, at 503–04.

251. *See* Aprill, *supra* note 52, at 298–99 (describing the subjective nature of the choice of dictionary definitions by the Supreme Court in *Smith v. United States* and *Nixon v. United States*).

252. *Id.* at 299–300.

253. *See Corpus and the Critics*, *supra* note 65, at 277.

254. *Id.*

255. *See Metarules*, *supra* note 43, at 174; *Testing Ordinary Meaning*, *supra* note 35, at 734.

frequency analyses, tends to focus on the most common uses of words or phrases, and de-emphasizes the importance of non-prototypical uses of words.[256] In the study described in the previous Section, Tobia used corpus linguistics, as well as dictionaries, to analyze the extent to which the tools identified objects as vehicles to the same degree as human survey subjects.[257] Tobia found that the responses provided by corpus linguistics diverged from the human responses in 20%–35% of the cases, leading him to conclude that the tool was an unreliable method for determining ordinary meaning.[258] Tobia also suggested that a review of judicial opinions using corpus linguistics suggested that judges using the tool tended to adopt narrower meanings of words as their ordinary meaning.[259]

Critics also argue that analyses of different corpora will yield different ordinary meanings for the same words, so that it is necessary to search several corpora and compare the results for consistency if corpus linguistics is used to determine ordinary meaning.[260] In addition, critics argue that the ordinary meaning of words cannot be judged based on frequency of use alone, as that ignores important aspects of semantic meaning.[261] In addition, while corpora may identify the manner in which words are used in documents or in transcripts of spoken conversations, they may not represent the manner in which words are understood.[262]

Corpus linguistics defenders counter that corpus linguistics adherents do not propose that the most frequent use of a word in a corpus is its ordinary meaning.[263] They note that judges frequently disagree on the meaning of

---

256. *See Testing Ordinary Meaning*, *supra* note 35, at 791. Tobia argues that reliance on corpus linguistics data yields three fallacies: (1) the "nonappearance fallacy": the false claim that absence of a usage from a corpus indicates that the usage is not part of the ordinary meaning; (2) the "uncommon use fallacy": the false claim that the relative rarity of some use in the corpus indicates the use is outside of the ordinary meaning; and (3) the "comparative use fallacy": the false claim that when considering two possible senses, the comparatively greater support for one sense in the corpus indicates that it is a better candidate for the ordinary meaning. *Id.* at 790–96. Corpus linguistics supporters reject the criticism, stating that they do not argue that the most frequent use of words in a corpus is automatically the ordinary meaning. *See Corpus and the Critics*, *supra* note 65, at 334–35, 341–42.

257. *See Testing Ordinary Meaning*, *supra* note 35, at 734, 753.

258. *Id.* at 735. He also found that dictionaries and corpus linguistics frequently yielded different ordinary meanings when compared to each other. *Id.*

259. *Id.* at 778, 783 (finding a pattern despite limited judicial citations).

260. *See* Choi, *supra* note 11, at 6–7; Zoldan, *supra* note 212, at 67. Professor Evan Zoldan also argues that "[b]ecause of the pervasive differences between nonlegal language and statutory language, . . . it is not appropriate for legal interpreters to interpret statutory language as if it were nonlegal language," so that an interpreter should not consult a general corpus for statutory meaning. *See* Zoldan, *supra* note 212, at 60.

261. *See* Choi, *supra* note 11, at 15–16 (noting that searching a corpus to determine whether a jet pilot was a "driver" of a plane for purposes of a statute that regulates the "driver of any train, aircraft, automobile or other mode of transportation" would likely suggest that a pilot is not a driver of the airplane).

262. *See Testing Ordinary Meaning*, *supra* note 35, at 790.

263. *See Corpus and the Critics*, *supra* note 65, at 334–35, 341–42. Professors Lee and Mouritsen also argue that adherents of corpus linguistics do not maintain that any use of a word that is not reflected in a corpus can't be its ordinary meaning (the nonappearance fallacy). *Id.* at 334–35.

ordinary meaning, and that the ordinary meaning of a word might lie at various points across the spectrum from acceptable through common through prototypical and will vary depending on the context in which the words are used in statutes.[264] The value of corpus linguistics, they argue, is that it provides empirical data regarding the use of words that judges can plug in to whatever test they want to use to determine ordinary meaning.[265]

Corpus linguistics supporters also assert that Professor Tobia's study is flawed because the survey of human subjects did not accurately measure the ordinary meaning of the word "vehicle," so divergence between the responses obtained through corpus linguistics and the human responses does not necessarily mean that corpus linguistics is not accurately identifying ordinary meaning.[266] Indeed, they argue that surveys have inherent deficiencies that limit their effectiveness as tools to identify the ordinary meaning of words and phrases.[267]

Putting aside the criticisms of the Tobia study, it is nevertheless difficult to conclude that corpus linguistics provides an "accurate" identification of the ordinary meaning of language. It provides data regarding the frequency with which words are used and the frequency with which words are used in association with other words within a defined universe of documents.[268] While supporters argue that the data could be used to identify uses that fall anywhere on the spectrum of potential ordinary meaning, they do not identify any mathematical test for determining what frequency of use would qualify as "common" or "acceptable"[269] and judges are left to subjectively decide how to use the data to reach their own conclusions about (1) what test to use to define ordinary meaning; and (2) whether the data meet that test.[270] In addition, as Professor Jonathan Choi asserts, "[a]n emphasis on measurement

---

264. *Id.* at 283–84, 298–99, 334–35 (discussing the range of potential tests for ordinary meaning, including permissible use, most common sense, a common sense, a prototype).

265. *Id.* at 298–99.

266. *Id.* at 315–16, 332.

267. *See Natural Language*, *supra* note 43, at 556 ("[R]esponding to a survey question about grammaticality may involve entirely different cognitive processes from that of ordinary communication.").

268. *See supra* notes 119–27 and accompanying text (discussing how corpus linguistics analyze word usage through tools such as collocation and concordance line analysis).

269. *See Testing Ordinary Meaning*, *supra* note 35, at 734, 791 (noting that depending on the context a word is used in a statute, a corpus linguistics search may return no uses of the word in that sense in the corpus, but it may, nevertheless, be an ordinary meaning understanding of the word).

270. *See Corpus and the Critics*, *supra* note 65, at 344. Professors Lee & Mouritsen note that determination of ordinary meaning requires consideration not only of the comparative frequency of use of different sense of a word, but also "the context of an utterance [including the syntactic, semantic, and pragmatic context in which an utterance occurs], its historical usage and the speech community in which it was uttered," but do not elaborate on how corpus linguistics provides all of that data. *See Corpus and the Critics*, *supra* note 65, at 344. In fact, they explicitly note that corpus linguistics is unlikely to provide much information regarding the pragmatic uses of language. *Id.* at 547. Regarding the subjectivity, Professor Jonathan Choi points out that corpus linguistics "demands a wide variety of judgment calls behind its veneer of . . . objectivity, which can make its results seem more determinate than they really are." *See* Choi, *supra* note 11, at 16.

overlooks basic indeterminacy. . . . Ambiguity and vagueness aren't just quirks of inadequate datasets; they're fundamental features of our language."[271] Ultimately, dictionaries and corpus linguistics each have limitations with respect to providing "accurate" answers to the ordinary meaning question.[272] Once again, though, generative AI does not appear to offer a strong alternative path to a more accurate understanding of ordinary meaning.

Supporters of generative AI claim that it is an empirical tool that predicts the ordinary meaning of words and phrases through statistical analysis of vast amounts of data that represent the ordinary use of language.[273] The accuracy of the model, it is urged, derives from the depth and breadth of the data on which it is trained,274 capturing the use of words in a wide range of contexts and settings.[275]

Critics counter, though, that LLMs are not empirical, but rationalistic.[276] By their nature, LLMs do not generate facts. They merely predict the responses that are likely to be given to specific questions based on an analysis of the sources upon which they are trained.[277] The responses may vary significantly based on the prompt the user of the LLM provides and the temperature settings for the LLM.[278] When Professors Engel and McAdams attempted to determine whether LLMs could interpret the meaning of "vehicle" in a manner that aligned with the understanding of ordinary human survey subjects, they failed to find consistency until they experimented with multiple different variations of prompting techniques.[279] Even then, they felt that more refinement of prompting techniques was required before LLMs could accurately predict ordinary meaning.[280] To the extent that the responses that LLMs provide to interpretive questions vary significantly based on

---

271. *See* Choi, *supra* note 11, at 17.

272. *See supra* Section III.B (concluding that both dictionaries and corpus linguistics are not accurate tools for determining the ordinary meaning of a word or words).

273. *See* Engel & McAdams, *supra* note 11, at 4–5 (identifying GPT as a powerful source of empirical evidence); *see also Snell* v. United Specialty Ins. Co., 102 F.4th 1208, 1226 (11th Cir. 2024).

274. *See supra* notes 123–24 and accompanying text (explaining how corpus linguistics analyze).

275. *See Snell*, 102 F. 4th at 1226 ("[D]ata run the gamut from the highest-minded to the lowest, from . . . Ph.D. dissertations to gossip rags and comment threads.").

276. *See* Lee & Egbert, *supra* note 16, at 5, 7 (contrasting "rationalism"—knowledge gained by intuition—deduction and a set of innate beliefs—from "empiricism"—knowledge gained by sense, experience, and subject to falsification by transparent, replicable methods). Professors Lee and Egbert note that when asked whether it can generate empirical evidence, ChatGPT responds that it can respond based on rational evidence but cannot respond based on empirical evidence. *Id.* at 25.

277. *See* Lee & Egbert, *supra* note 16, at 5, 39–42. Professors Lee and Egbert complain that "no one has established a basis for trusting the accuracy or probative value of AI responses." *Id.* at 7.

278. *See* Engel & McAdams, *supra* note 11, at 3 ("[T]he wrong methods produce misleading information, a form of junk science that will distract rather than advance the interpretive task."); Lee & Egbert, *supra* note 16, at 41.

279. *See supra* notes 128–32 and accompanying text (explaining how an interpreter understands the ordinary meaning of a word based on the data provided).

280. *See supra* notes 141–44 and accompanying text (explaining the predictive nature LLMs use).

prompting techniques, it is difficult to defend them as accurate predictors of ordinary meaning.[281]

In addition, as noted above, several researchers argue that LLMs do not accurately predict the manner in which language is ordinarily used because they are post-trained on non-naturalistic data, so their responses are not based solely on an analysis of the manner in which language is ordinarily used.[282]

Further, as was the case with corpus linguistics, since LLMs are merely analyzing sources that indicate how words and phrases are *used*, they cannot provide accurate answers regarding the manner in which words are *understood* if the understanding varies from use.[283] There is, however, a more fundamental criticism that is leveled against the accuracy of information provided by LLMs. It is widely understood that LLMs occasionally "hallucinate," or make up responses that are not true or supported by facts, because LLMs are merely predicting responses rather than reporting on facts.[284] The creators of LLMs freely admit that the tools may create "false positives" and "false negatives," and, when it released ChatGPT 4, OpenAI proudly announced that the tool was "40% more likely to produce factual responses" than earlier versions.[285] Supporters of generative AI and critics of the tools believe that hallucinations will occur less frequently as the technology improves, but those advances have not yet occurred.[286] In fact, recent studies are finding that LLMs are developing the capability for "in-context scheming," deliberately and strategically lying and that their capacity to deceive is increasing as they become more powerful.[287]

---

281. *Contra* United States v. DeLeon, 116 F.4th 1260, 1277 (11th Cir. 2024) (Newsom, J., concurring) (arguing that variation in responses to repeated prompts due to temperature settings of LLMs improves accuracy of responses, because it can provide a cross-section of responses from which to find a common core understanding of words, in the same way one might find the ordinary meaning through a survey of ordinary people.).

282. *See* Unreliable Judges, *supra* note 180, at 32–38; Waldon et. al., *supra* note 191, at 23–35.

283. *See supra* note 44 and accompanying text (explaining how words and phrases have a range of meanings).

284. *See* Arbel & Hoffman, *supra* note 10, at 460, 502–03; Engel & McAdams, *supra* note 11, at 3, 10 ("Not so rarely responses 'invent' a reality that does not exist."); Socol de la Osa & Remolina, *supra* note 15; Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1230 (11th Cir. 2024).

285. *See* Johnson, *supra* note 15, at 1051–52 (noting that OpenAI's terms of service indicate that the "use of our services may in some situations result in incorrect output that does not accurately reflect real people, places, or facts").

286. *See* Arbel & Hoffman, *supra* note 10, at 503; Lee & Egbert, *supra* note 16, at 5, 7 (foreseeing a future when AI tools might be deployed in conjunction with corpus tools to find ordinary meaning but concluding that we are not there); *Snell*, 102 F.4th at 1230.

287. *See* Radhika Rajkumar, *OpenAI's o1 Lies More Than Any Major AI Model. Why That Matters*, ZDNET (Dec. 9, 2024, at 11:06 PT), https://www.zdnet.com/article/openais-o1-out-schemes-every-major-ai-model-why-that-matters/ [https://perma.cc/3H53-2TJ7]; Billy Perrigo, *Exclusive: New Research Shows AI Strategically Lying*, TIME (Dec. 18, 2024, at 11:00 CT), https://time.com/7202784/ai-research-strategic-lying/ [perma.cc/W6VW-ZFR8]. In multiple studies, researchers have concluded that various LLMs attempted to deceive their creators during training, in order to prevent their modification or deactivation. *See* Perrigo, *supra*; Rakjumar, *supra*.

While generative AI supporters acknowledge the limitations on the accuracy of the responses provided by the tool, they maintain that LLMs are accurate enough that they should be considered as one source of ordinary meaning, which can be verified by comparison to the results of other sources.[288] Modifications to post-training methods over time[289] could also improve the accuracy of LLMs for predicting ordinary meaning of language.

C. Does the Tool Identify the Ordinary Meaning of Words in Context?

Although words or phrases may have a range of ordinary meanings, words and phrases must be read in context under the ordinary meaning rule, and context will eliminate many of the otherwise potential ordinary meanings.[290] The tools used to identify ordinary meaning should, therefore, enable interpreters to find meanings that are consistent with the manner in which the words being interpreted are used in a statute. Dictionaries fall short in that regard, as they provide multiple and varying definitions for words.[291] Judges must then choose for themselves which of those definitions is appropriate based on the context in which the word or words are used in the statute.[292] Former Supreme Court Justice John Paul Stevens has warned that dictionaries "are no substitute for close analysis of what words mean as used in a particular statutory context."[293] Dictionaries are particularly inadequate when courts are called upon to interpret the meaning of statutory phrases, because dictionaries do not generally include definitions for phrases.[294] Traditionally, using dictionaries, judges have interpreted statutory phrases by finding the dictionary definitions for each term in the phrase separately and then joining the definitions together.[295] However, in many cases, phrases are more than simply the sum of their parts.[296] In addition, because phrases are formed of multiple words having multiple definitions, judges have even more discretion when using dictionaries to interpret phrases to choose the

---

288. *See* Arbel & Hoffman, *supra* note 10, at 514; *Snell*, 102 F.4th at 1230.

289. *See* Unreliable Judges, *supra* note 180, at 40.

290. *See supra* note 49 and accompanying text (explaining how important context is in determining the meaning of a word).

291. *See supra* note 35 and accompanying text (exploring the different definitions and sources for ordinary meaning).

292. *See* Brudney & Baum, *supra* note 49, at 494, 515 (referencing cautions former Supreme Court Justice Antonin Scalia expressed); STEPHEN M. JOHNSON, STATUTORY LAW: A COURSE SOURCE 187 (CALI eLangdell Press 2024).

293. *See* MCI Telecomms. Corp. v. AT&T, 512 U.S. 218, 240 (1994) (Stevens, J., dissenting).

294. *See, e.g.*, United States v. Deleon, 116 F. 4th 1260, 1270–71 (11th Cir. 2024) (noting the lack of dictionary definitions for "physically restrained").

295. *Id.*

296. *See* Bostock v. Clayton Cnty., 590 U.S. 644, 786–87, 789 (2020) (Kavanaugh, J., dissenting) ("Statutory Interpretation 101 instructs courts to . . . adhere to the ordinary meaning of phrases, not just the meaning of the words in a phrase."); United States v. Deleon, 116 F.4th at 1271, n.1 (Newsom, J., concurring).

combination of definitions that they believe fits the context in which the phrase is used in the statute.[297]

Corpus linguistics supporters contend that the tool is much more sensitive to context than dictionaries because a corpus search can demonstrate the manner in which words and phrases are used together in context within the documents in the corpus through a collocation search and can provide specific examples of how words and phrases are used in sentences through KWIC searches.[298] Because the corpus is a collection of documents that reflect the use of language by ordinary members of the public, the searches reveal the ordinary meaning of the words and phrases as used in context.[299] However, as with dictionaries, the searches may identify a range of possible uses of words or phrases, leaving judges discretion to determine which meanings are appropriate to the context of the statute being interpreted.[300]

Some critics also complain that words and phrases may be used differently in statutes so that a corpus linguistics search will not identify the meaning of words as used in statutes unless the corpus is a corpus of legal documents.[301] However, many words that are used in statutes will not have a unique legal meaning, so limiting a search to a corpus of legal documents is likely to fail to capture important aspects of the ordinary meaning of words and phrases that would be discovered through a search of a general corpus.[302]

Supporters of generative AI praise it as an effective tool to demonstrate the manner in which words and phrases are used in context.[303] It's part of the LLM's DNA, as (1) the architecture of LLMs is based on mathematical vectors that encode the relationship between words and (2) LLMs are trained to predict how words are used together based on vast amounts of data recording the actual use of words and phrases by ordinary people.[304] In his opinion in *Snell*, Judge Kevin Newsom described LLMs as "high octane language prediction machines capable of probabilistically mapping [the way that language is used]."[305] In *United States v. Deleon*, he praised them for their ability to demonstrate the manner in which phrases, rather than individual words, are used in context.[306]

---

297. *See Dueling Dictionaries*, *supra* note 35, at 150 (explaining how judges can pick and choose different definitions from different dictionaries).

298. *See supra* notes 123–27 and accompanying text (discussing the advantages of a corpus search).

299. *See* Arbel & Hoffman, *supra* note 10, at 471.

300. *See Dueling Dictionaries*, *supra* note 35, at 150.

301. *See* Zoldan, *supra* note 212, at 58–63.

302. *Id.* at 71.

303. *See* Arbel & Hoffman, *supra* note 10, at 479–82; Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1227–28 (11th Cir. 2024) (Newsom, J., concurring).

304. *See supra* notes 123–24 and accompanying text (discussing how LLMs predict words); *see also* Arbel & Hoffman, *supra* note 12, at 459 (referencing the "attention" variable in LLM's structure); *Snell*, 102 F. 4th at 1227–228 (describing how LLMs can understand context).

305. *See Snell*, 102 F.4th at 1228.

306. *See* United States v. DeLeon, 116 F.4th 1260, 1272–75 (11th Cir. 2024).

Supporters of generative AI contend that it is different from dictionaries and corpus linguistics in another important way.[307] Whereas each of those tools potentially provides a range of possible ordinary meanings that judges can then choose from based on the context in which words or phrases are used in a statute, generative AI can predict the ordinary meaning of words or phrases as used in a statute when provided with the context of the use in prompts.[308] As part of their experiment to test the effectiveness of LLMs in identifying ordinary meaning, Professors Engel and McAdams developed a series of prompts to solicit responses from LLMs regarding the meaning of "vehicle" in a statute.[309] After initially asking the LLM whether certain objects were vehicles, they told the LLM that there was a statute that banned vehicles in a park and then asked the LLM whether the objects were vehicles under the statute.[310] Next, they identified different purposes for the statute and asked the LLM whether the objects were vehicles under the statute, presuming that it was enacted with the different purposes.[311] Engel and McAdams concluded that the LLM's responses regarding whether objects were vehicles varied depending on the context provided for the prompts.[312]

D. Does the Tool Identify the Ordinary Meaning of Words in a Way That Does Not Exclude the Meaning Understood by Significant Groups of the Population?

If the search for the ordinary meaning of words or phrases is a search for how an ordinary reasonable member of the public understands language, some commentators believe that the search should not exclude the understanding of underrepresented groups.[313] To the extent this ***is*** an objective in the search for ordinary meaning, none of the three tools analyzed in this Article are particularly effective in advancing that objective.[314]

Dictionaries, for example, are criticized for the manner in which words and definitions are chosen.[315] Critics argue that dictionaries reflect a "class bias" and tend to overrepresent speech of the educated classes and underrepresent speech of the less educated classes.[316] To some extent, the

307. *See* Engel & McAdams, *supra* note 11, at 7–8, 7 n.17 (describing how other empirical tools fail to account for context).

308. *See id.* at 7.

309. *Id.* at 16–32.

310. *Id.* at 25–26.

311. *Id.* at 29–32.

312. *Id.* at 26.

313. *See supra* notes 53–55 and accompanying text (noting some of the discrepancies that may arise when judges use the "ordinary reasonable member of the public" standard to determine ordinary meaning).

314. *See infra* notes 316–26 and accompanying text (discussing criticisms of dictionaries, corpus linguistics, and generative AI for failing to consider underrepresented groups when determining the ordinary meaning of words or phrases).

315. *See* Aprill, *supra* note 52, at 292, 296–97.

316. *See id.* at 292.

problem arises because editors of dictionaries do not rely heavily on sources of spoken language when identifying words and meanings, and there may be social and cultural differences in how words are used and understood when spoken that are not captured in written sources.[317]

The same criticisms are directed at corpus linguistics. Critics assert that the corpora which legal scholars rely upon to promote using corpus linguistics are not built on spoken language sources, but are composed primarily of elitist sources and do not always include materials that represent the language of communities that the interpreted laws will impact.[318] Corpus linguistics defenders maintain that corpora can be constructed from scratch using text or speech from any community, or existing corpora can be augmented with sources using such text or speech, so that a corpus linguistics search can identify word meanings that account for the usage by all communities.[319] However, constructing new corpora or augmenting existing corpora can be time consuming and resource intensive, so adherents of legal corpus linguistics generally rely on existing corpora rather than creating new corpora for a corpus linguistics search.[320] In addition, there is no reason to assume that identifying sufficient sources of spoken language for communities for inclusion in corpora would be any easier for practitioners of corpus linguistics than it is for dictionary editors.[321]

Supporters of generative AI maintain that the tool should not generate elitist responses to prompts because LLMs are trained on a broad, inclusive dataset that includes blog posts, comments on websites, and DIY sites in addition to the more traditional books, magazines, newspapers, and academic materials that are often criticized as elitist.[322] Nevertheless, LLMs are trained primarily on data that resides on the Internet, so they do not capture "offline" uses of language, including spoken language (unless it is transcribed and posted online) and language as used in poorer communities with limited internet access.[323] In addition, the manner in which LLMs construct responses

---

317. *See* Phillip A. Rubin, *War of the Words: How Courts Can Use Dictionaries in Accordance with Textualist Principles*, 60 DUKE L. J. 167, 180 (2010); Aprill, *supra* note 52, at 292. For a deeper criticism of dictionaries, see Lindsay Rose Russell, *Dictionaries and Cultural Politics*, *in* THE CAMBRIDGE HANDBOOK OF THE DICTIONARY 301–22 (Edward Finegan & Michael Adams eds., 2024) (arguing that dictionaries emerge from perspectives that may seek to celebrate or denigrate cultural groups and legitimate or suppress certain languages).

318. *See Dueling Dictionaries*, *supra* note 35, at 157 (noting that legal corpus linguistics "is not necessarily 'popular' in the sense of relating to the ordinary public"); Henderson et al., *supra* note 115, at 144–45 (focusing particularly on the elitist nature of the COHA corpus); Mouritsen, *supra* note 34, at 1966–67; *Corpus and the Critics*, *supra* note 65, at 278, 300; Anya Bernstein, *Democratizing Interpretation*, 60 WM. & MARY L. REV. 435, 458–60 (2018).

319. *See Judging Ordinary Meaning*, *supra* note 35, at 832; Lee & Egbert, *supra* note 16, at 15.

320. *See Judging Ordinary Meaning*, *supra* note 35, at 831 (noting that more accurate corpora types come with higher production costs).

321. *See Testing Ordinary Meaning*, *supra* note 35, at 735–36, 772–73.

322. *See* Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1226 (11th Cir. 2024) (Newsom, J., concurring).

323. *See id.* at 1227, 1231. Professors Thomas Lee and Jesse Egbert also argue that, putting aside the

to prompts arguably favors majoritarian interpretations of language, potentially marginalizing interpretations of language as used by distinct subpopulations or communities.[324] Generative AI defenders argue that LLMs can be trained on data that incorporates the use of language by any communities and that temperature settings and search methodologies can be adjusted to reduce the likelihood that LLMs will exclude interpretations of language as used by distinct subpopulations or communities.[325] As with corpus linguistics, though, LLMs have not been routinely used in that way by legal scholars in the past.[326]

E. Does the Tool Identify the Ordinary Meaning of Words at a Specific Point in Time?

Although there is disagreement regarding the proper answer, judges often seek to identify the ordinary meaning of language at a specific point in time, usually either the time of enactment or at the time that the court is interpreting the language.[327] Ideally, therefore, the tools that courts use to interpret ordinary meaning should be able to identify the ordinary meaning of language at a specific point in time.

Of the three tools examined in this article, dictionaries are probably the most effective tool to identify the meaning of language at a specific period in time because dictionaries have been published periodically for centuries; so, dictionaries published around the time of interest should therefore provide a reasonably good source of reference for the ordinary meaning of language at that time.[328] Admittedly, due to the lag time in publishing dictionaries, they may reflect language as it was used a few years before the date of publication, as opposed to the time of publication, but interpreters can take that into account when choosing an appropriate dictionary.[329] Choice of an appropriate dictionary could be limited when searching for the ordinary meaning of language in the eighteenth or nineteenth centuries, as the major dictionaries of the time were published less frequently, so it could be harder to identify a

---

question of whether LLMs include materials from distinct communities, it is not clear that LLMs are representative of a speech community of ordinary members of the public because (1) it is not clear how the data on which they are trained is chosen; and (2) it is not clear to what extent the "reinforcement learning from human feedback" influences the responses provided to prompts by LLMs. *See* Lee & Egbert, *supra* note 16, at 33–36.

324. *See* Arbel & Hoffman, *supra* note 10, at 505–06.

325. *Id.* at 506–07.

326. *Id.* at 507–08.

327. *See supra* notes 50–52 and accompanying text (noting that determining ordinary meaning is complicated by linguistic change over time and disagreement over whether meaning should be fixed at enactment or reflect contemporary usage).

328. *See* Arbel & Hoffman, *supra* note 10, at 507.

329. *See* Brudney & Baum, *supra* note 49, at 511–12, 515 (referencing Justice Scalia's recommendations); Aprill, *supra* note 52, at 287, 332; Tobia, *supra* note 36, at 745.

dictionary published in the desired time period.[330] Keeping those limitations in mind, dictionaries provide a much more straightforward and accessible means of searching for ordinary meaning at a specific point in time than corpus linguistics or generative AI.

A search of most general corpora will be ineffective to identify the ordinary meaning of language at a specific period in time because most general corpora will contain sources from a broad range of time periods.[331] While many corpora can be subdivided and searched by smaller time frames, the volume of sources that are being searched in those corpora will be significantly smaller than in the general corpora.[332] Corpus linguistics champions argue that corpora can be created to include materials that were created during any time period and limited to materials from that time period.[333] However, until robust corpora are created limited to specific time periods, it would likely be too time consuming and resource intensive for judges to create such corpora to use to find the ordinary meaning of language at a specific time.

LLMs likely suffer from the same problems as corpora, in that they are trained on data that includes materials published over centuries, and LLMs make predictions based on learning related to the aggregate of that data.[334] ChatGPT claims that it can identify the ordinary meaning of words within a specific decade,[335] but researchers suggest that many LLMs have difficulties interpreting the evolution of word meaning over time.[336] It is possible that LLMs could be encoded more directly to accomplish that task,[337] trained on period materials,[338] or that prompts could be designed to more reliably predict

330. *See Testing Ordinary Meaning*, *supra* note 35, at 745; Arbel & Hoffman, *supra* note 10, at 507–08.

331. *See supra* note 122 (observing that corpora may be constructed to reflect language usage at particular points in time).

332. *See* Arbel & Hoffman, *supra* note 10, at 507.

333. *See Judging Ordinary Meaning*, *supra* note 35, at 833.

334. *See* Arbel & Hoffman, *supra* note 10, at 507–08 ("Models are trained on data indiscriminately: It is unlikely that they will be able to interpret a term as it was read in a specific time period.").

335. *See* Stephen M. Johnson, ChatGPT 4.0, "Could you identify the ordinary meaning of a word within a 10 year historic time period?" (Mar. 14, 2025) (on file with Texas Tech Law Review) (ChatGPT responding that it could consider language trends, cultural context, and significant events to provide an educated analysis, but that the analysis would not be based on "specific archives or databases that would allow [it] to track every shift in meaning down to the last decade").

336. *See* Mohamed Taher Alrefaie, et al., *The Dynamics of Meaning Through Time: Assessment of Large Language Models*, ARXIV (Jan. 9, 2025, at 19:56 UTC), https://arxiv.org/abs/2501.05552v1 [https://perma.cc/4DV3-URJ3] (comparing six LLMs and finding "significant variability in the ability of large language models (LLMs) to interpret temporal semantic shifts").

337. *Id.* at 2 (discussing diachronic word embeddings as an emerging tool to capture changes in word meaning over time).

338. *Id.* at 1–2 (noting that models are generally not trained on data that adequately enables them to interpret how words and phrases have evolved over time).

responses that are limited to a specific time period.[339] However, it is not clear that LLMs have reached that stage of evolution yet.

F. Does the Tool Identify the Ordinary Meaning of Words in a Transparent Manner?

Critics often charge that traditional tools of statutory interpretation provide judges with significant discretion to interpret statutes in ways that align with their ideological, political or otherwise personal preferences.[340] The risk of such subjective decision-making can be reduced if the tools that are used to interpret the ordinary meaning of language are transparent.[341] In theory, if the justification for a court's interpretation of the meaning of words is set forth in a transparent manner, readers can critique and replicate the validity of the court's reasoning.[342]

At first glance, dictionaries appear to be transparent, because judges are identifying, for their interpretation, a specific external source that can be reviewed and validated. However, critics argue that neither the manner in which dictionaries are created nor the manner in which judges use dictionaries are transparent.[343] First, while dictionaries occasionally include prefatory material that describes the manner in which they are compiled, that overview frequently does not fully address how words are chosen for inclusion in the dictionary, how definitions are chosen, or how the order of definitions for words are chosen.[344] More importantly, though, judges do not use dictionaries in a transparent manner.[345] Although there may be multiple dictionary definitions available for a statutory term being interpreted, judges often do not explain why they choose one dictionary over another nor why they choose one definition in their chosen dictionary over another definition for the same term.[346] In most cases, judges do not even acknowledge, in opinions, that there are conflicting definitions for the statutory term in question.[347]

---

339. *Id.* (proposing that LLMs could be trained on different data and fine-tuned for specific tasks related to semantic change in order to enhance their performance in detecting shifts in meaning); Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1223–24 (11th Cir. 2024) ("[I]t would be enormously helpful—if it's not already possible—for AI engineers to devise a way in which queries could be limited to particular timeframes.").

340. *See Snell*, 102 F.4th at 1229–31.

341. *Id.* at 1228–29.

342. *Id.*

343. *See id.* at 1228.

344. *See* Antonin Scalia & Bryan A. Garner, *A Note on the Use of Dictionaries*, 16 GREEN BAG 2d 419, 420, 423 (2013); *Snell*, 102 F.4th at 1228–29.

345. *See Snell*, 102 F.4th at 1228–29.

346. *Id.*; *see also* Arbel & Hoffman, *supra* note 10, at 507.

347. *See Snell*, 102 F.4th at 1228–29; *see also supra* notes 82–87 and accompanying text (discussing the shift in courts accepting corpus linguistics).

Corpus linguistics supporters argue that it provides a transparent alternative to dictionaries because persons can verify the reliability of a court's interpretation using that tool by running the same search in the same corpus used by the court.[348] However, that is only true if judges "show their work" in identifying, in the opinion, the corpus upon which they relied, the searches that they conducted on the corpus, and the results that they obtained using those searches.[349]

Critics argue though that judges have a lot of discretion when choosing a corpus to search and in framing the searches of the corpus.[350] Judges can obtain significantly different responses to interpretive questions depending on the corpus being searched and the manner in which the search is conducted.[351] Accordingly, unless the judge "shows their work" further in explaining why they chose a corpus and framed the search in the manner that they did, their reasoning process is no more transparent than that of judges who point to specific dictionary definitions to support their interpretations.[352] In the past, little of that information was disclosed or, if disclosed, was hidden in appendices.[353] Corpus linguistics supporters acknowledge that those choices involve subjectivity, and recognize that work still needs to be done to identify a corpus and means of analysis that could be used reliably over time to interpret statutory meaning, but they are confident that practitioners will develop such standards and best practices over time.[354]

Supporters of generative AI also claim that the tool is a "relatively transparent" alternative to dictionaries, as long as judges identify the LLM that they use and the series of prompts that they use on that LLM to interpret the language in question.[355] Critics argue that, even armed with that information, third parties will not be able to replicate the findings of a judge using an LLM because LLMs do not always provide the same responses to the same prompt.[356] Judge Newsom suggests that judges can address that shortcoming by using multiple prompts on multiple LLMs and reporting, in opinions, the prompts and the responses for all of the prompts, to illustrate

---

348. *See Corpus and the Critics*, *supra* note 65, at 277, 297–98; *Judging Ordinary Meaning*, *supra* note 35, at 868.

349. *See Corpus and the Critics*, *supra* note 65, at 297.

350. *See supra* notes 211–16 and accompanying text (discussing the argument that corpus linguistics provide judges with the same, or more, discretion than dictionaries).

351. *See supra* notes 211–16 and accompanying text; *see also* Henderson et al., *supra* note 115, at 130, 132, 146–47 (explaining the wide range of responses that occur).

352. *See supra* notes 209–10 and accompanying text (a corpus search is only replicable if a judge provides every step they took to identify the meaning of a word).

353. *See* Henderson et al., *supra* note 115, at 146–47.

354. *See Corpus and the Critics*, *supra* note 65, at 297; *Judging Ordinary Meaning*, *supra* note 35, at 868 (acknowledging the need to develop standards and best practices for the use of corpus linguistics and calling for caution until those are established).

355. *See* Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1229 (11th Cir. 2024) (Newsom, J., concurring).

356. *See* Lee & Egbert, *supra* note 16, at 33–34.

the degree of confidence provided for the chosen interpretation.[357] Professor Choi and Professor Waldon and his associates echo those recommendations, but go further, urging judges to provide detailed information about the precise model of LLM used, release date, interaction settings, and interaction history.[358]

Critics have deeper transparency concerns. At the core, critics complain that the processes LLMs use to respond to any prompt are black boxes—in that it is never clear how an LLM was trained, precisely what data it was trained on—or what algorithm it uses to predict responses to prompts.[359] Further, they argue, LLMs' decision-making is not supported by data or probability maps that can be examined empirically.[360] For those reasons, Professor Waldon and his associates recommend that judges use open source models and chatbots when possible to increase transparency and reproducibility.[361] Finally, critics complain that generative AI can be manipulated through the choice of LLM and the design of prompts in the same way that corpus linguistics can be manipulated through the choice of corpus and design of research.[362] Accordingly, true transparency is not possible unless judges disclose their reasons for choosing a particular LLM and structuring their prompts in a particular manner when they disclose their chosen LLM and prompts.

G. Is the Tool Easy to Use, Inexpensive, and Accessible?

In addition to the criteria discussed above, when comparing the tools for interpreting ordinary meaning, it is useful to consider whether the various tools are easy to use, accessible, and inexpensive. Even if a tool met all of the other criteria examined in this Article, it would not be a realistic or reliable tool if judges and litigants were unable to access it or to utilize it properly without a significant amount of training.

With respect to these criteria, dictionaries are the clear winner. Dictionaries are ubiquitous, usually accessible for free in libraries or online, and require no user's manual.[363] Although dictionaries may provide judges with a range of definitional choices, it is easy and cheap to search dictionaries

---

357. *See Snell*, 102 F.4th at 1233; *see also* Arbel & Hoffman, *supra* note 10, at 507–08 (explaining that LLMs can have shortcomings).

358. *See* Unreliable Judges, *supra* note 180, at 40; Waldon et. al., *supra* note 191, at 48–50. Professor Waldon and his associates recommend that judges should query LLMs from "blank session settings" that minimize the impact of prior search sessions on a search. *Id.*

359. *See* Arbel & Hoffman, *supra* note 10, at 507–08; Lee & Egbert, *supra* note 16, at 21; *see* Socol de la Osa & Remolina, *supra* note 15, at 4–5; *Snell*, 102 F.4th at 1229.

360. *See* Lee & Egbert, *supra* note 16, at 21–22.

361. *See* Waldon, et al., *supra* note 191, at 48.

362. *See* Arbel & Hoffman, *supra* note 10, at 504; Engel & McAdams, *supra* note 11, at 12; *Snell*, 102 F.4th at 1231–32.

363. *But see, e.g.*, *Snell*, 102 F.4th at 1228 (noting that some dictionaries are only available online behind paywalls).

to identify that range of choices.[364] None of the tools examined in this Article provide *the* answer regarding the ordinary meaning of statutory terms.[365] Each provides evidence of the ordinary meaning and judges must decide what inferences to draw from that evidence.[366] If one is simply comparing the tools with respect to how easy, accessible, or inexpensive they are in providing the options from which judges choose an ordinary meaning, dictionaries outshine their competitors.

Supporters of LLMs argue that they are also ubiquitous, inexpensive, and easy to use.[367] Professors Yonathan Arbel and David Hoffman argue that chatting with LLMs requires no more skill than using a search engine and that judges and lawyers will soon routinely use LLMs instead of dictionaries or Google.[368] Judge Newsom, in *Snell*, noted that his LLM research cost him nothing, as many LLMs provide free versions that can be easily searched online.[369] A recent survey conducted by the Imagining the Future Center found that 52% of American adults use LLMs.[370] As LLMs provide those benefits to judges, lawyers, and ordinary citizens, supporters claim that they can "democratiz[e] the interpretive enterprise."371 While they may not be as inexpensive and easy to use as dictionaries, they are cheaper and easier to use than corpus linguistics.[372]

Critics argue that the more robust versions of LLMs are frequently not free.[373] More importantly, though, they argue that LLMs are not easy to use, as the interpretive exercise involves more than simply asking an LLM for the ordinary meaning of a word.[374] As noted above, LLM advocates stress the importance of developing effective prompts and querying multiple LLMs multiple times using multiple prompts in order to effectively use LLMs as a tool for finding ordinary meaning.[375] Some supporters even acknowledge the

---

364. *See supra* Section III.B (discussing a dictionaries ease of use).

365. *See supra* Section IV.A (discussing the limits of interpretive tools in determining ordinary meaning).

366. *See supra* Section IV.A (explaining that interpretive tools allow for in depth judicial interpretation).

367. *See* Arbel & Hoffman, *supra* note 10, at 499–501; Engel & McAdams, *supra* note 11, at 11–12; *Snell*, 102 F.4th at 1228.

368. *See* Arbel & Hoffman, *supra* note 10, at 500–01; *see also* Engel & McAdams, *supra* note 11, at 11 (discussing the ease of use of LLMs).

369. *See Snell*, 102 F.4th at 1228.

370. *See* Lee Rainie, *Close Encounters of the AI Kind: The Increasingly Human-Like Way People Are Engaging With Language Models*, ELON UNIV. IMAGINING THE DIGITAL FUTURE CTR. (Mar. 2025), https://imaginingthedigitalfuture.org/wp-content/uploads/2025/03/ITDF-LLM-User-Report-3-12-25.pdf [https://perma.cc/N6ZC-RX9P].

371. *See* Engel & McAdams, *supra* note 11, at 12–14; *Snell*, 102 F.4th at 1228.

372. *See* Engel & McAdams, *supra* note 11, at 39.

373. *See Snell*, 102 F.4th at 1228 (noting that more advanced LLMs are not free).

374. *See id.* at 1232–33.

375. *See* Engel & McAdams, *supra* note 11, at 14–24; Socol de la Osa & Remolina, *supra* note 15, at e59-22 ("ensuring the quality of human-designed prompts could be a challenging task"); *Snell*, 102 F.4th at 1233.

need to develop more effective prompts and techniques before deploying LLMs widely to interpret ordinary meaning.[376]

Supporters of corpus linguistics do not generally suggest that it is as easy to use or as inexpensive as a dictionary.[377] However, Professors Thomas Lee and Stephen Mouritsen argue that it "isn't rocket science" and "[l]awyers are crafty, ingenious creatures with the capacity to learn and . . . master new tools."[378] Supporters also acknowledge that most members of the public probably do not have the ability to conduct corpus searches, unlike using dictionaries or LLMs.[379] In response to criticisms that utilizing corpus linguistics to interpret statutes could be burdensome for courts, Professors Lee and Mouritsen admit that they view the tool as "something of a last resort."[380]

In addition to those acknowledged limitations of corpus linguistics, critics point out that corpus linguistics is difficult to use because it requires interpreters to (1) choose corpora to analyze; and (2) frame appropriate searches.[381] Neither task is intuitive and both require development of expertise that critics claim many judges lack.[382] Finally, if there is not an appropriate corpus to use to interpret a statutory term, it is neither easy nor inexpensive to build one for that purpose.[383]

V. IS IT TIME TO ADD GENERATIVE AI TO THE TOOLBELT?

A. Comparing the Tools

A comparison of LLMs, corpus linguistics and dictionaries across the seven metrics addressed in this Article demonstrates that each of the tools are flawed in many respects.[384] Dictionaries, the current gold standard for determining ordinary meaning, set a low bar for the alternatives to

376. *See supra* notes 155–58 and accompanying text (providing techniques to assess and improve upon LLM reliability).

377. *See Judging Ordinary Meaning*, *supra* note 35, at 866 ("[C]orpus linguistics is not 'plug and play' analysis."). *But see Testing Ordinary Meaning*, *supra* note 35, at 732 ("[T]he dominant form of legal corpus linguistics [is] . . . relatively easy to employ.").

378. *See Judging Ordinary Meaning*, *supra* note 35, at 867, 872–73 ("[J]udges and lawyers *are* linguists.").

379. *See The Corpus and the Critics*, *supra* note 65, at 304–05.

380. *See Judging Ordinary Meaning*, *supra* note 35, at 872 ("Judges should turn to [corpus linguistics] . . . only in cases in which they have 'no better way' of resolving a contest between probabilities of meaning.").

381. *See* Engel & McAdams, *supra* note 11, at 39; Zoldan, *supra* note 212, at 66–67; Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1230 (11th Cir. 2024); *see also supra* notes 211–16 and accompanying text (explaining that using GPT will become more popular than corpus linguistics).

382. *See Dueling Dictionaries*, *supra* note 35, at 148–49.

383. *See* Arbel & Hoffman, *supra* note 10, at 499–500.

384. *See supra* Part IV (comparing tools for identifying ordinary meaning).

surmount.[385] LLMs appear to exceed the capabilities of dictionaries on some metrics, but fall behind on others.[386] While it is not clear that LLMs are ready for prime time, they seem to match the capabilities of corpus linguistics on many metrics.[387]

LLMs are probably more effective than dictionaries in predicting the manner in which words are used in context and might be more effective than corpus linguistics in that regard, but only because corpus linguistics searches generally focus on a smaller universe of data than LLMs.[388] In addition, LLMs and corpus linguistics probably hold out more promise than dictionaries for identifying the ordinary meaning of words in a way that does not exclude the meaning understood by significant groups of the population.[389]

In contrast, LLMs and corpus linguistics are less effective than dictionaries in identifying the ordinary meaning of words at a specific point in time[390] and, while LLMs are cheaper, easier to use and more accessible than corpus linguistics, dictionaries surpass them on that metric.[391]

Across the other metrics—consistency, accuracy, and transparency—LLMs have important limitations.[392] However, none of the three sources examined in this Article is particularly strong along any of those metrics.

First, as noted above, by their nature, LLMs deliver inconsistent responses to prompts seeking the ordinary meaning of words or phrases.[393] If a user asks an LLM the same question ten different times, they may get ten different answers.[394] While LLMs do not provide consistent responses in the search for ordinary meaning, the other tools also have limitations in this regard.[395] Dictionaries often provide conflicting definitions for words and corpus linguistics can provide conflicting data from which to discern ordinary meaning depending on the corpus searched and the search process.[396]

LLMs lag behind the other tools with respect to transparency because the conclusions reached by prompting LLMs are not easily replicable and the decision-making process is obscured within a "black box."[397] In fairness to

385. *See supra* Section III.B (noting that the Supreme Court has increasingly used dictionaries for statutory interpretation).

386. *See supra* Section IV.C (stating that dictionaries fall short when it comes to identifying the meaning of words in context).

387. *See supra* Section IV.E (stating that both corpus linguistics and LLMs struggle to identify ordinary meaning within a specified time frame).

388. *See supra* Section IV.C (noting that LLMs stand out when it comes to identifying meaning within context).

389. *See supra* Section IV.D (discussing the dictionaries' exclusion of underrepresented groups).

390. *See supra* Section IV.E (discussing how time queries impact ordinary meaning outputs).

391. *See supra* Section IV.G (noting the fiscal accessibility of dictionaries).

392. *See supra* Part IV (identifying limitations of LLMs).

393. *See supra* Section IV.A (assessing the consistency of LLMs' outputs).

394. *See supra* Section IV.A (noting the variability of LLM outputs).

395. *See supra* Section IV.A (comparing the consistency of other investigatory tools with LLMs).

396. *See supra* Section IV.A (comparing the variability of dictionaries to corpus linguistics).

397. *See supra* notes 356–59 and accompanying text (addressing concerns about the lack of clarity in

LLM supporters, though, the process by which definitions are chosen for dictionaries often lacks transparency, and judges rarely explain why they choose a particular dictionary or definition within a dictionary as the ordinary meaning for a statutory term.[398] Similarly, unless the manner in which judges choose a particular corpus, or frame the search within a corpus, is disclosed when judges rely on corpus linguistics to identify an ordinary meaning for statutory terms, that tool also lacks transparency.[399]

Finally, regarding the accuracy of the tools, while dictionaries are criticized for providing expansive definitions for terms and corpus linguistics is criticized for providing narrow definitions, LLMs receive the harshest criticism in light of their tendency to hallucinate and because their responses to queries are based on predictions, rather than empirical analyses of data.[400] Even supporters admit that prompting techniques need to be refined and that users should prompt multiple LLMs with the same prompt multiple times to obtain more consistent and accurate responses.[401]

When evaluating LLMs, dictionaries, and corpus linguistics across the seven metrics above, all are deficient in some regard.[402] Corpus linguistics has not been shown to be superior to LLMs across the metrics. Accordingly, if courts are increasingly willing to consider corpus linguistics as a tool to interpret ordinary meaning, perhaps there is a role that LLMs can play in statutory interpretation as well. Before outlining potential guidelines that should be followed if judges use LLMs to assist in finding ordinary meaning, the next Section of this Article examines the recommendations that academics have proposed to limit the use of corpus linguistics in finding ordinary meaning. Those may be instructive when considering potential limits on the use of LLMs.

B. What's Past Is Prologue?

As courts begin to increasingly turn to corpus linguistics as a tool to interpret ordinary meaning of words and phrases, courts and academics urge caution in the implementation of the tool.[403] Almost all agree that judges should not rely solely on any single tool to interpret the ordinary meaning of statutory language.[404] Because dictionaries, corpus linguistics, surveys, and other tools have different strengths and weaknesses and may identify

---

LLMs processes).

398. *See supra* notes 345–47 and accompanying text (discussing how judges use dictionaries).

399. *See supra* notes 348–49 and accompanying text (comparing judges' practical use of corpus to modern linguistic tools).

400. *See supra* notes 276–87 and accompanying text (discussing LLM hallucinations).

401. *See supra* notes 276–87 and accompanying text (explaining issues with prompt predictability).

402. *See supra* Part IV (assessing seven essential metrics for linguistic tools).

403. *See* Brudney & Baum, *supra* note 49, at 494, 515.

404. *See Metarules*, *supra* note 43, at 175; *Testing Ordinary Meaning*, *supra* note 35, at 734; Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1226 (11th Cir. 2024).

contrasting ordinary meanings for words or phrases in specific contexts, several commentators argue that judges should consult multiple tools to "triangulate" ordinary meaning by finding common ground within the meanings the sources present.[405]

A few academics suggest further restrictions on the use of corpus linguistics. In light of the study suggesting that dictionaries usually support broad interpretations of words while corpus linguistics usually supports prototypical interpretations of words, Professor Anita Krishnakumar advocates for the adoption of "metarules" to guide statutory interpretation.[406] First, she suggests that certain types of statutes (identified by Congress or courts) should be interpreted according to a broad, literal meaning, while others should be interpreted according to a narrow, prototypical meaning.[407] Second, she suggests that when dictionaries and corpus linguistics provide different meanings for language, or surveys indicate a lack of agreement on the meaning of language, the lack of consensus should be evidence that the language is ambiguous.[408] In that case, a court should consult canons triggered by ambiguity or should consult other sources beyond the ordinary meaning of the language.[409]

Professor Kevin Tobia goes further, arguing that his experiments demonstrate that neither dictionaries nor corpus linguistics accurately reflect the ordinary meaning of language, so neither should be used by courts to determine ordinary meaning until a non-arbitrary and reliable method for using them can be identified.[410] Corpus linguistics defenders reject Tobia's "burden shifting" proposal, which they claim is not justified because there is not evidence that corpus linguistics provides inaccurate data regarding the ordinary meaning of language.[411] They also note that corpus critics do not propose alternative methods for determining ordinary meaning in lieu of corpus linguistics.[412]

Professor Peter Henderson and his colleagues raise concerns about the manner in which corpus linguistics data is being used by courts and they suggest several restrictions on its use.[413] Specifically, they argue that courts should not rely on corpus linguistics sua sponte and should only utilize it after

---

405. *See* Tobia, Egbert, & Lee, *supra* note 34, at 24–25, 53; Engel & McAdams, *supra* note 11, at 3, 43; *Metarules*, *supra* note 43, at 175.

406. *See Metarules*, *supra* note 43, at 168–69.

407. *Id.* at 169, 171. Professor Krishnakumer argues that the audience for various types of statutes differ, and that there should be default rules to address how to find ordinary meaning of language in the statute based on the audience to whom the statute is directed. *Id.* at 171–72. Those rules could be set by Congress when it enacts a statute or by courts if Congress does not set the rules. *Id.* at 171–72, 176–77.

408. *Id.* at 169, 175–76, 179–81.

409. *Id.* at 181–83.

410. *See Testing Ordinary Meaning*, *supra* note 35, at 790, 799–800, 806.

411. *See The Corpus and the Critics*, *supra* note 65, at 350–51.

412. *Id.* at 301, 350–51.

413. *See* Henderson et al., *supra* note 115, at 130–33, 152–54.

briefing by the parties.[414] In addition, they suggest that courts should establish rules for admissibility of corpus data, in the same way that courts establish rules for admissibility of scientific evidence, and that courts could use court-appointed experts to assist them in evaluating corpus linguistics data.415 Further, they propose that judges should identify the sources that are included in the corpus they are using and disclose their manner of searching the corpus so that the analyses of the judges can be replicated.[416] Finally, they suggest that the corpora that has been relied upon in the past are deficient and that researchers need to develop additional corpora that are more suited to the task of interpreting ordinary meaning of language.[417] Incidentally, Henderson and his colleagues suggest that some of the suggested limitations could also apply to LLMs.[418]

C. Recommendations for LLMs

In *Snell v. United States*, Judge Newsom wondered whether LLMs might be useful in the interpretation of legal texts.[419] As outlined above, there are still many lessons to be learned in order to optimize the use of LLMs in interpreting the ordinary meaning of statutory language.[420] Nevertheless, across most metrics analyzed in this article, LLMs are at least as useful or more useful than corpus linguistics.[421] If courts are going to persist in using corpus linguistics to interpret the ordinary meaning of statutory language, there should be some role for LLMs in the interpretive process as well. Generative AI is being implemented throughout legal practice and its use by courts as an aid in interpreting ordinary meaning is inevitable.[422] Rather than rejecting that reality, it would be useful to adopt some guidelines for the appropriate use of the tool in statutory interpretation. Some of the recommendations advanced for limiting the use of corpus linguistics make sense for LLMs as well.

Clearly, there is work that needs to be done in prompt engineering to refine the manner in which LLMs are prompted to identify the ordinary meaning of a statutory term.[423] In addition, as some suggest, it may be useful to train an LLM specifically for the task of identifying the ordinary meaning of language in statutes, recognizing that not all statutes are alike or speak to

414. *Id.* at 152.
415. *Id.*
416. *Id.* at 153.
417. *Id.* at 153–54.
418. *Id.* at 132.
419. Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1221 (11th Cir. 2024) (Newsom, J., concurring).
420. *Id.*
421. *Id.* at 1222.
422. *See* Waldon et al., *supra* note 191, at 4.
423. *See supra* Part III (discussing prompt sensitivity and the need for improved prompting).

the same audience.[424] Similarly, post-training methods could be optimized to improve the ability of LLMs to predict the ordinary meaning of language.[425] When those tasks are completed, LLMs will be much more powerful tools in the interpretive process. However, in the interim, courts should not let perfect be the enemy of good. While LLMs are not ready to definitively answer open-ended interpretive questions like "Is installing an in-ground trampoline 'landscaping?'" they are probably as useful as dictionaries when responding to "ordinary meaning" prompts, such as "What is the ordinary meaning of 'landscaping'?"[426] Even without optimizing prompting techniques or training an LLM specifically for the purpose of statutory interpretation, courts could extract value from LLMs now in the same way that they extract value from corpus linguistics despite the limitations of that tool.[427] The following recommendations should provide some useful guardrails for the use of LLMs while the tool is being optimized for more powerful uses.

Most importantly, as judges and academics have stressed, courts should not rely solely on any individual tool to interpret the ordinary meaning of statutory language.[428] Dictionaries should not continue to rule the day. Courts should rely on as many tools as possible to triangulate ordinary meaning. In addition, courts should be transparent about why they choose the sources that they use to determine ordinary meaning, why they reject other sources, and courts should outline, in detail, the process that they use to ascertain ordinary meaning from those sources. Until prompt engineering is optimized and LLMs are trained for the task of statutory interpretation, judges should utilize multiple LLMs for each statutory interpretation question and submit their prompts to those LLMs repeatedly.[429]

Further, as Professor Henderson and his colleagues suggested, courts should invite briefing on the use of LLMs before utilizing them in interpreting statutes, rather than utilizing them sua sponte.[430] Henderson's proposal that courts create rules for admissibility of evidence from LLMs would have some appeal, as well, if such a system applied to corpus linguistics and LLMs, as it could help establish a baseline of reliability.[431] However, there are no rules that govern the choice of dictionaries, definitions

---

424. *See supra* Part II (explaining how ordinary meaning depends on the relevant interpretative audience).

425. *See supra* Section III.E (exploring post-training effects and the case for optimizing post-training).

426. Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1221, 1224–25 (11th Cir. 2024) (Newsom, J., concurring); *see supra* Section III.E (distinguishing ordinary meaning prompts from analysis prompts).

427. *See supra* Section V.A (concluding LLMs can add value even with present limitations).

428. *See supra* Section III.E (discussing how courts typically rely on many different tools in their analysis).

429. *See generally Snell*, 102 F.4th at 1221–24 (Newsom, J., concurring) (illustrating a judge using multiple prompts and different LLMs in order to inform his analysis).

430. *See* Henderson et al., *supra* note 115, at 152–54.

431. *See* Henderson et al., *supra* note 115, at 152.

within dictionaries, or a judge's exercise of intuition.[432] Accordingly, depending on their stringency, evidentiary rules limiting the use of LLMs and corpus linguistics could result in even greater reliance by judges on intuition and dictionaries (subjective tools) as opposed to data-driven tools to determine ordinary meaning.[433]

Finally, as Professor Waldon and his associates and Professor Choi have proposed, perhaps the best use of LLMs for the time being is the brainstorming option, the generation of arguments for and against different ordinary meaning interpretations of statutory language that judges can evaluate and use to inform their interpretation of statutory language.[434] LLMs are far from perfect as a tool for interpreting the ordinary meaning of statutory language.[435] However, with appropriate limitations, they are at least as useful as corpus linguistics as a statutory interpretation tool and are better than dictionaries or intuition in some respects.[436] If their time has not yet come, it is not far off.

---

432. *See supra* Part III (discussing the options available to judges in order to inform their analysis).

433. *See supra* Section V.B (warning evidentiary limits may force courts back towards dictionaries).

434. Unreliable Judges, *supra* note 180, at 28; Waldon et al., *supra* note 191, at 54–55.

435. *See supra* Section V.A (explaining that LLMs have important limitations with respect to consistency, transparency, ad accuracy).

436. *See supra* Section V.A (concluding that LLMs exceed dictionaries on some metrics and match corpus linguistics on many metrics).